

## Research Paper

# De re/de dicto distinction: a logicians' perspective on data anonymity

Junli Jiang <sup>1</sup> and Pavel Naumov <sup>2,\*</sup><sup>1</sup>Institute of Logic and Intelligence, Southwest University, Tian Sheng Road, Chongqing 400715, China<sup>2</sup>School of Electronics and Computer Science, University of Southampton, University Road, Southampton, S017 1BJ, United Kingdom\*Corresponding author. School of Electronics and Computer Science, University of Southampton, University Road, Southampton, S017 1BJ, United Kingdom. E-mail: [pgn2@cornell.edu](mailto:pgn2@cornell.edu)

Received 16 December 2023; revised 30 July 2024; accepted 31 October 2024

## Abstract

The distinction between *de re* (of the thing) and *de dicto* (of what is said) readings of sentences has long been the topic of studies in logic and philosophy of language. The article proposes to apply these concepts to anonymity. It argues that, in the proposed setting, *de dicto* knowledge preserves anonymity, while *de re* knowledge does not. The article also considers a third, “overt,” form of knowledge. The main technical result is a sound and complete logical system that captures the interplay between a data traceability expression and the *de re*, *de dicto*, and overt knowledge modalities. The article also shows that the three knowledge modalities are not definable through each other even in the presence of the traceability expression.

**Keywords:** *de re*; *de dicto*; data anonymity; axiomatization; completeness; undefinability

## Introduction

Imagine that you have a nice neighbor whom you really like as a person. At the same time, you happen to hate rich people. Your neighbor is rich, but you do not know this. Thus, you like your neighbor and you hate him at the same time. However, you have these feelings in two distinct senses. You like him as an individual, but you hate the designator (“rich person”) that applies to him. In the philosophy of language, this distinction is often called *de re/de dicto* distinction [1–6]. You like the neighbor *de re* (of the thing) and hate him *de dicto* (of what is said). The *de re/de dicto* distinction is an umbrella term that is understood differently by different scholars and applies not only to preferences. As another example, imagine that you are travelling on a train and a fellow passenger starts giving you surprising medical advice. You most likely will not trust the stranger *de re*. However, you might be the type of person who trusts all doctors. If the passenger, unknown to you, happened to be a doctor, then you would trust her *de dicto* despite not trusting *de re*.

In addition to our examples of preferences and trust, *de re/de dicto* distinction also manifests itself in knowledge. Imagine a young woman Ann who is walking through a park. Ann was given up for adoption as a baby and she has never met her mom. Ann is curious about who her mom might be, so she has asked a close friend to

look for the mom. While walking through the park, Ann sees an old woman sitting on a bench. Unknown to both of them, the woman is Ann’s mother. Let us now consider two possible scenarios of what might happen next. In the first scenario, the woman tells Ann that she, the woman, is sick. In this case, Ann knows that the woman is sick, the woman is Ann’s mother, thus, Ann *de re* knows that her mom is sick. In the second scenario, while chatting with the woman on the bench about the weather, Ann gets a text from her friend. The text tells her that the friend was able to trace Ann’s mother and has learned that the mom is sick. In this case, Ann knows *de dicto* that her mom is sick. Note that these two forms of knowledge are very different. Under the first scenario, Ann might tell the woman that Ann is sorry that she, the woman, is sick. She might recommend a doctor. It would be strange for Ann to recommend a doctor under the second scenario.

As our final example, imagine that Ann is your neighbor and that you are working as a technician at a health clinic. Ann visits the clinic, and a nurse there draws Ann’s blood sample and sends it to you with a doctor’s order to test if “the patient” is pregnant. You conduct the test and learn that the patient indeed is pregnant. The patient is your neighbor Ann, but you don’t know this. In spite of this, you now *de dicto* know that Ann is pregnant. Because you know this *de*

dicto, your knowledge is really about the designator “patient,” not about Ann as a human being. This knowledge, as long as it stays de dicto, does not violate Ann’s *anonymity*. However, if you can *trace* the blood sample to your neighbor Ann, then you will know de re that she is pregnant. This knowledge will break Ann’s anonymity and violate her privacy. As this example shows, de re/de dicto distinction, when applied to knowledge based on data (such as test results), gives a way to characterize anonymity. Of course, privacy/anonymity is a broad term whose different aspects might be captured by different formal approaches. We are not suggesting here that de re/de dicto distinction can replace other, already existing, approaches to defining anonymity. But we think it can complement them in a meaningful way.

In this article, we propose a formal logical system that describes the interaction between de re (non-anonymized) knowledge based on data and de dicto (anonymized) forms of such knowledge. In addition to modalities representing those two forms of knowledge, the system also includes *data tracability* expression and a third, *overt*, knowledge modality. To define a formal semantics of this logical system, we combine two previously independent research areas: two-dimensional egocentric logic and data-informed epistemic logic. The main technical results of this work are the undefinability of the proposed modalities through each other and a completeness theorem for the logical system describing their interplay.

The rest of the article is organized as follows. First, we discuss egocentric logics and data-informed knowledge. Then, we give several more formal examples illustrating the difference between de re, de dicto, and overt knowledge. In the “Syntax and semantics” section, we introduce the language of our system and its formal semantics. In the “Undefinability results” section, we show that the modalities capturing these three forms of knowledge are not definable through each other. In the “Axioms” section, we state the axioms and the inference rules of our logical system. We prove their soundness and completeness in the next two sections. In the “Future work” section, we discuss *k*-anonymity and how it can be captured in our setting.

## Preliminaries

### Egocentric logic

Traditionally, the semantics of modal logics is defined in terms of a satisfaction relation  $w \Vdash \varphi$  between a possible world  $w$  and a formula  $\varphi$ . Under such semantics, the formula  $\varphi$  captures a property of the world  $w$ . Prior proposed to consider egocentric logics that capture the *properties of agents* rather than the properties of possible worlds [7]. The semantics of such logics can be defined in terms of a satisfaction relation  $a \Vdash \varphi$  between an agent  $a$  and a formula  $\varphi$ . For example, to express the fact that agent  $a$  is pregnant, we can write  $a \Vdash$  “is pregnant”. In the egocentric setting, the Boolean connectives can be used in the standard way. The statement  $a \Vdash$  “is pregnant”  $\wedge$  “lives in Bath” means that agent  $a$  is pregnant and lives in Bath. Seligman, Liu, and Girard suggested to consider friendship modality  $F$  (read “for each friend”) in the egocentric setting [8,9]. For example, they read the statement  $a \Vdash F$  “lives in Bath” as “each friend of agent  $a$  lives in Bath”. As usual, modalities can be combined with Boolean connectives. For instance, the statement  $a \Vdash F \neg$  (“is pregnant”  $\wedge$  “lives in Bath”) means that agent  $a$  does not have a pregnant friend who lives in Bath. At the same time,  $a \Vdash \neg F$  (“is pregnant”  $\wedge$  “lives in Bath”) means that agent  $a$  does have such a friend. Modality  $F$  is also used in [10,11]. We previously introduced modality  $L$ , which stands for “likes those who” [12]. For example, the statement  $a \Vdash$

$L$  “lives in Bath” means that agent  $a$  likes those who live in Bath. Modalities can be nested. For instance, the following statement  $a \Vdash LL$  “lives in Bath” means that agent  $a$  likes those who like people living in Bath.

Independently from the works on egocentric logics, a new class of two-dimensional semantics have been proposed in the philosophy of language [13]. Under such semantics, the truth value of a sentence depends not only on the current world but on some other parameter sometimes called indexical. Stalnaker suggested visualizing world/indexical combinations as cells in a two-dimensional matrix [14]. Grove and Halpern combined the egocentric logic and two-dimensional semantics ideas by considering a formal logical system where the role of an indexical is played by an agent [15–17]. The formal semantics of their logical system is defined in terms of a ternary satisfaction relation  $w, a \Vdash \varphi$ . For example, the following statement:  $w, a \Vdash$  “is pregnant” means that agent  $a$  is pregnant in world  $w$  (but might not be in some other worlds). In this setting, one can define the “knows about herself” modality  $K$ . Using such a modality, for example, one can say that, in world  $w$ , agent  $a$  knows that she is pregnant:  $w, a \Vdash K$  “is pregnant”. In this article, we refer to the Grove-Halpern class of semantics as (two-dimensional) egocentric semantics. Such semantics has been used to define “know-who” [18,19] and “know how to tell apart” [20] modalities. Besides the introduction of a new modality, an important contribution of [20] is a new technique for proving completeness technique for two-dimensional semantics. The technique builds on Stalnaker [14] matrices to recursively define the canonical model as a limit of an infinite chain of matrices of maximal consistent sets of formulae. In the current article, we further develop this construction by specifying tree structures on the cells of the matrices.

### Data-informed knowledge

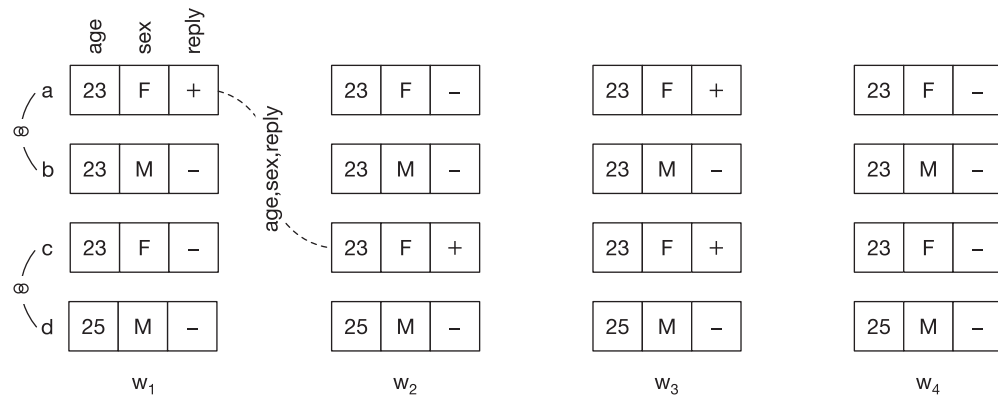
Imagine that there is a digital clock on a wall. We use  $t$  to denote the time shown by the clock. Suppose that in world  $w$  the clock shows 13:24. Thus, anyone who sees the clock would conclude that it is afternoon now. This conclusion can be made by any agent as long as the agent has access to the value of  $t$ . We say that this knowledge is *informed by data variable  $t$*  and write it as

$$w \Vdash K_t \text{ “It is afternoon”}.$$

This type of knowledge in the setting where  $t$  is a Boolean variable has been proposed by Grossi, Lorini, and Schwarzenruber [21]. For arbitrary variables, it was introduced by Baltag and van Benthem [22]. The term “data-informed knowledge” is suggested by us [23,24]. We also considered data-informed beliefs [25,26]. In this work, we consider data-informed knowledge in two-dimensional egocentric settings.

### De re/de dicto data-informed knowledge

Imagine a hypothetical Family Planning survey that asks participants just three questions: their age, sex, and if they expect a child. To keep the example simple, let us assume that the survey has been conducted only among four people, agents  $a, b, c$ , and  $d$ . Suppose that it is common knowledge that agent  $d$  is 25 and the rest are 23. Also, it is common knowledge that agents  $a$  and  $c$  are female and agents  $b$  and  $d$  are male. Assuming only women can be pregnant, there are 4 possible ways the survey can be answered. We call such combinations “possible worlds” and depict them in Fig. 1 as worlds  $w_1, w_2, w_3$ , and  $w_4$ . The figure also shows that  $a$  and  $b$  as well as  $c$  and  $d$  are



**Figure 1.** Possible worlds in a Family Planning survey. Reply + means that the participant is expecting a child.

married couples, but this becomes important only later in our story. Finally, suppose now that the real world is  $w_1$ . In other words, only  $a$  is expecting a child.

If somebody has access only to the anonymized results of agent  $a$ 's answers,  $(23, F, +)$ , then that person would know that the respondent is pregnant. Since the respondent, in this case, is agent  $a$ , one can say that the  $a$ 's survey data *informs the knowledge* that  $a$  is pregnant:

$$w_1, a \models K_{\text{age,sex,reply}} \text{ "is pregnant"}. \quad (1)$$

At the same time, the observer of the anonymized survey data  $(23, F, +)$  would not be able to distinguish this  $a$ 's data in world  $w_1$  from the  $c$ 's data in world  $w_2$ , see Fig. 1. Thus, one can argue, that the  $a$ 's survey data *does not inform* the knowledge that  $a$  is pregnant:

$$w_1, a \not\models K_{\text{age,sex,reply}} \text{ "is pregnant"}. \quad (2)$$

The inconsistency between statements (1) and (2) comes from the fact that they refer to *two different types of knowledge*. In statement (1), modality  $K$  captures the knowledge about the name "participant," which, in the current world, happens to refer to agent  $a$ . In statement (2), modality  $K$  captures the knowledge about the actual participant. As we discussed in the introduction, in the philosophy of language, this distinction between a statement about the name of an object and an actual object is known as *de dicto/de re* distinction. *De dicto* ("of what is said") specifies that a statement refers to the name (such as "participant" in our case) and *de re* ("of the thing") specifies that a statement refers to the actual object. In this article, instead of a single knowledge modality  $K$ , we use modalities  $D$  and  $R$  for *de dicto* and *de re* knowledge, respectively. Thus, in our example,

$$w_1, a \models D_{\text{age,sex,reply}} \text{ "is pregnant"},$$

$$w_1, a \not\models R_{\text{age,sex,reply}} \text{ "is pregnant"}.$$

Because *de re* data-informed knowledge of a formula  $\varphi$  reveals that  $\varphi$  is true about the actual person, it does not preserve the anonymity of the participant and potentially violates her privacy. At the same time, if personal data only informs *de dicto* knowledge, it reveals  $\varphi$  about "the participant," not an actual person. As a result, generally speaking, it preserves the anonymity of the participant.

At this point, the reader might think that *de re* data-informed knowledge is stronger than *de dicto* one. In other words, knowing something *de re* implies knowing *de dicto*. Perhaps surprisingly, this is not true.

Indeed, let us modify our Family Planning survey as shown in Fig. 2. Here, the agents  $a, b, c$ , and  $d$  are the same two couples of

the same age as in the previous example. They are again the only participants in a family planning survey. Let us suppose that instead of sex, the survey asks about the city in which the participant lives. As one can see from the figure, agents  $a$  and  $b$  live in Bath and agents  $c$  and  $d$  live in York. We assume that the city in which each couple lives, just as their ages, is public information. In addition, instead of "Are you expecting a child?," the survey asks "Is your family expecting a child?". The four possible worlds in this new setting are depicted in Fig. 2. Just like before, we assume that only agent  $a$  is pregnant. In other words, the current world is  $w_1$ .

Note that by looking at agent  $a$ 's anonymized survey data,  $(23, \text{Bath}, +)$ , one would not be able to conclude that "the participant" is pregnant because  $a$ 's data is indistinguishable from the survey data for agent  $b$  in the very same world  $w_1$  and agent  $b$  is not pregnant:

$$w_1, a \not\models D_{\text{age,city,reply}} \text{ "is pregnant"}. \quad (3)$$

At the same time, it is easy to see that in *any* possible world, if the survey data for *any* agent is  $(23, \text{Bath}, +)$ , then the agent  $a$  is pregnant in that world, see Fig. 2. Thus, any observer of the anonymized survey data  $(23, \text{Bath}, +)$  would know that agent  $a$  is pregnant. In other words, any observer of agent  $a$ 's anonymized survey data in world  $w_1$  would *de re* know that agent  $a$  is pregnant:

$$w_1, a \models R_{\text{age,city,reply}} \text{ "is pregnant"}. \quad (4)$$

## Overt data-informed knowledge and traceability

In the previous section, we introduced two forms of knowledge that can be informed by *anonymized* data. As we have seen from the two examples, neither of these two forms is stronger than the other in the sense that knowing something *de dicto* does not imply knowing *de re* and vice versa. In this section, we consider the knowledge informed by *non-anonymized* data. That is, we consider what can be concluded from personal data when not only the data is given, but it is also explicitly stated about which agent this data is. We refer to this third form of knowledge as *overt* data-informed knowledge and represent it by modality  $O$ .

To illustrate the difference between the three forms of data-informed knowledge, consider yet another variation of our Family Planning survey depicted in Fig. 3. The only difference from Fig. 2 is that  $c$  had to relocate to Bath while still being married to agent  $d$  who remains in York. Using the same argument as we gave for statement (3), we can see that agent  $a$ 's anonymized personal data does

|   | age   | city | reply |       |      |   |       |      |   |       |      |   |
|---|-------|------|-------|-------|------|---|-------|------|---|-------|------|---|
| a | 23    | Bath | +     | 23    | Bath | - | 23    | Bath | + | 23    | Bath | - |
|   | 23    | Bath | +     | 23    | Bath | - | 23    | Bath | + | 23    | Bath | - |
| c | 23    | York | -     | 23    | York | + | 23    | York | + | 23    | York | - |
|   | 25    | York | -     | 25    | York | + | 25    | York | + | 25    | York | - |
|   | $w_1$ |      |       | $w_2$ |      |   | $w_3$ |      |   | $w_4$ |      |   |

**Figure 2.** Possible worlds in a survey. Reply + means that the participant's family is expecting a child.

|       | age | city | reply |                |    |       |   |    |       |   |    |      |   |
|-------|-----|------|-------|----------------|----|-------|---|----|-------|---|----|------|---|
| a     | 23  | Bath | +     | age,city,reply | 23 | Bath  | - | 23 | Bath  | + | 23 | Bath | - |
|       | 23  | Bath | +     |                | 23 | Bath  | - | 23 | Bath  | + | 23 | Bath | - |
| c     | 23  | Bath | -     |                | 23 | Bath  | + | 23 | Bath  | + | 23 | Bath | - |
|       | 25  | York | -     |                | 25 | York  | + | 25 | York  | + | 25 | York | - |
| $w_1$ |     |      | $w_2$ |                |    | $w_3$ |   |    | $w_4$ |   |    |      |   |

**Figure 3.** Possible worlds in a survey. Reply + means that the participant's family is expecting a child.

not inform de dicto knowledge that she is pregnant:

$$w_1, a \not\models D_{\text{age,city,reply}} \text{ "is pregnant"}. \quad (5)$$

At the same time, in world  $w_1$ , any observer looking only at  $a$ 's anonymized data,  $(23, \text{Bath}, +)$ , will also not be able to infer *de re* that agent  $a$  is pregnant:

$$w_1, a \not\models R_{\text{age,city,reply}} \text{ "is pregnant"}. \quad (6)$$

This is because the personal data of agent  $a$  in world  $w_1$  is the same as the personal data of agent  $c$  in world  $w_2$ , where agent  $a$  is *not* pregnant. Thus, by relocating agent  $c$  to Bath we made statement (4) false.

Observe now that anyone who, in world  $w_1$ , has access to agent  $a$ 's personal data  $(23, \text{Bath}, +)$  and is also told that this data belongs to  $a$ , knows that she is pregnant:

$$w_1, a \models O_{\text{age,city,reply}} \text{ "is pregnant"}. \quad (7)$$

This is because agent  $a$  is pregnant in each world where her personal data is  $(23, \text{Bath}, +)$ .

