Journal of Philosophical Logic
https://doi.org/10.1007/s10992-023-09740-5

A Dynamic Logic of Data-Informed Knowledge

Kaya Deuser[1] · Junli Jiang[2] · Pavel Naumov[3] · Wenxuan Zhang[4]

Received: 26 April 2023 / Accepted: 15 December 2023

Abstract
With agents relying more and more on information from central servers rather than their own sensors, knowledge becomes property not of a specific agent but of the data that the agents can access. The article proposes a dynamic logic of data-informed knowledge that describes an interplay between three modalities and one relation capturing the properties of this form of knowledge. The main technical results are the undefinability of two dynamic operators through each other, a sound and complete axiomatisation, and a model checking algorithm.

1 Introduction

"I was desperate," Mario Costeja Gonzalez says about his feelings after a client suggested that Gonzalez googles his own name. Typing his full name into the search engine would show that Gonzalez, a professional financial advisor, had his own house foreclosed in 1998 [3]. That conversation between Gonzalez and his client led to the famous 13 May 2014 "right to be forgotten" ruling by the European Court of Justice, ordering Google to remove the link to the foreclosure information. Once Google removed the link from its search results, it made the foreclosure information, essen-

✉ Pavel Naumov
p.naumov@soton.ac.uk

Kaya Deuser
kd469@cornell.edu

Junli Jiang
walk08@swu.edu.cn

Wenxuan Zhang
wenxzhang.work@gmail.com

1 Cornell University, New York, USA

2 Institute of Logic and Intelligence, Southwest University, Chongqing, China

3 School of Electronics and Computer Science, University of Southampton, Southampton, UK

4 Scripps College, Claremont, California, USA

Published online: 16 January 2024

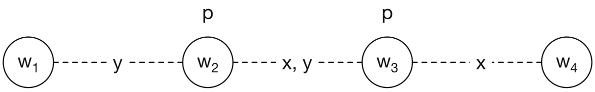

Fig. 1 Epistemic Model

tially, no longer publicly available. As a result, knowing the name of Mr Gonzalez no longer revealed the discrediting information about his past. But the story of Mr Gonzalez does not end here. His legal case became widely discussed in the media, which itself made enough information publicly available for people to be able to learn about the 1998 foreclosure of his house. Subsequently, Mr Gonzalez tried and failed to conceal the information about his legal case from public knowledge[1].

Google's publicising information about Mr Gonzalez's foreclosure, then removing it from the search results, and then publicising it again as the court case information are examples of public revelations and concealments that affect knowledge. In this article, we propose a dynamic epistemic logic that captures the properties of the interplay between knowledge and the public epistemic actions of revelation and concealment.

1.1 Data-Informed Knowledge

Traditionally, knowledge has been associated with individual agents. However, in the real world, just like in Mr Gonzalez's example, the agents rely more and more on access to data on servers and in databases rather than their own observations. In such a setting, it is more natural to associate knowledge with data rather than individual agents. In a previous work, we introduced the term *data-informed knowledge* to refer to such form of knowledge [24].

Formally, data-informed knowledge could be defined using epistemic models like the one depicted in Fig. 1. This model has four possible worlds: w_1, w_2, w_3, and w_4 and two data variables: x and y. Data variables can be Boolean, integer, strings, or any other type. We assume that each data variable has a value in each possible world. However, as we will see later, the specific values of data variables are not important for the definition of data-informed knowledge. It is only important in which worlds the values are different. We say that two worlds are *indistinguishable* by a data variable if the variable has the same value in both worlds. The indistinguishability relation for each data variable is shown in Fig. 1 using dashed lines. For example, data variable x in this epistemic model has the same value in worlds w_2, w_3, w_4 and a different value in world w_1.

Note that atomic proposition p is true in worlds w_2 and w_3 and is false in worlds w_1 and w_4, see Fig. 1. Recall that the value of data variable x in world w_1 is unique to this world. Thus, in world w_1, knowing just the value of x in the current world informs the knowledge that proposition p is false in the current world. We write this as:

$$w_1, \varnothing \Vdash \mathsf{K}_x \neg p.$$

[1] http://cyberlaw.stanford.edu/blog/2015/10/no-more-right-be-forgotten-mr-costeja-says-spanish-data-protection-authority

We will explain the meaning of the symbol $\varnothing$ in the above statement later. Observe that although proposition p is true in world w_2, it is not true in world w_4, which is indistinguishability from the world w_2 by data variable x. Thus, in world w_2 knowing just the value of x in the current world does not inform the knowledge that proposition p is true in the current world:

$$w_2, \varnothing \Vdash \neg \mathsf{K}_x p. \tag{1}$$

Imagine next that the value of data variable y is publicly revealed (announced). In this case, knowing the value of variable x alone in world w_2 informs the knowledge of p. Indeed, only worlds w_2 and w_3 are indistinguishability from world w_2 by data variable x and y, and in both of these worlds atomic proposition p is true. Thus, if the value of data variable y is publicly announced, then, in world w_2, knowing just the value of x in the current world informs the knowledge that proposition p is true in the current world:

$$w_2, \{y\} \Vdash \mathsf{K}_x p. \tag{2}$$

In general, we consider the satisfaction relation $w, U \Vdash \varphi$ where U is a finite set of data variables. We read it as "formula φ is satisfied in world w when the set of publicly revealed data variables is U". Using essentially the same argument as the one we use to justify statement (2), we can also show that

$$w_2, \{x\} \Vdash \mathsf{K}_y p.$$

Finally, note that even if no data variable is publicly revealed, then knowing the values of both data variables, x and y, in world w_2 informs that p is true in the current world:

$$w_2, \varnothing \Vdash \mathsf{K}_{x,y} p.$$

In this article, we consider modality K_X for an arbitrary *dataset* (a finite set of data variables) X. We read $\mathsf{K}_X\varphi$ as "dataset X informs the knowledge of statement φ". We also call K_X a "data-informed knowledge modality".

Grossi, Lorini