In this article, we investigate the dependencies between *de re*, *de dicto*, and *overt* forms of knowledge. To understand the relation between them, it is also important to consider the notion of *traceability*. We say that a set of data variables  $X$  is traceable to agent  $a$  in world  $w$  if knowing the values of data variables for agent  $a$  in world  $w$  informs the knowledge of which agent the data belongs to. We denote this by  $w, a \models X$ . For example,  $w_1, b \models \{ \text{age, sex, reply} \}$  in the exam-

ple depicted in Fig. 1. Indeed, in that example, there is only one agent of age 23 who is male. At the same time,  $w_1, a \not\models \{ \text{age, sex, reply} \}$  for the same example, because data  $(23, F, +)$  could come from agent  $a$  in world  $w_1$  or, for instance, from agent  $c$  in world  $w_2$ .

## Syntax and semantics

We start by defining the class of models that will be used to give formal semantics of our logical system. Throughout the rest of the article, we assume a fixed set of data variables  $V$  and a fixed set of atomic propositions. Examples of data variables from the two previous sections are *age*, *sex*, *city*, and *reply*. Examples of atomic propositions are "is pregnant" and "lives in Bath." By a *dataset* we mean any subset of  $V$ .

Intuitively, we assume that data variables have certain values that depend on the world and the agent. In the example depicted in Fig. 3, in world  $w_1$  for agent  $a$ , the value of data variable *reply* is + and the value of data variable *age* is 23. As we will see in Definition 2, the actual values of data variables are not important for our semantics. The only important thing is if the data variable has the same or different values for any two given world-agent combinations. Hence, formally, it is more convenient to view a data variable as an equivalence relation on such pairs. We adopt this approach in the definition below.

### Definition 1.

A model is a tuple  $(W, \mathcal{A}, \{ \sim_x \}_{x \in V}, \pi)$ , where

- (1)  $W$  is a (possibly empty) set of worlds,
- (2)  $\mathcal{A}$  is a (possibly empty) set of agents,
- (3)  $\sim_x$  is an equivalence relation on  $W \times \mathcal{A}$  for each data variable  $x \in V$ ,
- (4)  $\pi(p) \subseteq W \times \mathcal{A}$  for each atomic proposition  $p$ .

We write  $(u_1, a_1) \sim_X (u_2, a_2)$  if  $(u_1, a_1) \sim_x (u_2, a_2)$  for each data variable  $x \in X$ . In particular, observe that the following statement is always true:  $(u_1, a_1) \sim_\emptyset (u_2, a_2)$ .

Note that, unlike the traditional approach in modal logic, valuation  $\pi(p)$  is a set of pairs rather than a set of possible worlds. For example, if “is pregnant” is one of the atomic propositions, then  $(u, a) \in \pi$  (“is pregnant”) means that agent  $a$  is pregnant in world  $u$ .

The language  $\Phi$  of our logical system is defined by the following grammar:

$$\varphi := p \mid !X \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathbf{O}_X\varphi \mid \mathbf{D}_X\varphi \mid \mathbf{R}_X\varphi,$$

where  $p$  is an atomic proposition and  $X \subseteq V$  is a dataset. We read  $!X$  as “dataset  $X$  is traceable,”  $\mathbf{O}_X\varphi$  as “dataset  $X$  informs *overt* knowledge of  $\varphi$ ,”  $\mathbf{D}_X\varphi$  as “dataset  $X$  informs *de dicto* knowledge of  $\varphi$ ,” and  $\mathbf{R}_X\varphi$  as “dataset  $X$  informs *de re* knowledge of  $\varphi$ .” We assume that implication  $\rightarrow$  and biconditional  $\leftrightarrow$  as well as the Boolean constants true  $\top$  and false  $\perp$  are defined through negation and disjunction in the usual way.

#### Definition 2.

For any model  $(W, \mathcal{A}, \{\sim_x\}_{x \in V}, \pi)$ , world  $w \in W$ , agent  $a \in \mathcal{A}$ , and formula  $\varphi \in \Phi$ , the satisfaction relation  $w, a \Vdash \varphi$  is defined as follows:

- (1)  $w, a \Vdash p$ , if  $(w, a) \in \pi(p)$ ,
- (2)  $w, a \Vdash !X$ , if  $a = b$  for each world  $u \in W$  and each agent  $b \in \mathcal{A}$  such that  $(w, a) \sim_X (u, b)$ ,
- (3)  $w, a \Vdash \neg\varphi$ , if  $w, a \not\Vdash \varphi$ ,
- (4)  $w, a \Vdash \varphi \vee \psi$ , if either  $w, a \Vdash \varphi$  or  $w, a \Vdash \psi$ ,
- (5)  $w, a \Vdash \mathbf{O}_X\varphi$ , if  $u, a \Vdash \varphi$  for each world  $u \in W$  such that  $(w, a) \sim_X (u, a)$ ,
- (6)  $w, a \Vdash \mathbf{D}_X\varphi$ , if  $u, b \Vdash \varphi$  for each world  $u \in W$  and each agent  $b \in \mathcal{A}$  such that  $(w, a) \sim_X (u, b)$ ,
- (7)  $w, a \Vdash \mathbf{R}_X\varphi$ , if  $u, a \Vdash \varphi$  for each world  $u \in W$  and each agent  $b \in \mathcal{A}$  such that  $(w, a) \sim_X (u, b)$ .

Recall that, informally, the relation  $(w, a) \sim_X (u, b)$  means that all variables in dataset  $X$  have the same values for agent  $a$  in world  $w$  as for agent  $b$  in world  $u$ . Thus, item 2 above captures the fact that in world  $w$  dataset  $X$  about agent  $a$  uniquely identifies the agent.

Item 5 of the above definition assumes that the identity of agent  $a$  is revealed, so it only considers worlds  $u$  in which dataset  $X$  for this agent have the same values as in world  $w$ .

Item 6 and item 7 do not assume that the identity is revealed. Thus, they consider all possible world-agent combinations  $(u, b)$  for which the value of dataset  $X$  is the same as for  $(w, a)$ .

## Undefinability results

In this section, we show that none of the three modalities is definable through the two others even if the traceability expression is used. We use the “truth sets algebra” technique for proving undefinability, which has been proposed in [27] and used in [24,28,29]. Unlike the more traditional “bisimulation” method, this technique uses a single model. The results in this section use the two definitions below:

#### Definition 3.

For any given model, the truth set  $\llbracket \varphi \rrbracket$  of a formula  $\varphi \in \Phi$  is the set  $\{(w, a) \mid w, a \Vdash \varphi\}$ .

#### Definition 4.

Formulae  $\varphi, \psi \in \Phi$  are semantically equivalent if  $\llbracket \varphi \rrbracket = \llbracket \psi \rrbracket$  in each model.

Without loss of generality, in this section, we suppose that language  $\Phi$  contains a single atomic proposition  $p$  and a single data variable  $x$ . Alternatively, in the models we construct below, the valuation of all atomic propositions could be defined the same way as for  $p$ . Similarly, equivalence relations for other data variables can be defined the same as for data variable  $x$ .

## Undefinability of R through !, O, and D

Consider a model that contains three worlds:  $w, u$ , and  $v$ , and two agents  $a$  and  $b$ . Suppose that equivalence relation  $\sim_x$  divides set  $\{w, u, v\} \times \{a, b\}$  into two equivalence classes:  $\{(w, a), (u, b), (u, b)\}$  and  $\{(u, a), (v, a), (v, b)\}$ . Informally, it means that data variable  $x$  has one value, say  $+$ , for each pair in the first class and another value, say  $-$ , for each pair in the second class. We depict these values in the left-most diagram in Fig. 4. Let  $\pi(p) = \{(w, a), (u, b), (u, b)\}$ .

We visualize the truth sets of various formulae as  $3 \times 2$  tables whose rows are indexed with worlds and whose columns are indexed with agents. A cell  $(w, a)$  in such a table is colored gray if the pair  $(w, a)$  belongs to the truth set. The four diagrams in the center of Fig. 4 visualize the truth sets  $\llbracket p \rrbracket$ ,  $\llbracket \perp \rrbracket$ ,  $\llbracket \neg p \rrbracket$ , and  $\llbracket \top \rrbracket$ .

The idea behind the truth set algebra technique is to show that any formula that does not use modality  $R$  has a truth set equal to one of the sets  $\llbracket p \rrbracket$ ,  $\llbracket \perp \rrbracket$ ,  $\llbracket \neg p \rrbracket$ , and  $\llbracket \top \rrbracket$ . At the same time, a formula that uses modality  $R$  can have a different truth set.

#### Lemma 1.

$\llbracket \mathbf{O}_\emptyset\varphi \rrbracket, \llbracket \mathbf{D}_\emptyset\varphi \rrbracket, \llbracket \mathbf{O}_x\varphi \rrbracket, \llbracket \mathbf{D}_x\varphi \rrbracket \in \{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$  for any formula  $\varphi \in \Phi$  such that  $\llbracket \varphi \rrbracket \in \{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$ .

#### Proof.

Suppose that  $\llbracket \varphi \rrbracket = \llbracket p \rrbracket$ . First, we show that  $\llbracket \mathbf{O}_\emptyset\varphi \rrbracket = \llbracket \perp \rrbracket = \emptyset$ . Indeed, assume the opposite. Thus, there is a world  $w' \in \{w, u, v\}$  and an agent  $a' \in \{a, b\}$  such that  $(w', a') \in \llbracket \mathbf{O}_\emptyset\varphi \rrbracket$ . Hence,  $w', a' \Vdash \mathbf{O}_\emptyset\varphi$  by Definition 3. Then,  $v, a' \Vdash \varphi$  by item 5 of Definition 2 because  $(w', a') \sim_\emptyset (v, a')$ . Thus,  $(v, a') \in \llbracket \varphi \rrbracket$  by Definition 3. Hence, by the assumption  $\llbracket \varphi \rrbracket = \llbracket p \rrbracket$ , we have  $(v, a') \in \llbracket p \rrbracket$ . Observe in Fig. 4 that the truth set  $\llbracket p \rrbracket$  does not contain any single pair  $(x, y)$ , where  $x = v$ . This is a contradiction. Therefore,  $\llbracket \mathbf{O}_\emptyset\varphi \rrbracket = \llbracket \perp \rrbracket$ .

The proof that  $\llbracket \mathbf{D}_\emptyset\varphi \rrbracket = \llbracket \perp \rrbracket$  is similar, but it uses item 6 of Definition 2 instead of item 5. In Fig. 4, we visualize both results by a directed edge from diagram  $\llbracket p \rrbracket$  to diagram  $\llbracket \perp \rrbracket$  labeled with  $\mathbf{O}_\emptyset, \mathbf{D}_\emptyset$ .

Next, we show that  $\llbracket \mathbf{O}_x\varphi \rrbracket = \llbracket p \rrbracket$ .

( $\subseteq$ ): Let us consider any world  $w' \in \{w, u, v\}$  and any agent  $a' \in \{a, b\}$  such that  $(w', a') \in \llbracket \mathbf{O}_x\varphi \rrbracket$ . Thus,  $w', a' \Vdash \mathbf{O}_x\varphi$  by Definition 3. Note that  $(w', a') \sim_x (w', a')$ . Hence,  $w', a' \Vdash \varphi$  by item 5 of Definition 2. Then,  $(w', a') \in \llbracket \varphi \rrbracket$  again by Definition 3. Therefore,  $(w', a') \in \llbracket p \rrbracket$  by the assumption  $\llbracket \varphi \rrbracket = \llbracket p \rrbracket$ .

( $\supseteq$ ): Consider any  $w' \in \{w, u, v\}$  and any agent  $a' \in \{a, b\}$  such that

$$(w', a') \in \llbracket p \rrbracket. \quad (8)$$

By Definition 3, it suffices to show that  $w', a' \Vdash \mathbf{O}_x\varphi$ . Toward this proof, consider any  $w'' \in \{w, u, v\}$  such that  $(w', a') \sim_x (w'', a')$ . By item 5 of Definition 2, it suffices to establish that  $w'', a' \Vdash \varphi$ .

Observe that the assumption  $(w', a') \sim_x (w'', a')$  implies that  $(w', a') \in \llbracket p \rrbracket$  iff  $(w'', a') \in \llbracket p \rrbracket$ , see Fig. 4. Then,  $(w'', a') \in \llbracket p \rrbracket$  by statement (8). Thus, by the assumption  $\llbracket \varphi \rrbracket = \llbracket p \rrbracket$  it follows that  $(w'', a') \in \llbracket \varphi \rrbracket$ . Therefore,  $w'', a' \Vdash \varphi$  by Definition 3.

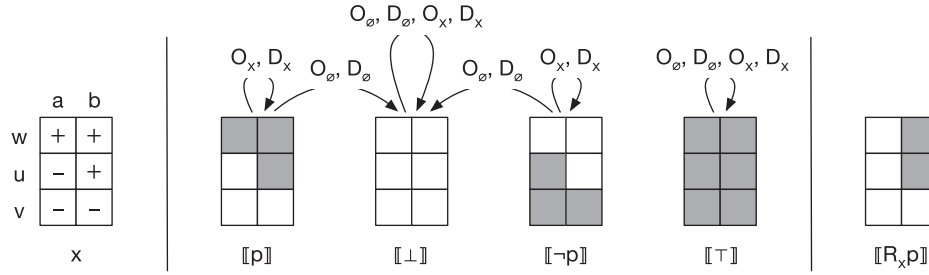


Figure 4. Toward the proof of Theorem 1.

The proof that  $\llbracket D_x \varphi \rrbracket = \llbracket p \rrbracket$  is similar, but it uses item 6 of Definition 2 instead of item 5. In Fig. 4, we visualize both results by a directed loop edge from diagram  $\llbracket p \rrbracket$  back to the same diagram  $\llbracket p \rrbracket$  labeled with  $O_x, D_x$ .

The proofs in the remaining three cases,  $\llbracket \varphi \rrbracket = \llbracket \perp \rrbracket$ ,  $\llbracket \varphi \rrbracket = \llbracket \neg p \rrbracket$ , and  $\llbracket \varphi \rrbracket = \llbracket \top \rrbracket$ , are similar. We show the corresponding directed edges in Fig. 4.  $\square$

#### Lemma 2.

$\llbracket \varphi \rrbracket \in \{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$  for any formula  $\varphi \in \Phi$  that does not contain modality  $R$ .

#### Proof.

We prove the statement of the lemma by induction on the structural complexity of formula  $\varphi$ .

If  $\varphi$  is an atomic proposition  $p$ , then  $\llbracket \varphi \rrbracket \in \{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$  because truth set  $\llbracket p \rrbracket$  is an element of the set  $\{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$ .

Suppose that formula  $\varphi$  has the form  $\neg X$ , where dataset  $X$  is either  $\emptyset$  or  $\{x\}$ . Observe from the left-most diagram in Fig. 4 that for any  $w' \in \{w, u, v\}$  and any  $a' \in \{a, b\}$  there is  $w'' \in \{w, u, v\}$  and an agent  $a'' \in \{a, b\}$  such that  $(w', a') \sim_X (w'', a'')$  and  $a' \neq a''$ . Thus, by item 2 of Definition 2, it follows that  $w', a' \not\models X$  for each world  $w' \in \{w, u, v\}$  and each agent  $a' \in \{a, b\}$ . Hence, truth set  $\llbracket \neg X \rrbracket$  is empty by Definition 2. Then,  $\llbracket \neg X \rrbracket = \llbracket \perp \rrbracket$ , see Fig. 4. Therefore,  $\llbracket \neg X \rrbracket \in \{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$ .

Suppose that formula  $\varphi$  has the form  $\neg \psi$ . Thus, the truth set  $\llbracket \varphi \rrbracket$  is the complement of the truth set  $\llbracket \psi \rrbracket$  by Definition 3 and item 3 of Definition 2. By the induction hypothesis,  $\llbracket \psi \rrbracket \in \{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$ . Observe in Fig. 4 that the complement of each truth sets in the family  $\{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$  belongs to the same family. Therefore,  $\llbracket \varphi \rrbracket \in \{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$ .

Suppose that formula  $\varphi$  has the form  $\psi_1 \vee \psi_2$ . Thus, the truth set  $\llbracket \varphi \rrbracket$  is the union of the truth sets  $\llbracket \psi_1 \rrbracket$  and  $\llbracket \psi_2 \rrbracket$  by Definition 3 and item 4 of Definition 2. Observe in Fig. 4 that the union of any two truth sets in the family  $\{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$  belongs to the same family. For example,  $\llbracket p \rrbracket \cup \llbracket \neg p \rrbracket = \llbracket \top \rrbracket$ . Note that, by the induction hypothesis,  $\llbracket \psi_1 \rrbracket, \llbracket \psi_2 \rrbracket \in \{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$ . Therefore,  $\llbracket \varphi \rrbracket \in \{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$ .

If formula  $\varphi$  has either the form  $O_x \psi$  or the form  $D_x \psi$ , then the statement of the lemma follows from the induction hypothesis and Lemma 1.  $\square$

Next, let us make an auxiliary observation based on the left-most diagram in Fig. 4.

#### Lemma 3.

For any world  $w' \in \{w, u, v\}$  and any agent  $a' \in \{a, b\}$ , if data variable  $x$  has value  $+$  at  $(w', a')$ , then data variable  $x$  has value  $+$  at  $(w', b)$ .

#### Lemma 4.

$\llbracket R_x p \rrbracket \notin \{\llbracket p \rrbracket, \llbracket \perp \rrbracket, \llbracket \neg p \rrbracket, \llbracket \top \rrbracket\}$ .

#### Proof.

It suffices to prove that the truth set  $\llbracket R_x p \rrbracket$  is the one depicted by the right-most diagram in Fig. 4. In other words it is enough to show that for any cell  $(w', a')$  of that diagram,  $w', a' \models R_x p$  iff cell  $(w', a')$  is gray.

( $\Rightarrow$ ): Suppose that the cell  $(w', a')$  is white at the right-most diagram in Fig. 4. We consider the following two cases separately:

Case 1:  $(w', a') = (w, a)$ . Note that  $(u, a) \notin \pi(p)$  by the choice of valuation function  $\pi$ . Hence,  $(u, a) \not\models p$  by item 1 of Definition 2. Additionally,  $(w, a) \sim_x (u, b)$ , see the left-most diagram in Fig. 4. Thus,  $(w, a) \not\models R_x p$  by item 7 of Definition 2. Therefore,  $(w', a') \not\models R_x p$  by the assumption  $(w', a') = (w, a)$  of the case.

Case 2:  $(w', a') \neq (w, a)$ . Thus, the assumption that the cell  $(w', a')$  is white at the right-most diagram in Fig. 4 implies that the same cell  $(w', a')$  is white at the second-from-left diagram in Fig. 4. Hence,  $(w', a') \notin \llbracket p \rrbracket$ . Then,  $(w', a') \not\models p$  by Definition 3. Therefore, it follows that  $(w', a') \not\models R_x p$  by item 7 of Definition 2 and because  $(w', a') \sim_x (w', a')$ .

( $\Leftarrow$ ): Suppose that the cell  $(w', a')$  is gray at the right-most diagram in Fig. 4. Thus  $a' = b$  and the value of data variable  $x$  is  $+$ , see Fig. 4. Consider any world  $w'' \in \{w, u, v\}$  and any agent  $a'' \in \{a, b\}$  such that  $(w', b) \sim_x (w'', a'')$ . By item 7 of Definition 2, it suffices to show that  $(w'', b) \models p$ .

Indeed, the assumption  $(w', b) \sim_x (w'', a'')$  means that the value of data variable  $x$  at cell  $(w'', a'')$  is the same as its value at cell  $(w', b)$ , which is  $+$ . Hence, the value of data variable  $x$  at cell  $(w'', b)$  is also  $+$  by Lemma 3. Hence,  $(w'', b) \in \llbracket p \rrbracket$ , see Fig. 4. Therefore,  $(w', b) \models p$  by Definition 3.  $\square$

The next result follows from Lemma 2 and Lemma 4.

#### Theorem 1 (undefinability).

Formula  $R_x p$  is not semantically equivalent to any formula in language  $\Phi$  that does not contain modality  $R$ .

#### Undefinability of $O$ and $D$

The proof of the next theorem is similar to the proof of Theorem 1 except that instead of Fig. 4 it uses Fig. 5.

#### Theorem 2 (undefinability).

Formula  $O_x p$  is not semantically equivalent to any formula in language  $\Phi$  that does not contain modality  $O$ .

The proof of the next theorem is also similar to the proofs of the two previous theorems except that it uses Fig. 6. Note that, in this case, data variable  $x$  has *three* possible values that we denote by  $-$ ,  $1$ , and  $2$ , see Fig. 6.

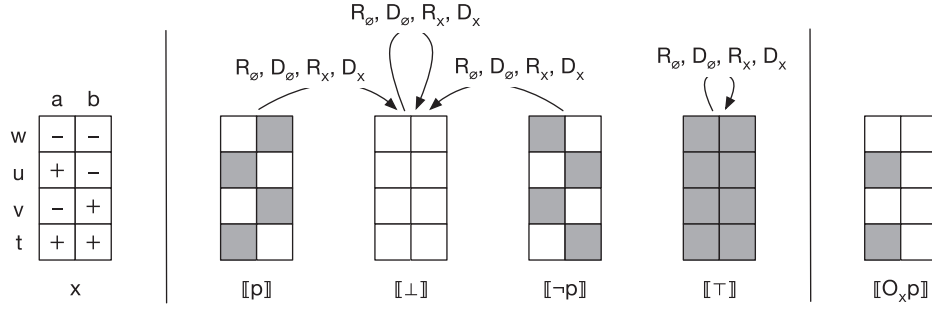


Figure 5. Toward the proof of Theorem 2.

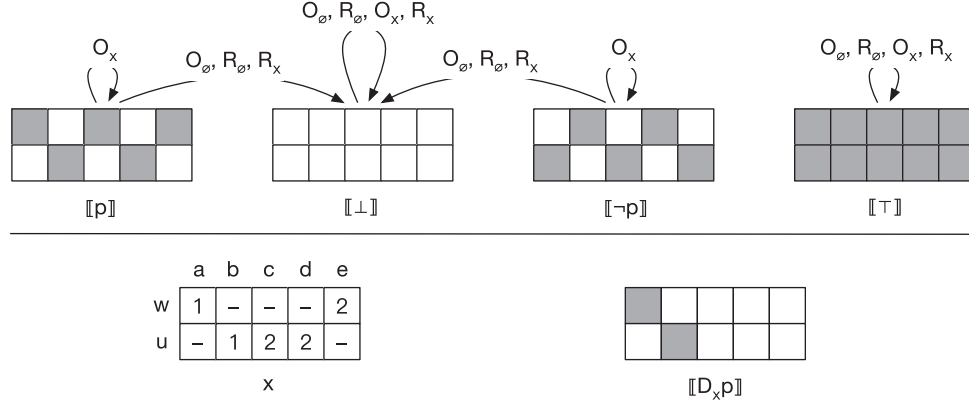


Figure 6. Toward the proof of Theorem 3.

**Theorem 3 (undefinability).**

Formula  $D_x p$  is not semantically equivalent to any formula in language  $\Phi$  that does not contain modality  $D$ .

**Axioms**

In addition to propositional tautologies in language  $\Phi$ , our logical system contains the following axioms:

- (1) Truth:  $O_X \varphi \rightarrow \varphi$ ,
- (2) Introspection of De Re Knowledge:  $R_X \varphi \rightarrow O_X R_X \varphi$ ,
- (3) Negative Introspection:  
 $\neg \Box_X \varphi \rightarrow \Box_X \neg \Box_X \varphi$ , where  $\Box \in \{O, D\}$ ,
- (4) Distributivity:  
 $\Box_X (\varphi \rightarrow \psi) \rightarrow (\Box_X \varphi \rightarrow \Box_X \psi)$ , where  $\Box \in \{O, D, R\}$ ,
- (5) Monotonicity:  
 $\Box_X \varphi \rightarrow \Box_Y \varphi$  and  $!X \rightarrow !Y$ , where  $X \subseteq Y$  and  $\Box \in \{O, D, R\}$ ,
- (6) Overt Knowledge:  $\Box_X \varphi \rightarrow O_X \varphi$ , where  $\Box \in \{D, R\}$ ,
- (7) Introspection of Traceability:  $!X \rightarrow D_X !X$ ,
- (8) Data-Free Knowledge:  $\Box_\emptyset \varphi \rightarrow R_\emptyset \varphi$ , where  $\Box \in \{O, D\}$ ,
- (9) Traceable Data:  $!X \rightarrow (O_X \varphi \rightarrow \Box_X \varphi)$ , where  $\Box \in \{D, R\}$ .

The Truth, the Negative Introspection, and the Distributivity axioms are the standard axioms of the epistemic logic. The Truth axiom is valid for all three modalities,  $O$ ,  $D$ , and  $R$ . We list it only for modality  $O$  because this principle for the other two modalities can be derived from using the Overt Knowledge axiom. At the same time, the Negative Introspection axiom is not universally valid for modality  $R$ . As usual in epistemic logic, the positive introspection principle is provable from the Negative Introspection and some other axioms. We show this in Lemma 6. The positive introspection principle is also

not universally valid for modality  $R$ . However, a weaker form of it, expressed by the Introspection of De Re Knowledge axiom, holds.

The Introspection of Traceability axiom states that if the data uniquely identifies the agent, then the data informs the de dicto knowledge of this. We state this axiom for modality  $D$ , the same property for modality  $O$  and  $R$  is derivable.

The Data-Free Knowledge axiom, when combined with the Overt Knowledge axiom, implies that the formulae  $O_\emptyset \varphi$  and  $R_\emptyset \varphi$  are equivalent. Note that the formula  $D_\emptyset \varphi$  is not equivalent to the two formulae above. Indeed,  $w, a \Vdash O_\emptyset \varphi$  means that  $u, a \Vdash \varphi$  for each world  $u$ , see Definition 2. At the same time,  $w, a \Vdash D_\emptyset \varphi$  means that  $u, b \Vdash \varphi$  is true for each world  $u$  and each agent  $b$ .

Finally, the Traceable Data axiom, when combined with the Overt Knowledge axiom, implies that there is no difference between the three forms of knowledge if the data is traceable.

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

$$\frac{\varphi}{D_X \varphi} \quad \frac{\varphi, \varphi \rightarrow \psi}{\psi}.$$

We omit the Necessitation inference rules for modalities  $O$  and  $R$  because these rules are derivable in our system. In addition to unary relation  $\vdash \varphi$ , we also consider binary relation  $\Gamma \vdash \varphi$  between a (possibly infinite) set of formulae  $\Gamma \subseteq \Phi$  and a formula  $\varphi \in \Phi$ . We say that  $\Gamma \vdash \varphi$  is true if formula  $\varphi$  is derivable from the *theorems* of our logical system and an additional set of assumptions  $\Gamma$  using the Modus Ponens inference rule only. Note that statements  $\emptyset \vdash \varphi$  and  $\vdash \varphi$  are equivalent. We say that a set of formulae  $\Gamma \subseteq \Phi$  is consistent if  $\Gamma \not\vdash \perp$ .

We conclude this section with a few technical observations that will be used in the proof of the completeness.

**Lemma 5.**

$\vdash \Box_X \varphi \rightarrow \varphi$ , where  $\Box \in \{D, R\}$ .

**Proof.**

Note that  $\vdash \Box_X \varphi \rightarrow O_X \varphi$  by the Overt Knowledge axiom. At the same time,  $\vdash O_X \varphi \rightarrow \varphi$  by the Truth axiom. Therefore,  $\vdash \Box_X \varphi \rightarrow \varphi$  by propositional reasoning.  $\square$

The next lemma states a well-known observation that the positive introspection principle is derivable from S5 axioms. To keep the article self-contained, we give its proof in the [appendix](#).

**Lemma 6.**

$\vdash \Box_X \varphi \rightarrow \Box_X \Box_X \varphi$ , where  $\Box \in \{O, D\}$ .

**Lemma 7.**

$\vdash \Box_X \varphi \leftrightarrow O_X \Box_X \varphi$ , where  $\Box \in \{D, R\}$ .

**Proof.**

Note that the formula  $O_X \Box_X \varphi \rightarrow \Box_X \varphi$  is an instance of the Truth axiom. Thus, it suffices to show that  $\vdash \Box_X \varphi \rightarrow O_X \Box_X \varphi$ . If  $\Box$  is modality R, then the last formula is an instance of the Introspection of De Re Knowledge axiom.

Finally, let us prove  $\vdash D_X \varphi \rightarrow O_X D_X \varphi$ . It follows from Lemma 6 that  $\vdash D_X \varphi \rightarrow D_X D_X \varphi$ . At the same time, the following formula is an instance of the Overt Knowledge axiom:  $D_X D_X \varphi \rightarrow O_X D_X \varphi$ . Therefore,  $\vdash D_X \varphi \rightarrow O_X D_X \varphi$  by propositional reasoning.  $\square$

The proof of the next standard lemma is also in the [appendix](#).

**Lemma 8.**

If  $\varphi_1, \dots, \varphi_n \vdash \psi$ , then  $\Box_X \varphi_1, \dots, \Box_X \varphi_n \vdash \Box_X \psi$ , for any modality  $\Box \in \{O, D, R\}$ .

**Lemma 9 (Lindenbaum).**

Any consistent set of formulae can be extended to a maximal consistent set of formulae.

**Proof.**

The standard proof of Lindenbaum's lemma [30, Proposition 2.14] applies here.

**Soundness****Theorem 4 (soundness).**

If  $\vdash \varphi$ , then  $w, a \Vdash \varphi$  for any world  $w$  and any agent  $a$  of any model  $(W, \mathcal{A}, \{\sim_x\}_{x \in V}, \pi)$ .

**Proof.**

The soundness of the Truth, the Negative Introspection, the Distributivity, the Monotonicity, and Overt Knowledge axioms as well as of the Necessitation and the Modus Ponens inference rules is straightforward. Below, we prove the soundness of each of the remaining axioms as a separate claim.

**Claim 1.**

If  $w, a \Vdash R_X \varphi$ , then  $w, a \Vdash O_X R_X \varphi$ .

**Proof of Claim.**

Let us consider any world  $u \in W$  such that  $(w, a) \sim_X (u, a)$ . By item 5 of Definition 2, it suffices to show that  $u, a \Vdash R_X \varphi$ . Toward this proof, consider any world  $v \in W$  and any agent  $b \in \mathcal{A}$  such that  $(u, a) \sim_X (v, b)$ . By item 7 of Definition 2, it suffices to prove that  $v, a \Vdash \varphi$ .

Note that  $(w, a) \sim_X (v, b)$  by the assumptions  $(w, a) \sim_X (u, a)$  and  $(u, a) \sim_X (v, b)$ . Thus,  $v, a \Vdash \varphi$  by the assumption  $w, a \Vdash R_X \varphi$  of the claim and item 7 of Definition 2.  $\square$

**Claim 2.**

If  $w, a \Vdash !X$ , then  $w, a \Vdash D_X !X$ .

**Proof of Claim.**

Consider any world  $u \in W$  and any agent  $b \in \mathcal{A}$  such that

$$(w, a) \sim_X (u, b). \quad (9)$$

By item 6 of Definition 2, it suffices to prove that  $u, b \Vdash !X$ . Toward this proof, consider any world  $v \in W$  and any agent  $c \in \mathcal{A}$  such that  $(u, b) \sim_X (v, c)$ . By item 2 of Definition 2, it suffices to show that  $b = c$ .

Note that statement (9) and the assumption  $(u, b) \sim_X (v, c)$  imply that  $(w, a) \sim_X (v, c)$ . Thus,  $a = c$  by the assumption  $w, a \Vdash !X$  of the claim and item 2 of Definition 2. Similarly,  $a = b$  by statement (9). Therefore,  $b = c$ .  $\square$

**Claim 3.**

If  $w, a \Vdash \Box_\emptyset \varphi$ , then  $w, a \Vdash R_\emptyset \varphi$ , where  $\Box \in \{O, D\}$ .

**Proof of Claim.**

Note that the statement  $(w, a) \sim_\emptyset (u, b)$  is true for any world  $u \in W$  and any agent  $b \in \mathcal{A}$ . Thus, by item 5 (or item 6) of Definition 2 the assumption  $w, a \Vdash \Box_\emptyset \varphi$  implies that  $u, a \Vdash \varphi$  for any world  $u \in W$ . Hence,  $w, a \Vdash R_\emptyset \varphi$  by item 7 of Definition 2.  $\square$

**Claim 4.**

If  $w, a \Vdash !X$  and  $w, a \Vdash O_X \varphi$ , then  $w, a \Vdash \Box_X \varphi$ , where  $\Box \in \{D, R\}$ .

**Proof of Claim.**

First, assume that  $\Box = D$ . Consider any world  $u \in W$  and any agent  $b \in \mathcal{A}$  such that  $(w, a) \sim_X (u, b)$ . By item 6 of Definition 2, it suffices to show that  $u, b \Vdash \varphi$ .

Note that the assumption  $w, a \Vdash !X$  of the claim implies that  $a = b$  by the assumption  $(w, a) \sim_X (u, b)$  and item 2 of Definition 2. Thus,  $(w, a) \sim_X (u, a)$  again by the assumption  $(w, a) \sim_X (u, b)$ . Hence,  $u, a \Vdash \varphi$  by the assumption  $w, a \Vdash O_X \varphi$  of the claim and item 5 of Definition 2. Therefore,  $u, b \Vdash \varphi$  because  $a = b$ .  $\square$

The proof in the case  $\Box = R$  is similar.  $\square$

This concludes the proof of the theorem.  $\square$

**Completeness**

As usual, the proof of completeness for our logical system consists of a canonical model construction and a proof of a “truth” lemma that connects the satisfaction relation of the canonical model with the provability in our logical system. Traditionally, the worlds of the canonical model are defined as the maximal consistent sets of formulae. This approach works for modal logic with a few exceptions. One such exception is the epistemic logic of distributed knowledge. To prove the completeness of this system, Fagin, Halpern, and Vardi use a tree construction in which nodes of the tree are labeled with maximal consistent sets [31]. Note that although our logical system does not have distributed knowledge *explicitly*, it has it implicitly. Indeed, for each agent  $a$ , one can consider a variable  $x_a$  that represents all information known to agent  $a$ . In such a setting, knowledge informed by a dataset  $\{x_a\}_{a \in G}$  is the distributed knowledge of group  $G$ .

The maximal consistent set construction also needs to be adjusted for the egocentric setting when the semantics is defined in terms of a world-agent pair positioned on the left side of the satisfaction relation  $\Vdash$ . In such a setting, a maximal consistent set captures the properties of a world-agent pair rather than just of a possible world. As a result, it is not, generally speaking, possible to define possible worlds simply as maximal consistent sets of formulae. One of the approaches to deal with this is suggested by Naumov and Tao [20]. They use a “matrix” technique in which the canonical model is recursively constructed as an infinite matrix whose cells are maximal consistent sets of formulae. The rows and columns of such a matrix correspond to

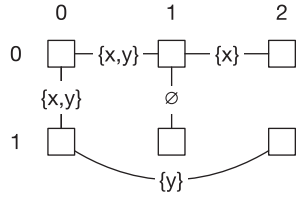


Figure 7. A Cartesian tree.

possible worlds and agents, respectively. The maximal consistent set  $S_{wa}$  in the  $w$ -th row and  $a$ -th column captures the properties that agent  $a$  has in world  $w$ .

In this article, we combine tree and matrix techniques to prove the completeness of our logical system. At the center of our construction is a tree structure on the cells of a matrix. Because the position of each cell in a matrix could be described by an element of the Cartesian product of the set of rows and the set of columns, we call such structure a *Cartesian tree*.

### Cartesian trees

Figure 7 depicts an example of a Cartesian tree. The nodes of this tree are the cells in  $2 \times 3$  matrix. As shown in the figure, we index rows and columns starting from 0. The nodes and the undirected edges between them form a tree (graph without cycles). As shown in the figure, edges in a Cartesian tree are labeled with datasets. The Cartesian tree depicted in Fig. 7 has a finite number of rows and columns. Our proof of completeness requires us to consider infinite matrices whose rows and columns can be indexed by the whole set of integer numbers (ordinal  $\omega$ ). To achieve this, we define cells of the matrix as elements of the Cartesian product  $\alpha \times \beta$ , where  $\alpha \leq \omega$  and  $\beta \leq \omega$  are two ordinals. For the example shown in Fig. 7, we have  $\alpha = 2 = \{0, 1\}$  and  $\beta = 3 = \{0, 1, 2\}$ . The formal definition of a Cartesian tree is below.

#### Definition 5.

A Cartesian tree is a tuple  $(\alpha, \beta, E, \ell)$ , where

- (1)  $\alpha, \beta \leq \omega$  are two ordinals,
- (2)  $E \subseteq (\alpha \times \beta)^2$  is an symmetric *adjacency* relation on  $\alpha \times \beta$  that forms an (undirected) tree structure,
- (3)  $\ell$  is a labeling function that to each edge  $e \in E$  assigns a dataset  $\ell(e) \subseteq V$ .

By  $\langle n_1, n_2 \rangle$  we denote the edge in the Cartesian tree between nodes  $n_1, n_2 \in \alpha \times \beta$ . We use corner brackets to differentiate *unordered* pairs from ordered pairs that we continue to denote by parentheses. We say that an edge  $\langle n_1, n_2 \rangle \in E$  is labeled with a variable  $x \in V$  if  $x \in \ell(\langle n_1, n_2 \rangle)$ . By a simple path, we mean a path in the tree without repeating nodes. We allow trivial paths that start and end at the same node. Recall that for any two nodes in a tree, there is a unique simple path between these two nodes.

#### Definition 6.

For any nodes  $n_1, n_2 \in \alpha \times \beta$  and any  $X \subseteq V$ , let  $n_1 \overset{X}{\sim} n_2$  if all edges along the simple path connecting nodes  $n_1$  and  $n_2$  are labeled with each data variable in set  $X$ .

As an example,  $(1, 0) \overset{\{x,y\}}{\sim} (0, 1)$  for the Cartesian tree depicted in Fig. 7. Also,  $(1, 0) \overset{\{x\}}{\sim} (0, 2)$  and  $(1, 1) \overset{\{x,y\}}{\sim} (1, 1)$ . In fact,  $n \overset{X}{\sim} n$  is true for each dataset  $X \subseteq V$  and each node  $n$  in any Cartesian tree.

### Complete loaded Cartesian trees

Cartesian trees provide a framework on which we construct the canonical model by “loading” a maximal consistent set of formulae

at each node of the tree. We expect there to be some correlation between maximal consistent sets and the tree structure. The formal definition of a loaded Cartesian tree is below.

#### Definition 7.

A loaded Cartesian tree is a tuple  $(\alpha, \beta, E, \ell, S)$  such that

- (1)  $(\alpha, \beta, E, \ell)$  is a Cartesian tree,
- (2)  $S$  is a node labeling function that maps each node  $(w, a)$  in set  $\alpha \times \beta$  into a maximal consistent set of formulae  $S(w, a)$ , which we denote by  $S_{wa}$ , such that
  - (a) if  $\neg X \in S_{wa}$  and  $(w, a) \overset{X}{\sim} (u, b)$ , then  $a = b$ ,
  - (b) if  $(w, a) \overset{X}{\sim} (u, a)$ , then  $\mathcal{O}_X \varphi \in S_{wa}$  iff  $\mathcal{O}_X \varphi \in S_{ua}$ ,
  - (c) if  $(w, a) \overset{X}{\sim} (u, b)$ , then  $\mathcal{D}_X \varphi \in S_{wa}$  iff  $\mathcal{D}_X \varphi \in S_{ub}$ ,
  - (d) if  $\mathcal{R}_X \varphi \in S_{wa}$  and  $(w, a) \overset{X}{\sim} (u, b)$ , then  $\varphi \in S_{ua}$ .

Informally, we say that one loaded Cartesian tree is an extension of another if the extended tree adds new nodes while preserving the edges, labels, and loads in the original tree. The formal definition is given below.

#### Definition 8.

A loaded Cartesian tree  $(\alpha', \beta', E', \ell', S')$  is an *extension* of a loaded Cartesian tree  $(\alpha, \beta, E, \ell, S)$  if

- (1)  $\alpha \leq \alpha'$ ,
- (2)  $\beta \leq \beta'$ ,
- (3)  $E = E' \cap (\alpha \times \beta)^2$ ,
- (4)  $\ell'(e) = \ell(e)$  for each edge  $e \in E$ ,
- (5)  $S'_{wa} = S_{wa}$  for each node  $(w, a) \in \alpha \times \beta$ .

In order to be able to convert a loaded Cartesian tree into a canonical model, the tree must be “complete.” Below, we give the formal definition of what “complete” means. In the next subsection, we prove that any finite loaded Cartesian tree can be extended to a (possibly infinite) complete loaded Cartesian tree. This result is stated at the end of the next subsection as Theorem 5. The non-trivial proof of this theorem constitutes the bulk of the proof of the completeness theorem.

#### Definition 9.

A loaded Cartesian tree  $(\alpha, \beta, E, \ell, S)$  is complete when for each ordinal  $u < \alpha$ , ordinal  $b < \beta$ , formula  $\varphi \in \Phi$ , and dataset  $X \subseteq V$ ,

- (1) if  $\neg X \notin S_{ub}$ , then there is a node  $(w, a) \in \alpha \times \beta$  such that  $(u, b) \overset{X}{\sim} (w, a)$  and  $b \neq a$ ,
- (2) if  $\mathcal{O}_X \varphi \notin S_{ub}$ , then there is a node  $(w, b) \in \alpha \times \beta$  such that  $(u, b) \overset{X}{\sim} (w, b)$  and  $\varphi \notin S_{wb}$ ,
- (3) if  $\mathcal{D}_X \varphi \notin S_{ub}$ , then there is a node  $(w, a) \in \alpha \times \beta$  such that  $(u, b) \overset{X}{\sim} (w, a)$  and  $\varphi \notin S_{wa}$ ,
- (4) if  $\mathcal{R}_X \varphi \notin S_{ub}$ , then there is a node  $(w, a) \in \alpha \times \beta$  such that  $(u, b) \overset{X}{\sim} (w, a)$  and  $\varphi \notin S_{wb}$ .

### Extension theorem

A loaded Cartesian tree  $(\alpha, \beta, E, \ell, S)$  is finite if  $\alpha, \beta < \omega$ . As mentioned above, in this subsection we show that each finite loaded Cartesian tree can be extended to a complete loaded Cartesian tree. To be complete, the tree must satisfy conditions 1–4 of Definition 9 for each dataset  $X$ , each formula  $\varphi$ , each row  $w$ , and each column  $a$ . Instead of constructing the complete tree at once, we define it as a limit of an infinite chain of extensions. Each of these extensions makes one (or more) of conditions 1–4 satisfied for some specific values of  $X$ ,  $\varphi$ ,  $w$ , and  $a$ . The existence of these single-step extensions is shown in the next two lemmas.

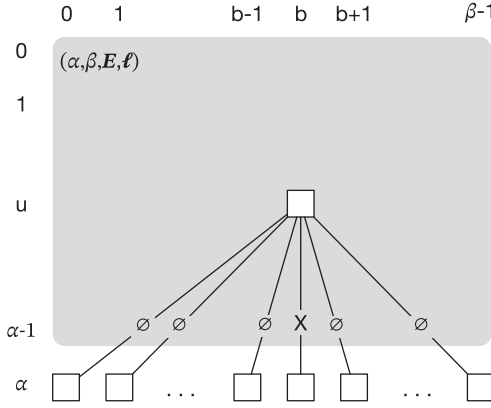


Figure 8. Cartesian tree  $(\alpha + 1, \beta, E', \ell')$ .

**Lemma 10.**

For any finite loaded Cartesian tree  $T = (\alpha, \beta, E, \ell, S)$ , any node  $(u, b) \in \alpha \times \beta$ , and any formula  $O_X \varphi \notin S_{ub}$ , there is an extension  $T' = (\alpha + 1, \beta, E', \ell', S')$  of tree  $T$  such that  $(u, b) \overset{X}{\circ} (\alpha, b)$  in tree  $T'$  and  $\varphi \notin S'_{\alpha b}$ .

**Proof.**

To define the loaded Cartesian tree  $T' = (\alpha + 1, \beta, E', \ell', S')$ , we first construct Cartesian tree  $(\alpha + 1, \beta, E', \ell')$  and then specify the load  $S'$ . The tree  $(\alpha + 1, \beta, E', \ell')$  is obtained from tree  $(\alpha, \beta, E, \ell)$  by an addition of a new row  $\alpha$ . Each node in this new row is connected by an edge to the cell  $(u, b)$ . All of the newly added edges are labeled with the empty set except the edge between nodes  $(u, b)$  and  $(\alpha, b)$ , which is labeled with set  $X$ , see Fig. 8. Formally, relation  $E'$  on the set  $(\alpha + 1) \times \beta$  is defined as:

$$E' = E \cup \{ \langle (u, b), (\alpha, a) \rangle \mid a < \beta \}. \quad (10)$$

and the labeling function  $\ell'$  defined as:

$$\ell'(e) = \begin{cases} X, & \text{if } e = \langle (u, b), (\alpha, b) \rangle, \\ \emptyset, & \text{if } e = \langle (u, b), (\alpha, a) \rangle, \text{ where } a \neq b, \\ \ell(e), & \text{otherwise.} \end{cases} \quad (11)$$

Note that  $(\alpha + 1, \beta, E', \ell')$  is a Cartesian tree by Definition 5.

Next, we proceed to define function  $S'$ . First, let us consider the set of formulae

$$Q^- = \{ \neg \varphi \} \cup \{ \psi \mid O_X \psi \in S_{ub} \}. \quad (12)$$

**Claim 5.**

$Q^-$  is a consistent set of formulae.

**Proof of Claim.**

Suppose the opposite. Then, there are formulae

$$O_X \psi_1, \dots, O_X \psi_n \in S_{ub} \quad (13)$$

such that  $\psi_1, \dots, \psi_n \vdash \varphi$ . Hence,  $O_X \psi_1, \dots, O_X \psi_n \vdash O_X \varphi$ , by Lemma 8. Thus,  $S_{ub} \vdash O_X \varphi$  by assumption (13). Then,  $O_X \varphi \in S_{ub}$  because  $S_{ub}$  is a maximal consistent set, which contradicts the assumption  $O_X \varphi \notin S_{ub}$  of the lemma.  $\square$

By Lemma 9, set  $Q^-$  can be extended to a maximal consistent set  $Q$ .

To finish the construction of the loaded Cartesian tree, we need to define the labeling function  $S'$  that maps pairs from Cartesian product  $(\alpha + 1) \times \beta$  into maximal consistent sets. The values of this function  $S'_{wa}$  are given by the matrix of size  $\alpha + 1$  by  $\beta$  shown in below

equation:

$$(S'_{wa}) = \begin{bmatrix} S_{00} & S_{01} & \dots & S_{0,b-1} & S_{0b} & S_{0,b+1} & \dots & S_{0,\beta-1} \\ S_{10} & S_{11} & \dots & S_{1,b-1} & S_{1b} & S_{1,b+1} & \dots & S_{1,\beta-1} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ S_{u0} & S_{u1} & \dots & S_{u,b-1} & S_{ub} & S_{u,b+1} & \dots & S_{u,\beta-1} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ S_{\alpha-1,0} & S_{\alpha-1,1} & \dots & S_{\alpha-1,b-1} & S_{\alpha-1,b} & S_{\alpha-1,b+1} & \dots & S_{\alpha-1,\beta-1} \\ S_{u0} & S_{u1} & \dots & S_{u,b-1} & Q & S_{u,b+1} & \dots & S_{u,\beta-1} \end{bmatrix}.$$

Matrix  $(S'_{wa})$  toward the proof of Lemma 10.

Although the visual definition of  $S'_{wa}$  through a matrix is easier to understand intuitively, in proofs, it is more convenient to refer to an equivalent definition in a more traditional form:

$$S'_{wa} = \begin{cases} S_{ua}, & \text{if } w = \alpha \text{ and } a \neq b, \\ Q, & \text{if } w = \alpha \text{ and } a = b, \\ S_{wa}, & \text{otherwise.} \end{cases} \quad (14)$$

This concludes the definition of the tuple  $T' = (\alpha + 1, \beta, E', \ell', S')$ .

**Claim 6.**

$(w_1, a_1) \overset{Y}{\circ} (w_2, a_2)$  in tree  $T$  iff  $(w_1, a_1) \overset{Y}{\circ} (w_2, a_2)$  in tree  $T'$ , for each dataset  $Y \subseteq V$ , each  $w_1, w_2 < \alpha$ , and each  $a_1, a_2 < \beta$ .

To show that  $T'$  is a loaded Cartesian tree, we need to verify conditions (a)–(d) of item 2 of Definition 7.

We verify these conditions in Claim 8, Claim 10, Claim 11, and Claim 13 below. First, let us prove an auxiliary statement used in the proof of Claim 8.

**Claim 7.**

$O_Y \psi \in S_{ub}$  iff  $O_Y \psi \in Q$ , for each dataset  $Y \subseteq X$  and each formula  $\psi \in \Phi$ .

**Proof of Claim.**

( $\Rightarrow$ ): By Lemma 6 and the Modus Ponens rule, the assumption  $O_Y \psi \in S_{ub}$  implies that  $S_{ub} \vdash O_Y \psi$ . Hence,  $S_{ub} \vdash O_X O_Y \psi$  by the Monotonicity axiom, the Modus Ponens inference rule, and the assumption  $Y \subseteq X$ . Thus,  $O_X O_Y \psi \in S_{ub}$  because  $S_{ub}$  is a maximal consistent set. Then,  $O_Y \psi \in Q^-$  by equation (12). Therefore,  $O_Y \psi \in Q$  because  $Q^- \subseteq Q$ .

( $\Leftarrow$ ): Suppose that  $O_Y \psi \notin S_{ub}$ . Thus,  $\neg O_Y \psi \in S_{ub}$  because  $S_{ub}$  is a maximal consistent set. Then,  $S_{ub} \vdash O_Y \neg O_Y \psi$  by the Negative Introspection axiom and the Modus Ponens inference rule. Hence,  $S_{ub} \vdash O_X \neg O_Y \psi$  by the Monotonicity axiom, the Modus Ponens inference rule, and the assumption  $Y \subseteq X$  of the claim. Thus,  $O_X \neg O_Y \psi \in S_{ub}$  because  $S_{ub}$  is a maximal consistent set. Then,  $\neg O_Y \psi \in Q^-$  by equation (12). Hence,  $\neg O_Y \psi \in Q$  because  $Q^- \subseteq Q$ . Therefore,  $O_Y \psi \notin Q$  because set  $Q$  is consistent.  $\square$

We are now ready to prove condition 2(b) of Definition 7.

**Claim 8.**

$O_Y \psi \in S'_{w_1 a}$  iff  $O_Y \psi \in S'_{w_2 a}$ , for each  $w_1, w_2 < \alpha + 1$ , each  $a < \beta$ , and each  $Y \subseteq V$  such that  $(w_1, a) \overset{Y}{\circ} (w_2, a)$  in  $T'$ .

**Proof of Claim.**

We consider the following three cases separately.

*Case 1:*  $w_1 = w_2$ . Then,  $O_Y \psi \in S'_{w_1 a}$  iff  $O_Y \psi \in S'_{w_2 a}$ .

*Case 2:*  $w_1, w_2 < \alpha$ . Then,  $(w_1, a) \overset{Y}{\circ} (w_2, a)$  in  $T$  by Claim 6 and the assumption  $(w_1, a) \overset{Y}{\circ} (w_2, a)$  in  $T'$  of the current claim. Hence,  $O_Y \psi \in S_{w_1 a}$  iff  $O_Y \psi \in S_{w_2 a}$  by condition 2(b) of Definition 7. Therefore,  $O_Y \psi \in S'_{w_1 a}$  iff  $O_Y \psi \in S'_{w_2 a}$  by equation (14).

*Case 3:* Exactly one of  $w_1, w_2$  is equal to  $\alpha$ . Without loss of generality, suppose that

$$w_1 < \alpha \quad \text{and} \quad w_2 = \alpha. \quad (15)$$

Then, the assumption  $(w_1, a) \overset{Y}{\sim} (w_2, a)$  in  $T'$  of the claim implies that  $(w_1, a) \overset{Y}{\sim} (\alpha, a)$  in  $T'$ . In other words, in tree  $T'$ , the unique path from node  $(w_1, a)$  to node  $(\alpha, a)$  is labeled with each data variable from set  $Y$ . Note, see Fig. 8, that this path must go through the node  $(u, b)$  and contain the edge  $\langle (u, b), (\alpha, a) \rangle$ . Thus,

$$Y \subseteq \ell'(\langle (u, b), (\alpha, a) \rangle) \quad (16)$$

and  $(w_1, a) \overset{Y}{\sim} (u, b)$  in tree  $T'$ . The last statement, by Claim 6, implies

$$(w_1, a) \overset{Y}{\sim} (u, b) \text{ in } T. \quad (17)$$

We further split this case into two subcases:

*Case 3A:  $a = b$ .* Then,  $\ell'(\langle (u, b), (\alpha, a) \rangle) = X$  by equation (11). Hence, by statement (16),

$$Y \subseteq X. \quad (18)$$

In addition, statement (17) and the assumption  $a = b$  of Case 3A imply that  $O_Y \psi \in S_{w_1a}$  iff  $O_Y \psi \in S_{ub}$  by item 2(b) of Definition 7. Hence,  $O_Y \psi \in S_{w_1a}$  iff  $O_Y \psi \in Q$  by Claim 7 and statement (18). Then,  $O_Y \psi \in S'_{w_1a}$  iff  $O_Y \psi \in S'_{\alpha b}$  by equation (14). Therefore,  $O_Y \psi \in S'_{w_1a}$  iff  $O_Y \psi \in S'_{w_2a}$  because of the part  $w_2 = \alpha$  of statement (15) and the assumption  $a = b$  of Case 3A.

*Case 3B:  $a \neq b$ .* Then,  $\ell'(\langle (u, b), (\alpha, a) \rangle) = \emptyset$  by equation (11). Hence, by statement (16),

$$Y = \emptyset. \quad (19)$$

By the assumption of the lemma  $u < \alpha$ . Thus, the node  $(u, a)$  also belongs to the tree  $T$ . Hence, vacuously, all edges along the unique simple path between the nodes  $(w_1, a)$  and  $(u, a)$  are labeled with each element of the empty set. In other words,  $(w_1, a) \overset{\emptyset}{\sim} (u, a)$  in  $T$ . Thus,  $(w_1, a) \overset{Y}{\sim} (u, a)$  in  $T$  by statement (19). Then,  $O_Y \psi \in S_{w_1a}$  iff  $O_Y \psi \in S_{ua}$  by item 2(b) of Definition 7. Hence,  $O_Y \psi \in S'_{w_1a}$  iff  $O_Y \psi \in S'_{\alpha a}$  by equation (14), the part  $w_1 < \alpha$  of statement (15), and the assumption  $a \neq b$  of Case 3B. Therefore,  $O_Y \psi \in S'_{w_1a}$  iff  $O_Y \psi \in S'_{w_2a}$  by the part  $w_2 = \alpha$  of statement (15).  $\square$

Claim 10 verifies condition 2(c) of Definition 7. The next claim is an auxiliary statement toward proving Claim 10.

#### Claim 9.

$D_Y \psi \in S'_{\alpha a}$  iff  $D_Y \psi \in S'_{ub}$ , for each formula  $\psi \in \Phi$ , each ordinal  $a < \beta$ , and each dataset  $Y \subseteq \ell'(\langle (\alpha, a), (u, b) \rangle)$ .

#### Proof of Claim.

We consider the following two cases separately:

*Case 1:  $a \neq b$ .* Then the assumption  $Y \subseteq \ell'(\langle (\alpha, a), (u, b) \rangle)$  of the claim implies, see Fig. 8,

$$Y = \emptyset. \quad (20)$$

The assumption  $(u, b) \in \alpha \times \beta$  of the lemma implies that  $u < \alpha$ . Thus, nodes  $(u, b)$  and  $(u, a)$  both belong to the tree  $T$ . Vacuously, all edges along the simple path connecting these nodes in tree  $T$  are labeled with each element of the empty set. Hence,  $(u, b) \overset{\emptyset}{\sim} (u, a)$  in tree  $T$ . Then,  $D_{\emptyset} \psi \in S_{ub}$  iff  $D_{\emptyset} \psi \in S_{ua}$  by item 2(c) of Definition 7. Thus,  $D_{\emptyset} \psi \in S'_{ub}$  iff  $D_{\emptyset} \psi \in S'_{\alpha a}$  by equation (14) and the assumption  $a \neq b$  of the claim. Therefore,  $D_Y \psi \in S'_{ub}$  iff  $D_Y \psi \in S'_{\alpha a}$  by equation (20).

*Case 2:  $a = b$ .* Then, the assumption  $Y \subseteq \ell'(\langle (\alpha, a), (u, b) \rangle)$  of the claim imply that  $(\alpha, a) \overset{Y}{\sim} (u, a)$  in  $T'$ . Thus,  $O_Y D_Y \psi \in S'_{\alpha a}$  iff  $O_Y D_Y \psi \in S'_{ua}$  by Claim 8. Hence,  $D_Y \psi \in S'_{\alpha a}$  iff  $D_Y \psi \in S'_{ua}$  by Lemma 7 because  $S'_{\alpha a}$  and  $S'_{ua}$  are maximal consistent sets. Therefore,  $D_Y \psi \in S'_{\alpha a}$  iff  $D_Y \psi \in S'_{ub}$  by the assumption  $a = b$  of the case.  $\square$

#### Claim 10.

$D_Y \psi \in S'_{w_1a_1}$  iff  $D_Y \psi \in S'_{w_2a_2}$  for each formula  $\psi \in \Phi$ , each  $w_1, w_2 < \alpha + 1$ , each  $a_1, a_2 < \beta$ , and each  $Y \subseteq V$  such that  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$ .

#### Proof of Claim.

We prove the claim by induction on the length of the simple path between the nodes  $(w_1, a_1)$  and  $(w_2, a_2)$  in tree  $T'$ .

*Base Case.*  $(w_1, a_1) = (w_2, a_2)$ . Then,  $w_1 = w_2$  and  $a_1 = a_2$ . Therefore,  $D_Y \psi \in S'_{w_1a_1}$  iff  $D_Y \psi \in S'_{w_2a_2}$ .

*Induction Step.* Consider the following two cases separately:

*Case 1:  $w_1, w_2 < \alpha$ .* Then,  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T$  by Claim 6 and the assumption  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$  of the current claim. Thus,  $D_Y \psi \in S_{w_1a_1}$  iff  $D_Y \psi \in S_{w_2a_2}$  by item 2(c) of Definition 7. Therefore,  $D_Y \psi \in S'_{w_1a_1}$  iff  $D_Y \psi \in S'_{w_2a_2}$  by equation (14) and the assumption  $w_1, w_2 < \alpha$  of the case.

*Case 2: At least one of  $w_1$  and  $w_2$  is equal to  $\alpha$ .* Without loss of generality, let  $w_1 = \alpha$ . Consider the simple path between the nodes  $(w_1, a_1) = (\alpha, a_1)$  and  $(w_2, a_2)$  in tree  $T'$ . Because we consider the induction step, this path has more than one node. Observe that the node  $(\alpha, a_1)$  in tree  $T'$  is only connected to node  $(u, b)$ , see Fig. 8. Thus, the path has the form:  $(\alpha, a_1), (u, b), \dots, (w_2, a_2)$ . Then, by the assumption  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$  of the claim,

$$Y \subseteq \ell'(\langle (\alpha, a_1), (u, b) \rangle). \quad (21)$$

$$(u, b) \overset{Y}{\sim} (w_2, a_2) \text{ in } T'. \quad (22)$$

Thus,  $D_Y \psi \in S'_{\alpha a_1}$  iff  $D_Y \psi \in S'_{ub}$  by Claim 9 and statement (21). Additionally, by the induction hypothesis, statement (22) implies that  $D_Y \psi \in S'_{ub}$  iff  $D_Y \psi \in S'_{w_2a_2}$ . Hence,  $D_Y \psi \in S'_{w_1a_1}$  iff  $D_Y \psi \in S'_{w_2a_2}$  because  $w_1 = \alpha$ .  $\square$

The next claim verifies condition 2(d) of Definition 7.

#### Claim 11.

If  $R_Y \psi \in S'_{w_1a_1}$  and  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$ , then  $\psi \in S'_{w_2a_1}$ , for each  $w_1, w_2 < \alpha + 1$ , each  $a_1, a_2 < \beta$ , each dataset  $Y \subseteq V$ , and each formula  $\psi \in \Phi$ .

#### Proof of Claim.

One of the following four cases must take place, see Fig. 8:

*Case 1: either  $Y = \emptyset$  or  $a_1 = a_2$ .* In the former case, all edges along the simple path between nodes  $(w_1, a_1)$  and  $(w_2, a_1)$  are labeled with all data variables from set  $Y$ . Hence,  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_1)$  in  $T'$ . In the latter case, the same statement is true by the assumption of the claim that  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$ . Thus,  $O_Y \psi \in S'_{w_1a_1}$  iff  $O_Y \psi \in S'_{w_2a_1}$  by Claim 8. Note also that  $O_Y \psi \in S'_{w_1a_1}$  by the assumption  $R_Y \psi \in S'_{w_1a_1}$  of the claim, the Overt Knowledge axiom, the Modus Ponens inference rule, and the maximality of the set  $S'_{w_1a_1}$ . Hence,  $O_Y \psi \in S'_{w_2a_1}$ . Therefore,  $\psi \in S'_{w_2a_1}$  by the Truth axiom, the Modus Ponens inference rule, and the maximality of the set  $S'_{w_1a_1}$ .

*Case 2:  $w_1, w_2 < \alpha$ .* Then,  $R_Y \psi \in S_{w_1a_1}$  by the assumption  $R_Y \psi \in S'_{w_1a_1}$  of the claim and equation (14). Additionally,  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T$  by the assumption  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$  of the current claim and Claim 6. Hence,  $\psi \in S_{w_2a_1}$  by item 2(d) of Definition 7. Therefore,  $\psi \in S'_{w_2a_1}$  by equation (14).

*Case 3:  $a_1 \neq a_2$ ,* one of the nodes  $(w_1, a_1)$  and  $(w_2, a_2)$  belongs to tree  $T$  and the other is the node  $(\alpha, b)$ . Then, the simple path between the nodes  $(w_1, a_1)$  and  $(w_2, a_2)$  must go through the node  $(u, b)$ , see Fig. 8. Furthermore, note that the assumption  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$

in  $T'$  of the claim implies that

$$(w_1, a_1) \overset{Y}{\sim} (u, b) \text{ in } T', \quad (23)$$

$$(u, b) \overset{Y}{\sim} (w_2, a_2) \text{ in } T'. \quad (24)$$

We further split this case into two subcases:

*Case 3A:*  $w_1 < \alpha$  and  $(w_2, a_2) = (\alpha, b)$ . Then,  $R_Y \psi \in S_{w_1 a_1}$  by the assumption  $R_Y \psi \in S'_{w_1 a_1}$  of the claim and equation (14). Note also that  $(w_1, a_1) \overset{Y}{\sim} (u, b)$  in  $T$  by statement (23) and Claim 6. Thus,  $\psi \in S_{u a_1}$  by item 2(d) of Definition 7. Then,  $\psi \in S'_{\alpha a_1}$  by equation (14) and the assumption  $a_1 \neq a_2$  of Case 3. Therefore,  $\psi \in S'_{w_2 a_1}$  by the assumption  $(w_2, a_2) = (\alpha, b)$  of Case 3A.

*Case 3B:*  $w_2 < \alpha$  and  $(w_1, a_1) = (\alpha, b)$ . By the Introspection of De Re Knowledge axiom and the Modus Ponens inference rule, the assumption  $R_Y \psi \in S'_{w_1 a_1}$  of the claim implies  $S'_{w_1 a_1} \vdash O_Y R_Y \psi$ . Then,  $O_Y R_Y \psi \in S'_{w_1 a_1}$  because  $S'_{w_1 a_1}$  is a maximal consistent set. Thus,  $O_Y R_Y \psi \in S'_{u b}$  by statement (23) and Claim 8. Hence,  $S'_{u b} \vdash R_Y \psi$  by the Truth axiom and the Modus Ponens inference rule. Then,  $R_Y \psi \in S'_{u b}$  because  $S'_{u b}$  is a maximal consistent set. Thus, by equation (14) and the assumption  $(u, b) \in \alpha \times \beta$  of the lemma,

$$R_Y \psi \in S_{u b}. \quad (25)$$

At the same time, statement (24), by Claim 6, the assumption  $(u, b) \in \alpha \times \beta$  of the lemma, and the assumption  $w_2 < \alpha$  of Case 3B, implies that  $(u, b) \overset{Y}{\sim} (w_2, a_2)$  in  $T$ . Thus,  $\psi \in S_{w_2 b}$  by item 2(d) of Definition 7 and statement (25). Hence,  $\psi \in S_{w_2 a_1}$  because  $b = a_1$  by the assumption  $(w_1, a_1) = (\alpha, b)$  of Case 3B. Therefore,  $\psi \in S'_{w_2 a_1}$  by equation (14) and the assumption  $w_2 < \alpha$  of Case 3B.  $\square$

The condition 2(a) of Definition 7 is verified in Claim 13. The next claim is an auxiliary statement used in the proof of Claim 13. Note that we have chosen to delay the verification of condition 2(a) because the proof of Claim 12 is using Claim 10.

#### Claim 12.

$!Y \in S'_{w_1 a_1}$  iff  $!Y \in S'_{w_2 a_2}$ , for each  $Y \subseteq V$ , each  $w_1, w_2 < \alpha + 1$ , and each  $a_1, a_2 < \beta$  such that  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$ .

#### Proof of Claim.

It suffices to show that if  $!Y \in S'_{w_1 a_1}$ , then  $!Y \in S'_{w_2 a_2}$ . Indeed, by the Introspection of Traceability axiom and the Modus Ponens inference rule, the assumption  $!Y \in S'_{w_1 a_1}$  implies that  $S'_{w_1 a_1} \vdash D_Y(!Y)$ . Hence,  $D_Y(!Y) \in S'_{w_1 a_1}$  because  $S'_{w_1 a_1}$  is a maximal consistent set. Thus,  $D_Y(!Y) \in S'_{w_2 a_2}$  by Claim 10 and the assumption  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$  of the current claim. Hence,  $S'_{w_2 a_2} \vdash !Y$  by Lemma 5. Therefore,  $!Y \in S'_{w_2 a_2}$  because  $S'_{w_2 a_2}$  is a maximal consistent set.  $\square$

#### Claim 13.

If  $!Y \in S'_{w_1 a_1}$  and  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$ , then  $a_1 = a_2$ , for each dataset  $Y \subseteq V$ , each  $w_1, w_2 < \alpha + 1$ , and each  $a_1, a_2 < \beta$ .

#### Proof of Claim.

We prove the statement of the claim by induction on the length of the simple path between the nodes  $(w_1, a_1)$  and  $(w_2, a_2)$  in tree  $T'$ .

*Base Case:* if the length of the path is zero, then  $(w_1, a_1) = (w_2, a_2)$ . Therefore,  $a_1 = a_2$ .

*Induction Step:* we further split the induction step into the following three cases:

*Case 1:*  $w_1, w_2 < \alpha$ . Then,  $!Y \in S_{w_1 a_1}$  by equation (14). Also, by Claim 6 and the assumption  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$  of the current claim,  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T$ . Therefore,  $a_1 = a_2$  by item 2(a) of Definition 7.

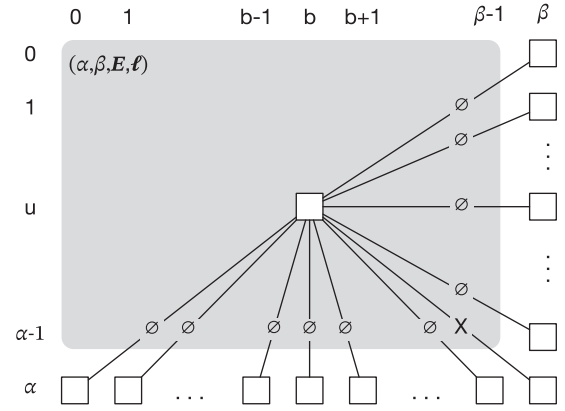


Figure 9. Toward a proof of Lemma 11.

*Case 2:* at least one of  $w_1, w_2$  is equal to  $\alpha$ . By Claim 12, without loss of generality, we can suppose that  $w_2 = \alpha$ . Note that  $(w_1, a_1) \neq (w_2, a_2)$  because we consider the induction step of the proof. Thus, the assumption  $w_2 = \alpha$  implies that the simple path in tree  $T'$  from the node  $(w_1, a_1)$  to the node  $(w_2, a_2) = (\alpha, a_2)$  must go through the node  $(u, b)$ , see Fig. 8. Then, the simple path from the node  $(w_1, a_1)$  to the node  $(u, b)$  is shorter than the simple path from the node  $(w_1, a_1)$  to the node  $(w_2, a_2)$ . In addition,

$$(w_1, a_1) \overset{Y}{\sim} (u, b) \text{ in } T', \quad (26)$$

$$Y \subseteq \ell'((u, b), (\alpha, a_2))) \quad (27)$$

by the assumption  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$  of the claim. We further split the proof into the following two subcases:

*Case 2A:*  $a_2 = b$ . Note that  $a_1 = b$  by the induction hypothesis, the assumption  $!Y \in S'_{w_1 a_1}$  of the claim and statement (26). Therefore,  $a_1 = a_2$  by the assumption  $a_2 = b$  of Case 2A.

*Case 2B:*  $a_2 \neq b$ . Then,  $\ell'((u, b), (\alpha, a_2))) = \emptyset$ , see Fig. 8. Thus,  $Y = \emptyset$  by statement (27). Then, vacuously, all edges along the simple path between the nodes  $(u, b)$  and  $(u, a_2)$  in tree  $T$  are labeled with all elements of the set  $Y$ . In other words,

$$(u, b) \overset{Y}{\sim} (u, a_2) \text{ in } T. \quad (28)$$

At the same time, the assumption  $!Y \in S'_{w_1 a_1}$  of the claim implies that  $!Y \in S'_{u b}$  by Claim 12 and statement (26). Thus,  $!Y \in S_{u b}$  by equation (14) and because  $u < \alpha$  due to the assumption  $(u, b) \in (\alpha, \beta)$  of the lemma. Therefore,  $a_2 = b$  by item 2(a) of Definition 7 and statement (28), which contradicts the assumption of Case 2B.  $\square$

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

#### Lemma 11.

For any finite loaded Cartesian tree  $T = (\alpha, \beta, E, \ell, S)$ , any node  $(u, b) \in \alpha \times \beta$ , and any formulae  $D_X \varphi \notin S_{u b}$  and  $R_X \psi \notin S_{u b}$  such that  $!X \notin S_{u b}$  there is an extension  $T' = (\alpha + 1, \beta + 1, E', \ell', S')$  of tree  $T$  such that  $(u, b) \overset{X}{\sim} (\alpha, \beta)$  in tree  $T'$ ,  $\varphi \notin S'_{\alpha \beta}$ , and  $\psi \notin S'_{\alpha b}$ .

#### Proof.

Following the general structure of the proof of Lemma 10, we first define the Cartesian tree  $(\alpha + 1, \beta + 1, E', \ell')$ . This tree is obtained by connecting all newly added nodes to the node  $(u, b)$  of the original tree  $(\alpha, \beta, E, \ell)$ , see Fig. 9. Formally, relation  $E'$  on the set  $(\alpha + 1) \times$

$(\beta + 1)$  is defined as follows:

$$E' = E \cup \{((u, b), (\alpha, \beta))\} \cup \{((u, b), (\alpha, a)) \mid a < \beta\} \\ \cup \{((u, b), (w, \beta)) \mid w < \alpha\}. \quad (29)$$

We label all newly added edges with the empty set except for the edge  $((u, b), (\alpha, \beta))$ , which is labeled with set  $X$ , see Fig. 9. Formally, we define edge labeling function  $\ell'$  as follows:

$$\ell'(e) = \begin{cases} X, & \text{if } e = ((u, b), (\alpha, \beta)), \\ \emptyset, & \text{if } e = ((u, b), (\alpha, a)), \text{ where } a < \beta, \\ \emptyset, & \text{if } e = ((u, b), (w, \beta)), \text{ where } w < \alpha, \\ \ell(e), & \text{otherwise.} \end{cases} \quad (30)$$

The above equation concludes the definition of the Cartesian tree  $(\alpha + 1, \beta + 1, E', \ell')$ . Toward the definition of the labeling function  $S'$ , let us first consider the following two sets of formulae:

$$H^- = \{\neg\varphi\} \cup \{\chi \in \Phi \mid D_X\chi \in S_{ub}\}, \quad (31)$$

$$G^- = \{\neg\psi\} \cup \{\chi \in \Phi \mid R_X\chi \in S_{ub}\}. \quad (32)$$

The proof of the next claim is the same as the proof of Claim 5 except that it uses modalities  $D$  and  $R$  instead of modality  $O$ .  $\square$

**Claim 14.**

$H^-$  and  $G^-$  are consistent sets of formulae.

By Lemma 9, sets  $H^-$  and  $G^-$  can be extended to maximal consistent sets  $H$  and  $G$ , respectively. We are now ready to define the labeling function  $S'$ . The values of the labelling function  $S'_{ua}$  are given by the  $(\alpha + 1) \times (\beta + 1)$  matrix shown in below equation:

$$(S'_{ua}) = \begin{bmatrix} S_{00} & S_{01} & \dots & S_{0,b-1} & S_{0b} & S_{0,b+1} & \dots & S_{0,\beta-1} & H^- \\ S_{10} & S_{11} & \dots & S_{1,b-1} & S_{1b} & S_{1,b+1} & \dots & S_{1,\beta-1} & H \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ S_{u0} & S_{u1} & \dots & S_{u,b-1} & S_{ub} & S_{u,b+1} & \dots & S_{u,\beta-1} & H \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ S_{\alpha-1,0} & S_{\alpha-1,1} & \dots & S_{\alpha-1,b-1} & S_{\alpha-1,b} & S_{\alpha-1,b+1} & \dots & S_{\alpha-1,\beta-1} & H \\ S_{u0} & S_{u1} & \dots & S_{u,b-1} & G & S_{u,b+1} & \dots & S_{u,\beta-1} & H \end{bmatrix}.$$

Matrix  $(S'_{ua})$  towards the proof of Lemma 11.

In other words, let

$$S'_{ua} = \begin{cases} S_{ua}, & \text{if } w = \alpha, a \neq b, \text{ and } a \neq \beta, \\ G, & \text{if } w = \alpha \text{ and } a = b, \\ H, & \text{if } a = \beta, \\ S_{wa}, & \text{otherwise.} \end{cases} \quad (33)$$

The tuple  $T' = (\alpha + 1, \beta + 1, E', \ell', S')$  is now defined. Next, we show that  $T'$  is a loaded Cartesian tree through verifying conditions 2(a)–(d) of Definition 7. We start with the following equivalent of Claim 6.

**Claim 15.**

$(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in tree  $T$  iff  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in tree  $T'$ , for each dataset  $Y \subseteq V$ , each  $w_1, w_2 < \alpha$ , and each  $a_1, a_2 < \beta$ .

Claim 17 verifies condition 2(b) of Definition 7. Claim 16 is an auxiliary statement used in the proof of Claim 17.

**Claim 16.**

$O_\emptyset\chi \in S_{ub}$  iff  $O_\emptyset\chi \in G$ , for each formula  $\chi \in \Phi$ .

**Proof of Claim.**

$(\Rightarrow)$ : Suppose that  $O_\emptyset\chi \in S_{ub}$ . Then, by Lemma 6 and the Modus Ponens inference rule,  $S_{ub} \vdash O_\emptyset O_\emptyset\chi$ . Hence,  $S_{ub} \vdash R_\emptyset O_\emptyset\chi$  by the Data-Free Knowledge axiom and the Modus Ponens inference rule. Then,  $S_{ub} \vdash R_X O_\emptyset\chi$  by the Monotonicity axiom and the Modus Ponens inference rule. Thus,  $R_X O_\emptyset\chi \in S_{ub}$  because  $S_{ub}$  is a

maximal consistent set of formulae. Therefore,  $O_\emptyset\chi \in G^- \subseteq G$  by equation (32).

$(\Leftarrow)$ : Suppose that  $O_\emptyset\chi \notin S_{ub}$ . Then,  $\neg O_\emptyset\chi \in S_{ub}$  because  $S_{ub}$  is a maximal consistent set of formulae. Thus,  $S_{ub} \vdash O_\emptyset \neg O_\emptyset\chi$  by the Negative Introspection axiom and the Modus Ponens inference rule. Hence,  $S_{ub} \vdash R_\emptyset \neg O_\emptyset\chi$  by the Data-Free Knowledge axiom and the Modus Ponens inference rule. Thus,  $S_{ub} \vdash R_X \neg O_\emptyset\chi$  by the Monotonicity axiom and the Modus Ponens inference rule. Then,  $R_X \neg O_\emptyset\chi \in S_{ub}$  because  $S_{ub}$  is a maximal consistent set of formulae. Thus,  $\neg O_\emptyset\chi \in G^- \subseteq G$  by equation (32). Therefore,  $O_\emptyset\chi \notin G$  because set  $G$  is consistent.  $\square$

**Claim 17.**

$O_Y\chi \in S'_{w_1a}$  iff  $O_Y\chi \in S'_{w_2a}$ , for each formula  $\chi \in \Phi$ , each  $w_1, w_2 < \alpha + 1$ , each  $a < \beta + 1$ , and each dataset  $Y \subseteq V$  such that  $(w_1, a) \overset{Y}{\sim} (w_2, a)$  in  $T'$ .

**Proof of Claim.**

We consider the following four cases separately:

*Case 1:*  $w_1 = w_2$ . Then,  $S'_{w_1a} = S'_{w_2a}$ . Therefore,  $O_Y\chi \in S'_{w_1a}$  iff  $O_Y\chi \in S'_{w_2a}$ .

*Case 2:*  $a = \beta$ . Then,  $S'_{w_1a} = H = S'_{w_2a}$  by equation (33). Therefore,  $O_Y\chi \in S'_{w_1a}$  iff  $O_Y\chi \in S'_{w_2a}$ .

*Case 3:*  $a < \beta$  and  $w_1, w_2 < \alpha$ . Then,  $(w_1, a) \overset{Y}{\sim} (w_2, a)$  in  $T$  by Claim 15 and the assumption  $(w_1, a) \overset{Y}{\sim} (w_2, a)$  in  $T'$  of the current claim. Hence,  $O_Y\chi \in S_{w_1a}$  iff  $O_Y\chi \in S_{w_2a}$  by item 2(b) of Definition 7. Therefore,  $O_Y\chi \in S'_{w_1a}$  iff  $O_Y\chi \in S'_{w_2a}$  by equation (33) and the assumptions  $a < \beta$  and  $w_1, w_2 < \alpha$  of the case.

*Case 4:*  $a < \beta$  and one of  $w_1, w_2$  is equal to  $\alpha$  while the other is less than  $\alpha$ . Without loss of generality, let  $w_1 < \alpha$  and  $w_2 = \alpha$ . Hence, the simple path between nodes  $(w_1, a)$  and  $(w_2, a)$  must contain the edge  $((u, b), (w_2, a))$ , see Fig. 9. In addition,  $Y \subseteq \ell'(((u, b), (w_2, a)))$  by the assumption  $(w_1, a) \overset{Y}{\sim} (w_2, a)$  in  $T'$  of the claim. Note that  $\ell'(((u, b), (w_2, a))) = \emptyset$  by either equation (30) or Fig. 9 and the assumption of the case that  $a < \beta$ . Thus,  $Y = \emptyset$ . Hence, it suffices to show that  $O_\emptyset\chi \in S'_{w_1a}$  iff  $O_\emptyset\chi \in S'_{w_2a}$ .

Observe that nodes  $(w_1, a)$  and  $(u, a)$  belong to the tree  $T$  by the assumption of Case 4 and the assumption  $u < \alpha$  of the lemma. Then, all edges along the simple path in tree  $T$  between these nodes are vacuously labeled with all elements of the empty set. Hence,  $(w_1, a) \overset{\emptyset}{\sim} (u, a)$  in  $T$ . Thus, by item 2(b) of Definition 7,

$$O_\emptyset\chi \in S_{w_1a} \text{ iff } O_\emptyset\chi \in S_{ua}. \quad (34)$$

We further split this case into two subcases:

*Case 4A:*  $a = b$ . Hence,  $O_\emptyset\chi \in S_{ua}$  iff  $O_\emptyset\chi \in G$  by Claim 16. Then,  $O_\emptyset\chi \in S_{w_1a}$  iff  $O_\emptyset\chi \in G$  by statement (34). Thus,  $O_\emptyset\chi \in S'_{w_1a}$  iff  $O_\emptyset\chi \in S'_{w_2a}$  by equation (33), the assumptions  $a < \beta$ ,  $w_1 < \alpha$ , and  $w_2 = \alpha$  of Case 4 and the assumption  $a = b$  of Case 4A.

*Case 4B:*  $a \neq b$ . Then,  $O_\emptyset\chi \in S_{ua}$  iff  $O_\emptyset\chi \in S'_{\alpha a}$  by equation (33), the assumption  $a < \beta$  of Case 4 and the assumption  $u < \alpha$  of the lemma. Thus,  $O_\emptyset\chi \in S_{w_1a}$  iff  $O_\emptyset\chi \in S'_{\alpha a}$  by statement (34). Therefore,  $O_\emptyset\chi \in S'_{w_1a}$  iff  $O_\emptyset\chi \in S'_{w_2a}$  by equation (33), the assumptions  $a < \beta$ ,  $w_1 < \alpha$ , and  $w_2 = \alpha$  of Case 4.  $\square$

Claim 19 verifies condition 2(c) of Definition 7. Claim 18 is an auxiliary statement used in the proof of Claim 19.

**Claim 18.**

$D_Y\chi \in S'_{ua}$  iff  $D_Y\chi \in S'_{ub}$ , for each  $\chi \in \Phi$ , each  $(w, a) \in (\alpha + 1) \times (\beta + 1)$  such that  $(w, a) \notin \alpha \times \beta$ , and each dataset  $Y \subseteq \ell'(((w, a), (u, b)))$ .

**Proof of Claim.**

We consider the following two cases separately:

*Case 1:*  $w = \alpha$ ,  $a < \beta$ . Then,  $\ell'((w, a), (u, b)) = \emptyset$  by equation (30); alternatively, see Fig. 9. Thus,  $Y = \emptyset$  by the assumption  $Y \subseteq \ell'((w, a), (u, b))$  of the claim. Hence, it suffices to prove that  $D_{\emptyset} \chi \in S'_{wa}$  iff  $D_{\emptyset} \chi \in S'_{ub}$ . We further split this case into two subcases.

*Case 1A:*  $a \neq b$ . The assumption  $(u, b) \in \alpha \times \beta$  of the lemma implies that  $u < \alpha$  and  $b < \beta$ . Hence,  $(u, a), (u, b) \in \alpha \times \beta$  by the assumption  $a < \beta$  of Case 1. Thus, nodes  $(u, a)$  and  $(u, b)$  belong to tree  $T$ . Note that all edges along the simple path between these two nodes are vacuously labeled with all elements of the empty set. Hence,  $(u, a) \stackrel{\emptyset}{\sim} (u, b)$  in tree  $T$ . Thus,  $D_{\emptyset} \chi \in S_{ua}$  iff  $D_{\emptyset} \chi \in S_{ub}$  by item 2(c) of Definition 7. Then,  $D_{\emptyset} \chi \in S'_{ua}$  iff  $D_{\emptyset} \chi \in S'_{ub}$  by equation (33) and the assumption  $w = \alpha$  of Case 1 and the assumption  $a \neq b$  of Case 1A. Hence,  $D_{\emptyset} \chi \in S'_{ua}$  iff  $D_{\emptyset} \chi \in S'_{ub}$  by equation (33) and the assumption  $(u, b) \in \alpha \times \beta$  of the lemma.

*Case 1B:*  $a = b$ . Note that, vacuously, all edges along the path between the nodes  $(w, a)$  and  $(u, b)$  in tree  $T'$  are labeled with all elements of the empty set. In other words,  $(w, a) \stackrel{\emptyset}{\sim} (u, b)$  in tree  $T'$ . Hence,  $O_{\emptyset} D_{\emptyset} \chi \in S'_{wa}$  iff  $O_{\emptyset} D_{\emptyset} \chi \in S'_{ub}$  by Claim 17 and the assumption  $a = b$  of Case 1B. Therefore,  $D_{\emptyset} \chi \in S'_{wa}$  iff  $D_{\emptyset} \chi \in S'_{ub}$  by Lemma 7 because  $S'_{wa}$  and  $S'_{ub}$  are maximal consistent sets of formulae.

*Case 2:*  $a = \beta$ . Then, the assumption  $Y \subseteq \ell'((w, a), (u, b))$  of the claim implies, see Fig. 9,

$$Y \subseteq X. \quad (35)$$

In addition, by equation (33) and the assumption  $(u, b) \in \alpha \times \beta$ , it suffices to prove that  $D_Y \chi \in H$  iff  $D_Y \chi \in S_{ub}$ .

$(\Rightarrow)$ : Suppose that  $D_Y \chi \notin S_{ub}$ . Then,  $\neg D_Y \chi \in S_{ub}$  because  $S_{ub}$  is a maximal consistent set. Hence,  $S_{ub} \vdash D_Y \neg D_Y \chi$  by the Negative Introspection axiom and the Modus Ponens inference rule. Thus,  $S_{ub} \vdash D_X \neg D_Y \chi$  by the Monotonicity axiom and statement (35). Hence,  $D_X \neg D_Y \chi \in S_{ub}$  because  $S_{ub}$  is a maximal consistent set. Then,  $\neg D_Y \chi \in H^- \subseteq H$  by equation (31). Therefore,  $D_Y \chi \notin H$  because set  $H$  is consistent.

$(\Leftarrow)$ : Let  $D_Y \chi \in S_{ub}$ . Then,  $S_{ub} \vdash D_Y D_Y \chi$  by Lemma 6 and the Modus Ponens inference rule. Thus,  $S_{ub} \vdash D_X D_Y \chi$  by the Monotonicity axiom and statement (35). Hence,  $D_X D_Y \chi \in S_{ub}$  because  $S_{ub}$  is a maximal consistent set of formulae. Then,  $D_Y \chi \in H^- \subseteq H$  by equation (31).  $\square$

The proof of the next claim is similar to the proof of Claim 10 except it uses Claim 15 and Claim 18 instead of Claim 6 and Claim 9.

#### Claim 19.

$D_Y \chi \in S'_{w_1 a_1}$  iff  $D_Y \chi \in S'_{w_2 a_2}$ , for each dataset  $Y \subseteq V$ , each formula  $\chi \in \Phi$ , each  $w_1, w_2 \leq \alpha + 1$  and each  $a_1, a_2 \leq \beta + 1$  such that  $(w_1, a_1) \stackrel{Y}{\sim} (w_2, a_2)$  in  $T'$ .

Next, we verify condition 2(d) of Definition 7.

#### Claim 20.

If  $R_Y \chi \in S'_{w_1 a_1}$  and  $(w_1, a_1) \stackrel{Y}{\sim} (w_2, a_2)$  in  $T'$ , then  $\chi \in S'_{w_2 a_1}$  for each dataset  $Y \subseteq V$ , each formula  $\chi \in \Phi$ , each  $w_1, w_2 \leq \alpha + 1$  and each  $a_1, a_2 \leq \beta + 1$ .

#### Proof of Claim.

The assumption  $(w_1, a_1) \stackrel{Y}{\sim} (w_2, a_2)$  in  $T'$  implies that one of the following cases must take place, see Fig. 9.

*Case 1:*  $(w_1, a_1) = (w_2, a_2)$ . The assumption  $R_Y \chi \in S'_{w_1 a_1}$  implies that  $S'_{w_1 a_1} \vdash \chi$  by Lemma 5. Hence,  $\chi \in S'_{w_1 a_1}$  because  $S'_{w_1 a_1}$  is a maximal consistent set. Therefore,  $\chi \in S'_{w_2 a_1}$  by the assumption of the case.

*Case 2:*  $w_1, w_2 < \alpha$  and  $a_1, a_2 < \beta$ . Thus, the assumption  $(w_1, a_1) \stackrel{Y}{\sim} (w_2, a_2)$  in  $T'$  implies that  $(w_1, a_1) \stackrel{Y}{\sim} (w_2, a_2)$  in  $T$  by Claim 15. In addition, the assumption  $R_Y \chi \in S'_{w_1 a_1}$  implies that  $R_Y \chi \in S_{w_1 a_1}$  by equation (33). Then,  $\chi \in S_{w_2 a_1}$  by item 2(d) of Definition 7. Therefore,  $\chi \in S'_{w_2 a_1}$  by equation (33) and the assumptions  $w_2 < \alpha$  and  $a_1 < \beta$  of the case.

*Case 3:*  $Y = \emptyset$ . Thus, by the assumption  $R_Y \chi \in S'_{w_1 a_1}$  of the claim,  $R_{\emptyset} \chi \in S'_{w_1 a_1}$ . Then,  $S'_{w_1 a_1} \vdash O_{\emptyset} \chi$  by the Overt Knowledge axiom and the Modus Ponens inference rule. Hence,  $O_{\emptyset} \chi \in S'_{w_1 a_1}$  because  $S'_{w_1 a_1}$  is a maximal consistent set of formulae. Observe that, vacuously, all edges along the simple path between the nodes  $(w_1, a_1)$  and  $(w_2, a_1)$  in tree  $T'$  are labeled with each data variable from the empty set. In other words,  $(w_1, a_1) \stackrel{\emptyset}{\sim} (w_2, a_1)$  in  $T'$ . Thus,  $O_{\emptyset} \chi \in S'_{w_2 a_1}$  by Claim 17. Hence,  $S'_{w_2 a_1} \vdash \chi$  by the Truth axiom and the Modus Ponens inference rule. Therefore,  $\chi \in S'_{w_2 a_1}$  because  $S'_{w_2 a_1}$  is a maximal consistent set.

*Case 4:*  $a_1 = \beta$ . The assumption  $R_Y \chi \in S'_{w_1 a_1}$  of the claim implies that  $S'_{w_1 a_1} \vdash \chi$  by Lemma 5. Thus,  $\chi \in S'_{w_1 a_1}$  because  $S'_{w_1 a_1}$  is a maximal consistent set. Then,  $\chi \in H$  by equation (33) and the assumption  $a_1 = \beta$  of the case. Therefore,  $\chi \in S'_{w_2 a_1}$  again by equation (33) and the assumption  $a_1 = \beta$  of the case.

*Case 5:*  $w_2 = \alpha$ ,  $w_1 < \alpha$ ,  $a_1 < \beta$ , and  $Y \subseteq X$ . Then, any path from the node  $(w_1, a_1)$  to the node  $(w_2, a_2)$  must go through the node  $(u, b)$ , see Fig. 9. Hence, the assumption  $(w_1, a_1) \stackrel{Y}{\sim} (w_2, a_2)$  in  $T'$  of the claim implies that

$$(w_1, a_1) \stackrel{Y}{\sim} (u, b) \text{ in } T'. \quad (36)$$

Hence, by Claim 15, the assumption  $(u, b) \in \alpha \times \beta$  of the lemma, and the assumptions  $w_1 < \alpha$  and  $a_1 < \beta$  of the case,

$$(w_1, a_1) \stackrel{Y}{\sim} (u, b) \text{ in } T. \quad (37)$$

We further split this case into two subcases:

*Case 5A:*  $a_1 = b$ . The assumption  $R_Y \chi \in S'_{w_1 a_1}$  of the claim implies  $S'_{w_1 a_1} \vdash O_Y R_Y \chi$  by the Introspection of De Re Knowledge axiom and the Modus Ponens rule. Hence,  $O_Y R_Y \chi \in S'_{w_1 a_1}$  because  $S'_{w_1 a_1}$  is a maximal consistent set. Thus,  $O_Y R_Y \chi \in S'_{ub}$  by Claim 17, statement (36) and the assumption  $a_1 = b$  of Case 5A. Then,  $O_Y R_Y \chi \in S_{ub}$  by equation (33) and the assumption  $(u, b) \in \alpha \times \beta$  of the lemma. Hence,  $S_{ub} \vdash R_Y \chi$  by the Truth axiom and the Modus Ponens inference rule. Thus,  $S_{ub} \vdash R_X \chi$  by the Monotonicity axiom, the Modus Ponens inference rule, and the assumption  $X \subseteq Y$  of Case 5. Then,  $R_X \chi \in S_{ub}$  because  $S_{ub}$  is a maximal consistent set. Hence,  $\chi \in G^- \subseteq G$  by equation (32). Thus,  $\chi \in S'_{w_2 a_1}$  by equation (33), the assumption  $w_2 = \alpha$  of Case 5 and the assumption  $a_1 = b$  of Case 5A.

*Case 5B:*  $a_1 \neq b$ . The assumption  $R_Y \chi \in S'_{w_1 a_1}$  of the claim implies  $R_Y \chi \in S_{w_1 a_1}$  by equation (33) and the assumptions  $w_1 < \alpha$  and  $a_1 < \beta$  of Case 5. Hence,  $\chi \in S_{u, a_1}$  by item 2(d) of Definition 7 and statement (37). Therefore,  $\chi \in S_{w_2, a_1}$  by equation (33), the assumptions  $w_2 = \alpha$  and  $a_1 < \beta$  of Case 5 and the assumption  $a_1 \neq b$  of Case 5B.

Finally, we verify condition 2(a) of Definition 7.

#### Claim 21.

If  $!Y \in S'_{w_1 a_1}$  and  $(w_1, a_1) \stackrel{Y}{\sim} (w_2, a_2)$  in  $T'$ , then  $a_1 = a_2$  for each dataset  $Y \subseteq V$ , each  $w_1, w_2 \leq \alpha + 1$  and each  $a_1, a_2 \leq \beta + 1$ .

#### Proof of Claim.

At least one of the following three cases must take place, see Fig. 9:

*Case 1:*  $(w_1, a_1) = (w_2, a_2)$ . Then,  $a_1 = a_2$ .

*Case 2:*  $w_1, w_2 < \alpha$  and  $a_1, a_2 < \beta$ . Then, the assumption  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$  implies that  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T$  by Claim 15. Also, the assumption  $!Y \in S'_{w_1 a_1}$  implies  $!Y \in S_{w_1 a_1}$  by equation (33). Therefore,  $a_1 = a_2$  by item 2(a) of Definition 7.

*Case 3:* The simple path between the nodes  $(w_1, a_1)$  and  $(w_2, a_2)$  goes through the node  $(u, b)$ . Then,

$$(w_1, a_1) \overset{Y}{\sim} (u, b) \text{ in } T', \quad (38)$$

$$Y \subseteq X. \quad (39)$$

by the assumption  $(w_1, a_1) \overset{Y}{\sim} (w_2, a_2)$  in  $T'$  of the claim. At the same time, the assumption  $!Y \in S'_{w_1 a_1}$  of the claim implies that  $S'_{w_1 a_1} \vdash D_Y !Y$  by the Introspection of the Traceability axiom and the Modus Ponens inference rule. Hence,  $D_Y !Y \in S'_{w_1 a_1}$  because  $S'_{w_1 a_1}$  is a maximal consistent set. Thus,  $D_Y !Y \in S'_{ub}$  by statement (38) and Claim 19. Then,  $S'_{ub} \vdash !Y$  by Lemma 5. Thus,  $S'_{ub} \vdash !X$  by the Monotonicity axiom, statement (39) and the Modus Ponens inference rule. Then,  $!X \in S'_{ub}$  because  $S'_{ub}$  is a maximal consistent set. Therefore,  $!X \in S_{ub}$  by equation (33) and the assumption  $(u, b) \in \alpha \times \beta$  of the lemma, which contradicts the assumption  $!X \notin S_{ub}$  of the lemma.  $\square$

This completes the proof of the lemma.  $\square$

For any tuples of sets  $(A_1, \dots, A_n)$  and  $(B_1, \dots, B_n)$  of the same length, let  $(A_1, \dots, A_n) \subseteq (B_1, \dots, B_n)$  if  $A_i \subseteq B_i$  for each  $i \leq n$ . For any infinite chain of tuples

$$(A_1^1, \dots, A_n^1) \subseteq (A_1^2, \dots, A_n^2) \subseteq (A_1^3, \dots, A_n^3) \subseteq \dots$$

by  $\bigcup_i (A_1^i, \dots, A_n^i)$  we mean the tuple  $(\bigcup_i A_1^i, \dots, \bigcup_i A_n^i)$ . Note that, by Definition 7, each loaded Cartesian tree is a tuple  $(\alpha, \beta, E, \ell, S)$  where  $\alpha$  and  $\beta$  are ordinals,  $E$  is a set of edges,  $\ell$  and  $S$  are labeling functions. Following the set theory tradition, we view ordinals and functions as sets. The next lemma follows from Definition 7.

**Lemma 12.**

For any infinite chain of loaded trees  $T_1 \subseteq T_2 \subseteq T_3 \subseteq \dots$ , if loaded tree  $T_{i+1}$  is an extension of the loaded tree  $T_i$  for each  $i \geq 1$ , then  $\bigcup_i T_i$  is a loaded tree that extends each of the loaded trees  $T_1, T_2, \dots$ .

Intuitively, Definition 9 requires a complete loaded tree to have no “deficiencies”. It is convenient to talk about each such deficiency separately. To be able to do this, below we introduce the notion of  $(u, b, \psi)$ -complete loaded tree.

**Definition 10.**

For any loaded Cartesian tree  $T = (\alpha, \beta, E, \ell, S)$ , any integers  $u < \alpha$  and  $b < \beta$ , and any formula  $\psi \in \Phi$ , tree  $T$  is  $(u, b, \psi)$ -complete if the following conditions are satisfied:

- (1) if formula  $\psi$  has the form  $!X$  and  $!X \notin S_{ub}$ , then there is  $(w, a) \in \alpha \times \beta$  such that  $(u, b) \overset{X}{\sim} (w, a)$  and  $b \neq a$ ,
- (2) if formula  $\psi$  has the form  $O_X \varphi$  and  $O_X \varphi \notin S_{ub}$ , then there is  $(w, b) \in \alpha \times \beta$  such that  $(u, b) \overset{X}{\sim} (w, b)$  and  $\varphi \notin S_{wb}$ ,
- (3) if formula  $\psi$  has the form  $D_X \varphi$  and  $D_X \varphi \notin S_{ub}$ , then there is  $(w, a) \in \alpha \times \beta$  such that  $(u, b) \overset{X}{\sim} (w, a)$  and  $\varphi \notin S_{wa}$ ,
- (4) if formula  $\psi$  has the form  $R_X \varphi$  and  $R_X \varphi \notin S_{ub}$ , then there is  $(w, a) \in \alpha \times \beta$  such that  $(u, b) \overset{X}{\sim} (w, a)$  and  $\varphi \notin S_{wb}$ .

The next lemma follows from Definition 8 and Definition 10.

**Lemma 13.**

If a loaded Cartesian tree  $T$  is  $(u, b, \psi)$ -complete, then any extension of tree  $T$  is also  $(u, b, \psi)$ -complete.

The lemma below follows from Definition 9 and Definition 10.

**Lemma 14.**

If a loaded Cartesian tree  $T = (\alpha, \beta, E, \ell, S)$  is  $(u, b, \psi)$ -complete for all integers  $u < \alpha$  and  $b < \beta$  and each formula  $\psi \in \Phi$ , then tree  $T$  is complete.

**Lemma 15.**

For any finite loaded Cartesian tree  $T = (\alpha, \beta, E, \ell, S)$ , any integers  $u < \alpha$  and  $b < \beta$ , and any formula  $\psi \in \Phi$ , if tree  $T$  is not  $(u, b, \psi)$ -complete, then there is a finite  $(u, b, \psi)$ -complete extension of tree  $T$ .

**Proof.**

Note that  $S_{ub} \not\vdash \perp$  because set  $S_{ub}$  is consistent. Then,  $S_{ub} \not\vdash D_X \perp$  and  $S_{ub} \not\vdash R_X \perp$  by Lemma 5 applied contrapositively. Thus,

$$D_X \perp \notin S_{ub} \text{ and } R_X \perp \notin S_{ub}. \quad (40)$$

To finish the proof of the lemma, we consider the cases corresponding to conditions 1-4 of Definition 10 separately.

*Case 1:* suppose that formula  $\psi$  has the form  $!X$  and  $!X \notin S_{ub}$ . Hence, by Lemma 11 and statement (40), there is an extension  $T' = (\alpha + 1, \beta + 1, E', \ell', S')$  of tree  $T$  such that  $(u, b) \overset{X}{\sim} (\alpha, \beta)$  in tree  $T'$ . Thus,  $T'$  is a  $(u, b, \psi)$ -complete extension of tree  $T$  by Definition 10.

*Case 2:* suppose formula  $\psi$  has the form  $O_X \varphi$  and  $O_X \varphi \notin S_{ub}$ . Then, by Lemma 10 there is an extension  $T' = (\alpha + 1, \beta, E', \ell', S')$  of tree  $T$  such that  $(u, b) \overset{X}{\sim} (\alpha, b)$  in tree  $T'$  and  $\varphi \notin S'_{\alpha b}$ . Thus,  $T'$  is a  $(u, b, \psi)$ -complete extension of tree  $T$  by Definition 10.

*Case 3:* suppose formula  $\psi$  has the form  $D_X \varphi$  and  $D_X \varphi \notin S_{ub}$ . We further slit this case into two subcases.

*Case 3A:*  $!X \in S_{ub}$ . The assumption  $D_X \varphi \notin S_{ub}$  of Case 3 implies  $S_{ub} \not\vdash D_X \varphi$  because  $S_{ub}$  is a maximal consistent set. Thus, it follows that  $S_{ub} \not\vdash O_X \varphi$  by the Traceable Data axiom applied contrapositively and the assumption  $!X \in S_{ub}$  of Case 3A. Hence,  $O_X \varphi \notin S_{ub}$  because  $S_{ub}$  is a maximal consistent set. Then, by Lemma 10 there is an extension  $T' = (\alpha + 1, \beta, E', \ell', S')$  of tree  $T$  such that  $(u, b) \overset{X}{\sim} (\alpha, b)$  in tree  $T'$  and  $\varphi \notin S'_{\alpha b}$ . Thus,  $T'$  is a  $(u, b, \psi)$ -complete extension of tree  $T$  by Definition 10.

*Case 3B:*  $!X \notin S_{ub}$ . Hence, by the assumption  $D_X \varphi \notin S_{ub}$  of Case 3, the part  $R_X \perp \notin S_{ub}$  of statement (40), and Lemma 11, there is an extension  $T' = (\alpha + 1, \beta + 1, E', \ell', S')$  of tree  $T$  such that  $(u, b) \overset{X}{\sim} (\alpha, \beta)$  in tree  $T'$  and  $\varphi \notin S'_{\alpha \beta}$ . Thus,  $T'$  is a  $(u, b, \psi)$ -complete extension of tree  $T$  by Definition 10.

*Case 4:* suppose formula  $\psi$  has the form  $R_X \varphi$  and  $R_X \varphi \notin S_{ub}$ . Similar to the previous case, we further slit this case into two subcases.

*Case 4A:*  $!X \in S_{ub}$ . By applying the argument similar to the one in Case 3A, but using the form  $!X \rightarrow (O_X \rightarrow R_X)$  of the Traceability axiom instead of the form  $!X \rightarrow (O_X \rightarrow D_X)$ , one can conclude that there is an extension  $T' = (\alpha + 1, \beta, E', \ell', S')$  of tree  $T$  such that  $(u, b) \overset{X}{\sim} (\alpha, b)$  in tree  $T'$  and  $\varphi \notin S'_{\alpha b}$ . Thus,  $T'$  is a  $(u, b, \psi)$ -complete extension of tree  $T$  by Definition 10.

*Case 4B:*  $!X \notin S_{ub}$ . Thus, by the assumption  $R_X \varphi \notin S_{ub}$  of Case 4, the part  $D_X \perp \notin S_{ub}$  of statement (40), and Lemma 11, there is an extension  $T' = (\alpha + 1, \beta + 1, E', \ell', S')$  of tree  $T$  such that  $(u, b) \overset{X}{\sim} (\alpha, \beta)$  in tree  $T'$  and  $\varphi \notin S'_{\alpha \beta}$ . Therefore,  $T'$  is a  $(u, b, \psi)$ -complete extension of tree  $T$  by Definition 10.  $\square$

**Theorem 5 (extension).**

Any finite loaded Cartesian tree can be extended to a complete loaded Cartesian tree.

**Proof.**

Let  $T$  be a finite loaded Cartesian tree. To prove the theorem, we will construct an infinite chain of finite loaded Cartesian trees  $T = T_1 \subseteq$

$T_2 \subseteq T_3 \subseteq \dots$  such that  $T_{i+1}$  is an extension of tree  $T_i$  for each integer  $i \geq 1$ . Toward the construction of this chain, consider an enumeration  $(u_1, b_1, \psi_1), (u_2, b_2, \psi_2), \dots$  of all triples  $(u, b, \psi)$  such that  $u$  and  $b$  are nonnegative integer numbers and  $\psi \in \Phi$  is a formula.

We define chain  $T = T_1 \subseteq T_2 \subseteq T_3 \subseteq \dots$  recursively. Suppose that tree  $T_k = (\alpha_k, \beta_k, E_k, \ell_k, S_k)$  is already defined. If  $T_k$  is  $(u, b, \psi)$ -complete for each  $u < \alpha_k, b < \beta_k$ , and  $\psi \in \Phi$ , then let  $T_{k+1} = T_k$ . Otherwise, let  $i_{\min}$  be the smallest  $i$  such that  $u_i < \alpha_k, b_i < \beta_k$ , and  $T_k$  is not  $(u_i, b_i, \psi_i)$ -complete. By Lemma 15, tree  $T_k$  can be extended to a finite  $(u_i, b_i, \psi_i)$ -complete tree  $T_{k+1}$ .

Let  $T' = \bigcup_{k \geq 1} T_k = (\alpha', \beta', E', \ell', S')$ . Note that  $T'$  is an extension of all trees  $T_k$  by Lemma 12.  $T'$  is  $(u, b, \psi)$ -complete for each  $u < \alpha'$  and  $b < \beta'$  by the construction of the chain  $T = T_1 \subseteq T_2 \subseteq T_3 \subseteq \dots$  and Lemma 13. Therefore,  $T'$  is complete by Lemma 14.  $\square$

### Canonical model

In this subsection, we define a canonical model based on a complete loaded Cartesian tree. Recall that we think about such a tree as a matrix. In our construction, the rows of the matrix become possible worlds of the model, and the columns of the matrix become the agents.

#### Definition 11.

For each complete loaded Cartesian tree  $T = (\alpha, \beta, E, \ell, S)$ , let canonical model  $M(T)$  be defined as the tuple  $(\alpha, \beta, \{\sim\}_{x \in V}, \pi)$ , such that

- (1)  $(w_1, a_1) \sim_x (w_2, a_2)$  if  $(w_1, a_1) \overset{[x]}{\sim} (w_2, a_2)$  for each  $x \in V$ , each  $w_1, w_2 \in \alpha$ , and each  $a_1, a_2 \in \beta$ ,
- (2)  $\pi(p) = \{(w, a) \in \alpha \times \beta \mid p \in S_{wa}\}$  for each  $p$ .

#### Lemma 16.

$(w_1, a_1) \sim_X (w_2, a_2)$  iff  $(w_1, a_1) \overset{X}{\sim} (w_2, a_2)$ , for each  $w_1, w_2 \in \alpha$ , each  $a_1, a_2 \in \beta$ , and each dataset  $X \subseteq V$ .

#### Proof.

By the definition, the statement  $(w_1, a_1) \sim_X (w_2, a_2)$  is equivalent to the statement that  $(w_1, a_1) \sim_x (w_2, a_2)$  for each data variable  $x \in X$ . By item 1 of Definition 11, the last statement is equivalent to the statement that  $(w_1, a_1) \overset{[x]}{\sim} (w_2, a_2)$  for each data variable  $x \in X$ . By Definition 6, the previous statement is equivalent to the statement that for each  $x \in X$  there is a simple path between nodes  $(w_1, a_1)$  and  $(w_2, a_2)$  such that all edges along this path are labeled with data variable  $x$ . Note that in any tree there is a *unique* simple path between any two nodes. Thus, the last statement is equivalent to the statement that all edges along this *unique* simple path between nodes  $(w_1, a_1)$  and  $(w_2, a_2)$  are labeled with each variable from set  $X$ . Finally, by Definition 6, the previous statement is equivalent to  $(w_1, a_1) \overset{X}{\sim} (w_2, a_2)$ .  $\square$

#### Lemma 17.

$w, a \models \varphi$  iff  $\varphi \in S_{wa}$ , for any formula  $\varphi \in \Phi$ , any world  $w \in \alpha$ , and any agent  $a \in \beta$  of any canonical model  $M(T)$  based on a complete loaded Cartesian tree  $T = (\alpha, \beta, E, \ell, S)$ .

#### Proof.

We prove the statement of the lemma by induction on structural complexity of formula  $\varphi$ . If formula  $\varphi$  is an atomic proposition, then the statement of the lemma follows from item 1 of Definition 2 and item 2 of Definition 11. The case when formula  $\varphi$  is a negation or a disjunction follows from items 3 and 4 of Definition 2 and the fact that  $S_{wa}$  is a maximal consistent set in the standard way.

Suppose that formula  $\varphi$  has the form  $!X$ .

$(\Rightarrow)$ : Assume that  $!X \notin S_{wa}$ . Then, by item 1 of Definition 9, there is a node  $(w', a')$  such that  $(w, a) \overset{X}{\sim} (w', a')$  and  $a \neq a'$ . Hence,  $(w, a) \sim_X (w', a')$  by Lemma 16. Therefore,  $w, a \not\models !X$  by item 2 of Definition 2.

$(\Leftarrow)$ : Assume that  $w, a \not\models !X$ . Then, by item 2 of Definition 2, there exist a world  $w'$  and an agent  $a'$  such that  $(w, a) \sim_X (w', a')$  and  $a \neq a'$ . Hence,  $(w, a) \overset{X}{\sim} (w', a')$  by Lemma 16. Therefore,  $!X \notin S_{wa}$  by item 2(a) of Definition 7 applied contrapositively.

Suppose that formula  $\varphi$  has the form  $O_X \psi$ .

$(\Rightarrow)$ : Assume that  $O_X \psi \notin S_{wa}$ . Then, by item 2 of Definition 9, there is a node  $(w', a)$  such that  $(w, a) \overset{X}{\sim} (w', a)$  and  $\psi \notin S_{w'a}$ . Hence,  $(w, a) \sim_X (w', a)$  by Lemma 16 and  $w', a \not\models \psi$  by the induction hypothesis. Therefore,  $w, a \not\models O_X \psi$  by item 5 of Definition 2.

$(\Leftarrow)$ : Suppose that  $O_X \psi \in S_{wa}$ . Consider any world  $w'$  such that  $(w, a) \sim_X (w', a)$ . By item 5 of Definition 2, it suffices to show that  $w', a \models \psi$ .

The assumption  $(w, a) \sim_X (w', a)$  implies  $(w, a) \overset{X}{\sim} (w', a)$  by Lemma 16. Then,  $O_X \psi \in S_{w'a}$  by item 2(b) of Definition 7 and the assumption  $O_X \psi \in S_{wa}$ . Hence,  $S_{w'a} \vdash \psi$  by the Truth axiom and the Modus Ponens inference rule. Thus,  $\psi \in S_{w'a}$  because  $S_{w'a}$  is a maximal consistent set. Therefore,  $w', a \models \psi$  by the induction hypothesis.

Suppose that formula  $\varphi$  has the form  $D_X \psi$ .

$(\Rightarrow)$ : Assume that  $D_X \psi \notin S_{wa}$ . Then, by item 3 of Definition 9, there is a node  $(w', a')$  such that  $(w, a) \overset{X}{\sim} (w', a')$  and  $\psi \notin S_{w'a'}$ . Hence,  $(w, a) \sim_X (w', a')$  by Lemma 16 and  $w', a' \not\models \psi$  by the induction hypothesis. Therefore,  $w, a \not\models D_X \psi$  by item 6 of Definition 2.

$(\Leftarrow)$ : Suppose that  $D_X \psi \in S_{wa}$ . Consider any world  $w'$  and any agent  $a'$  such that  $(w, a) \sim_X (w', a')$ . By item 6 of Definition 2, it suffices to show that  $w', a' \models \psi$ .

The assumption  $(w, a) \sim_X (w', a')$  implies  $(w, a) \overset{X}{\sim} (w', a')$  by Lemma 16. Then,  $D_X \psi \in S_{w'a'}$  by item 2(c) of Definition 7 and the assumption  $D_X \psi \in S_{wa}$ . Then,  $S_{w'a'} \vdash \psi$  by Lemma 5. Thus,  $\psi \in S_{w'a'}$  because  $S_{w'a'}$  is a maximal consistent set. Therefore,  $w', a' \models \psi$  by the induction hypothesis.

Suppose that formula  $\varphi$  has the form  $R_X \psi$ .

$(\Rightarrow)$ : Assume that  $R_X \psi \notin S_{wa}$ . Then, by item 4 of Definition 9, there is a node  $(w', a')$  such that  $(w, a) \overset{X}{\sim} (w', a')$  and  $\psi \notin S_{w'a'}$ . Hence,  $(w, a) \sim_X (w', a')$  by Lemma 16 and  $w', a' \not\models \psi$  by the induction hypothesis. Therefore,  $w, a \not\models R_X \psi$  by item 7 of Definition 2.

$(\Leftarrow)$ : Suppose that  $R_X \psi \in S_{wa}$ . Consider any world  $w'$  and any agent  $a'$  such that  $(w, a) \sim_X (w', a')$ . By item 7 of Definition 2, it suffices to show that  $w', a' \models \psi$ .

The assumption  $(w, a) \sim_X (w', a')$  implies  $(w, a) \overset{X}{\sim} (w', a')$  by Lemma 16. Then,  $\psi \in S_{w'a'}$  by item 2(d) of Definition 7 and the assumption  $R_X \psi \in S_{wa}$ . Therefore,  $w', a' \models \psi$  by the induction hypothesis.  $\square$

### Theorem 6 (strong completeness).

For any set of formulae  $\Gamma \subseteq \Phi$  and any formula  $\varphi \in \Phi$ , if  $\Gamma \not\vdash \varphi$ , then there is a world  $w \in W$  and an agent  $a \in \mathcal{A}$  of some model, such that  $w, a \models \gamma$  for each formula  $\gamma \in \Gamma$  and  $w, a \not\models \varphi$ .

#### Proof.

The assumption  $\Gamma \not\vdash \varphi$  implies that the set  $\Gamma \cup \{\neg\varphi\}$  is consistent. Let  $\Delta$  be any maximal consistent extension of this set. Consider the loaded Cartesian tree  $(1, 1, \emptyset, \emptyset, S)$ , where  $S_{00} = \Delta$ . In other words, the loaded cartesian tree consists of a single node  $(0,0)$  labeled with set  $\Delta$ . Because the set of edges is empty, the tuple  $(1, 1, \emptyset, \emptyset)$  is a Cartesian tree by Definition 5. Thus, to prove that the tuple  $(1, 1, \emptyset, \emptyset, S)$  is a loaded Cartesian tree, it suffices to show that conditions 2(a)-(d) of Definition 7 are satisfied. Conditions 2(a)-(c) are vacuously satisfied because the tree has only one node. To verify con-

dition 2(d), it suffices to show that if  $R_X \psi \in S_{00}$ , then  $\psi \in S_{00}$ . Indeed, the assumption  $R_X \psi \in S_{00}$  implies  $S_{00} \vdash \psi$  by Lemma 5 and the Modus Ponens inference rule. Therefore,  $\psi \in S_{00}$  because  $S_{00}$  is a maximal consistent set.

By Theorem 5, tree  $(1, 1, \emptyset, \emptyset, S)$  can be extended to a complete loaded Cartesian tree  $T' = (\alpha, \beta, E, \ell, S')$ . Note that  $S'_{00} = S_{00} = \Delta \supseteq \Gamma \cup \{\neg\varphi\}$  by item 5 of Definition 8. Consider canonical model  $M(T')$ . By Lemma 17, we have  $0, 0 \Vdash \gamma$  for each  $\gamma \in \Gamma$  and  $0, 0 \Vdash \neg\varphi$  because  $\Gamma \cup \{\neg\varphi\} \subseteq S'_{00}$ . Note that the statement  $0, 0 \Vdash \neg\varphi$  implies  $0, 0 \not\Vdash \varphi$  by item 3 of Definition 2.

## Future work: $k$ -anonymity

The term “ $k$ -anonymity” is usually used to refer to a setting where some information about an agent not only cannot be used to identify the agent precisely but it even cannot be used to identify a group of  $k$  agents to which the agent belongs. This concept has been extensively studied in the database literature [32,33]. It also was investigated by logicians in the context of knowledge [34] and actions [35]. In the setting of our article, there are at least two different notions that capture different aspects of  $k$ -anonymity.

First, we can generalize our traceability expression  $!X$  to  $k$ -traceability by modifying item 2 of Definition 2 as follows:

$w, a \Vdash !^k X$ , when there are agents  $a_1, \dots, a_k \in \mathcal{A}$  such that for each world  $u \in W$  and each agent  $b \in \mathcal{A}$ , if  $(w, a) \sim_X (u, b)$ , then  $b \in \{a_1, \dots, a_k\}$ .

Note that  $!^1 X$  is equivalent to the traceability expression  $!X$  we studied earlier in this article. The Monotonicity and the Introspection of Traceability axioms are also valid for  $k$ -traceability:

$$!^k X \rightarrow !^k Y, \text{ if } X \subseteq Y, \quad !^k X \rightarrow D_X !^k X.$$

Before discussing the other notion related to  $k$ -anonymity, let us go back to our de re modality. Recall from statement (4) that in world  $w_1$ , knowing *age*, *city*, and *reply* of agent  $a$  informs de re knowledge that agent  $a$  is pregnant. Let us now suppose that, in the same world  $w_1$ , instead of the questionnaire filled in by agent  $a$  we are looking at the questionnaire filled in by her husband, agent  $b$ . Although these two questionnaires are filled *identically*, see Fig. 2, agent  $b$ 's questionnaire, of course, does not inform the de re knowledge that he is pregnant:

$$w_1, b \not\Vdash R_{\text{age,city,reply}} \text{ “is pregnant”}.$$

Nevertheless, his questionnaire does inform de re knowledge that his wife, agent  $a$ , is pregnant. In other words, although the data in his questionnaire does not violate his own privacy, it does violate the privacy of his wife. The language of our original logical system cannot express this. However, one can introduce a new modality  $R^3$  (“de re knowledge about somebody”) that does it. The semantics of this modality can be defined as follows:

$w, a \Vdash R_X^3 \varphi$ , if there is a agent  $a' \in \mathcal{A}$  such that for each world  $u \in W$  and each agent  $b \in \mathcal{A}$ , if  $(w, a) \sim_X (u, b)$ , then  $u, a' \Vdash \varphi$ .

In our example,

$$w_1, b \Vdash R_{\text{age,city,reply}}^3 \text{ “is pregnant”}.$$

Some properties of the modality  $R^3$  are expected:

$$R_X \varphi \rightarrow R_X^3 \varphi, \quad R_X^3 \varphi \rightarrow R_Y^3 \varphi \text{ if } X \subseteq Y.$$

Among the *unexpected* ones is perhaps the introspection:

$$R_X^3 \varphi \rightarrow D_X R_X^3 \varphi.$$

Note that unlike the corresponding axiom for modality  $R$ , the introspection property for modality  $R^3$  is *stronger*, because it is true for de dicto knowledge, not just the overt knowledge.

Finally, let us get back to  $k$ -anonymity. Imagine a different survey where an agent is asked if one of the agent's children is getting married. Suppose that an agent  $a$ , who has  $k$  adult children, fills in the survey and answers this question positively. If the survey has enough information to trace the identity of the agent, then the dataset  $X$  in the survey informs the knowledge that one of the  $k$  children is getting married. We write this as  $R_X^k$  “is getting married”. In general, modality  $R^k$  can be defined as follows:

$w, a \Vdash R_X^k \varphi$ , if there are agents  $a_1, \dots, a_k \in \mathcal{A}$  such that for each world  $u \in W$  and each agent  $b \in \mathcal{A}$ , if  $(w, a) \sim_X (u, b)$ , then  $u, a_i \Vdash \varphi$  for some  $i \leq k$ .

Observe that the statement  $\neg R_X^k \varphi$  means that dataset  $X$  does not provide enough information to identify a group of  $k$  agents one of which has property  $\varphi$ . Thus, just like the expression  $\neg !^k X$  it captures an aspect of  $k$ -anonymity. Below are some of the properties of modality  $R^k$ :

$$\begin{aligned} R_X^1 \varphi &\leftrightarrow R_X^3 \varphi, & R_X^k \varphi &\rightarrow D_X R_X^k \varphi, \\ R_X^k \varphi &\rightarrow R_Y^k \varphi \text{ if } X \subseteq Y, & R_X^k \varphi &\rightarrow R_Y^l \varphi \text{ if } l \leq k. \end{aligned}$$

A complete axiomatization of the interplay between the expression  $!^k X$  and the modality  $R_X^k$  remains an open question.

## Conclusion

In this article, we observed that what one “knows” about an agent based on the agent's data can be defined in three different ways. We called them de re, de dicto, and overt forms of knowledge. Out of these three forms, only de dicto knowledge preserves the privacy of the agent. Our main technical results are the mutual undefinability of these three forms of knowledge in an egocentric setting even when the traceability expression  $!$  is used and a sound and complete logical system describing the interplay between the three forms of knowledge and the traceability expression.

Finally, in the “Future work” section, we discuss how our approach can be generalized to capture two forms of  $k$ -anonymity.

## Author contribution

Junli Jiang (Conceptualization-Equal, Formal analysis-Equal, Funding acquisition-Equal, Investigation-Equal, Methodology-Equal, Project administration-Equal, Resource, Writing - original draft-Equal, Writing - review & editing-Equal).

Pavel Naumov (Conceptualization-Equal, Formal analysis-Equal, Funding acquisition-Equal, Investigation-Equal, Methodology-Equal, Project administration-Equal, Resource, Writing - original draft-Equal, Writing - review & editing-Equal).

*Conflict of interest:* The authors wish to declare that they are not aware of any conflict of interest related to this work.

## Funding

J.J. acknowledges the support of the Key project of the National Social Science Fund of China (No. 24AZX019) and the China Scholarship Council (No. 202306990071).

## Appendix

To keep the article self-contained, in this [appendix](#), we give the proofs of the lemmas not proven in the main part of the article. We place them here because they are either straightforward or well-known.

**Lemma 18.**

The inference rule  $\frac{\varphi}{\Box_X \varphi}$ , where  $\Box \in \{O, R\}$ , is derivable.

**Proof.**

Suppose that  $\vdash \varphi$ . Thus,  $\vdash D_X \varphi$  by the Necessitation inference rule. Hence,  $\vdash O_X \varphi$  by the Overt Knowledge axiom and the Modus Ponens inference rule. This proves that the rule  $\frac{\varphi}{\Box_X \varphi}$  is derivable.

To prove derivability of the rule  $\frac{\varphi}{\Box_R \varphi}$ , assume again that  $\vdash \varphi$ . Thus,  $\vdash D_O \varphi$  by the Necessitation inference rule. Hence,  $\vdash R_O \varphi$  by the Data-Free Knowledge axiom and the Modus Ponens inference rule. Therefore,  $\vdash R_X \varphi$  by the Monotonicity axiom and the Modus Ponens inference rule.  $\square$

**Lemma 6.**

$\vdash \Box_X \varphi \rightarrow \Box_X \Box_X \varphi$ , where  $\Box \in \{O, D\}$ .

**Proof.**

Formula  $\Box_X \neg \Box_X \varphi \rightarrow \neg \Box_X \varphi$  is either an instance of the Truth axiom (if  $\Box$  is modality O) or Lemma 5 (if  $\Box$  is modality D). Thus, by the law of contraposition,  $\vdash \Box_X \varphi \rightarrow \neg \Box_X \neg \Box_X \varphi$ . Hence, taking into account the following instance of the Negative Introspection axiom:  $\neg \Box_X \neg \Box_X \varphi \rightarrow \Box_X \neg \Box_X \neg \Box_X \varphi$ , we have

$$\vdash \Box_X \varphi \rightarrow \Box_X \neg \Box_X \neg \Box_X \varphi. \quad (41)$$

At the same time,  $\neg \Box_X \varphi \rightarrow \Box_X \neg \Box_X \varphi$  is an instance of the Negative Introspection axiom. Thus,  $\vdash \neg \Box_X \neg \Box_X \varphi \rightarrow \Box_X \varphi$  by the law of contrapositive in the propositional logic. Hence,  $\vdash \Box_X (\neg \Box_X \neg \Box_X \varphi \rightarrow \Box_X \varphi)$  by either the Necessitation inference rule (if  $\Box$  is modality D) or Lemma 18 (if  $\Box$  is modality O). Thus, by the Distributivity axiom and the Modus Ponens inference rule,  $\vdash \Box_X \neg \Box_X \neg \Box_X \varphi \rightarrow \Box_X \Box_X \varphi$ . The latter, together with statement (41), implies the statement of the lemma by propositional reasoning.  $\square$

**Lemma 19 (deduction).**

If  $\Gamma, \varphi \vdash \psi$ , then  $\Gamma \vdash \varphi \rightarrow \psi$ .

**Proof.**

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

- (1)  $\vdash \psi_k$ , or
- (2)  $\psi_k \in \Gamma$ , or
- (3)  $\psi_k$  is equal to  $\varphi$ , or
- (4) there are  $i, j < k$  such that formula  $\psi_j$  is equal to  $\psi_i \rightarrow \psi_k$ .

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

*Case 1:*  $\vdash \psi_k$ . Note that  $\psi_k \rightarrow (\varphi \rightarrow \psi_k)$  is a propositional tautology, and thus, is an axiom of our logical system. Hence,  $\vdash \varphi \rightarrow \psi_k$  by the Modus Ponens rule. Therefore,  $\Gamma \vdash \varphi \rightarrow \psi_k$ .

*Case 2:*  $\psi_k \in \Gamma$ . Note again that  $\psi_k \rightarrow (\varphi \rightarrow \psi_k)$  is a propositional tautology, and thus, is an axiom of our logical system. Therefore, by the Modus Ponens rule,  $\Gamma \vdash \varphi \rightarrow \psi_k$ .

*Case 3:* formula  $\psi_k$  is equal to  $\varphi$ . Thus,  $\varphi \rightarrow \psi_k$  is a propositional tautology. Therefore,  $\Gamma \vdash \varphi \rightarrow \psi_k$ .

*Case 4:* formula  $\psi_j$  is equal to  $\psi_i \rightarrow \psi_k$  for some  $i, j < k$ . Thus, by the induction hypothesis,  $\Gamma \vdash \varphi \rightarrow \psi_i$  and  $\Gamma \vdash \varphi \rightarrow (\psi_i \rightarrow \psi_k)$ . Note that formula  $(\varphi \rightarrow \psi_i) \rightarrow ((\varphi \rightarrow (\psi_i \rightarrow \psi_k)) \rightarrow (\varphi \rightarrow \psi_k))$  is a propositional tautology. Therefore,  $\Gamma \vdash \varphi \rightarrow \psi_k$  by applying the Modus Ponens inference rule twice.  $\square$

**Lemma 8.**

If  $\varphi_1, \dots, \varphi_n \vdash \psi$ , then  $\Box_X \varphi_1, \dots, \Box_X \varphi_n \vdash \Box_X \psi$ , where  $\Box \in \{O, D, R\}$ .

**Proof.**

The assumption  $\varphi_1, \dots, \varphi_n \vdash \psi$ , by Lemma 19 applied  $n$  times, implies that

$$\vdash \varphi_1 \rightarrow (\varphi_2 \rightarrow \dots (\varphi_n \rightarrow \psi) \dots).$$

Thus, by either the Necessitation inference rule (if  $\Box$  is the modality O) or by Lemma 18 (if  $\Box$  is one of the modalities D and R),

$$\vdash \Box_X (\varphi_1 \rightarrow (\varphi_2 \rightarrow \dots (\varphi_n \rightarrow \psi) \dots)).$$

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

$$\vdash \Box_X \varphi_1 \rightarrow \Box_X (\varphi_2 \rightarrow \dots (\varphi_n \rightarrow \psi) \dots).$$

Then, again by the Modus Ponens rule,

$$\Box_X \varphi_1 \vdash \Box_X (\varphi_2 \rightarrow \dots (\varphi_n \rightarrow \psi) \dots).$$

Therefore,  $\Box_X \varphi_1, \dots, \Box_X \varphi_n \vdash \Box_X \psi$  by applying the previous steps ( $n-1$ ) more times.  $\square$

## References

1. Quine WV. Quantifiers and propositional attitudes. *J Philos* 1956; 53: 177–87.
2. Lewis D. Attitudes de dicto and de se. *Philos Review* 1979; 88: 513–43.
3. Chisholm R. Knowledge and belief: ‘De dicto’ and ‘de re’. *Philos Stud* 1976; 29: 1–20.
4. Keshet E, Schwarz F. De re/de dicto. In: *The Oxford Handbook of Reference*. Oxford, England: Oxford University Press. 2019, 167–202.
5. Wang Y, Seligman J. When names are not commonly known: epistemic logic with assignments. In: Bezhanishvili G, D’Agostino G, Metcalfe G al. et, (eds.). *Advances in Modal Logic 12, proceedings of the 12th conference on “Advances in Modal Logic”, held in Bern, Switzerland, August 27-31, 2018*. Rickmansworth: College Publications, 2018, 611–28.
6. Cohen M, Tang W, Wang Y. De Re Updates. In: Halpern JY, Perea A, (eds.). *Proceedings Eighteenth Conference on Theoretical Aspects of Rationality and Knowledge, TARK 2021, Beijing, China, June 25-27, 2021*. Vol. 335, Australia: Open Publishing Association Waterloo. 2021, 103–17.
7. Prior AN. Egocentric logic. *Noûs* 1968;2:91–207.
8. Seligman J, Liu F, Girard P. Logic in the community. In: *Logic and Its Applications*. New York: Springer. 2011. 178–88.
9. Seligman J, Liu F, Girard P. Facebook and the Epistemic Logic of Friendship. In: *14th conference on Theoretical Aspects of Rationality and Knowledge (TARK ‘13), January 2013, Chennai, India*. 229–38, 2013.
10. Christoff Z, Hansen JU. A logic for diffusion in social networks. *J Appl Logic* 2015; 13: 48–77.
11. Christoff Z, Hansen JU, Proietti C. Reflecting on social influence in networks. *J Logic Lang Inform* 2016; 25: 299–333.
12. Jiang J, Naumov P. A logic of higher-order preferences. *Synthese* 2024; 203: 210.
13. Schroeter L. Two-Dimensional Semantics. In: Zalta EN, (ed.). *The Stanford Encyclopedia of Philosophy*, Fall 2021ed. Stanford, California, United States: Metaphysics Research Lab, Stanford University, 2021. <https://plato.stanford.edu/archives/fall2021/entries/two-dimension-al-semantics/>.
14. Stalnaker RC. Assertion. In: *Pragmatics*, Leiden, The Netherlands: Brill, 1978, 315–32.
15. Grove AJ, Halpern JY. Naming and Identity in a Multi-Agent Epistemic Logic. In: Allen JF, Fikes R, Sandewall E, (eds.). *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR’91)*. Cambridge, MA, USA, April 22-25, 1991, 301–12, Burlington, Massachusetts: Morgan Kaufmann. 1991.
16. Grove AJ, Halpern JY. Naming and identity in epistemic logics Part I: the propositional case. *J Logic Comput* 1993; 3: 345–78.
17. Grove AJ. Naming and identity in epistemic logic Part II: a first-order logic for naming. *Artif Intell* 1995; 74: 311–50.
18. Epstein S, Naumov P. Epistemic Logic of Know-Who. In: *Proceedings of Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)*. Washington, DC USA: AAAI Press Washington. 2021.
19. Epstein S, Naumov P, Tao J. An egocentric logic of de dicto and de re knowing who. *J Logic Comput* 2023; 34: Issue 7, October 2024. 1347–1376.

20. Naumov P, Tao J. An egocentric logic of knowing how to tell them apart. *J Symbolic Logic* 2023; 1–33. <https://doi.org/10.1017/jsl.2023.45>.
21. Grossi D, Lorini E, Schwarzenrüder F. The ceteris paribus structure of logics of game forms. *J Artif Intel Res* 2015; 53: 91–126.
22. Baltag A, van Benthem J. A simple logic of functional dependence. *J Philos Logic* 2021; 50: 1–67.
23. Jiang J, Naumov P. Data-Informed Knowledge and Strategies. *Artif Intell* 2022; 309: 103727.
24. Deuser K, Jiang J, Naumov P. *et al.* A Dynamic Logic of Data-Informed Knowledge. *J Philos Logic* 2024; 53: 521–57.
25. Jiang J, Naumov P. A logic of trust-based beliefs. *Synthese* 2024; 204. <https://doi.org/10.1007/s11229-024-04671-3>.
26. Jiang J, Naumov P. The Logic of Doxastic Strategies. In: *Proceedings of 38th AAAI Conference on Artificial Intelligence (AAAI-24)*. Vol. 38. 19876–883, 2024.
27. Knight S, Naumov P, Shi Q. *et al.* Truth set algebra: a new way to prove undefinability. arXiv:220804422, 2022, preprint: not peer reviewed.
28. Shi Q. Responsibility in Extensive Form Games. In: *Proceedings of 38th AAAI Conference on Artificial Intelligence (AAAI-24)*. Washington, DC, USA: AAAI Press. 2024.
29. Cui X, Naumov P. Responsibility in Infinite Games. *Notre Dame Journal of Formal Logic* 2024.
30. Mendelson E. *Introduction to mathematical logic*. Boca Raton, Florida: CRC press, 2009. <https://www.amazon.co.uk/Introduction-Mathematical-Discrete-Mathematics-Applications/dp/1584888768>.
31. Fagin R, Halpern JY, Vardi MY. What can machines know? On the properties of knowledge in distributed systems. *J ACM (JACM)* 1992; 39: 328–76.
32. Sweeney L. K-anonymity: A Model for Protecting Privacy. *Int J Uncertain Fuzziness Knowl-Based Syst* 2002; 10: 557–70.
33. Domingo-Ferrer J, Sánchez D, Soria-Comas J. *Database anonymization: privacy models, data utility, and microaggregation-based inter-model connections* New York: Springer Nature. 2022.
34. Syverson PF, Stubblebine SG. Group principals and the formalization of anonymity. In: *International Symposium on Formal Methods*. 814–33, New York: Springer, 1999.
35. Halpern JY, O'Neill KR. Anonymity and information hiding in multi-agent systems. In: *Proceedings of the Sixteenth IEEE Computer Security Foundations Workshop*. 75–88, 2003.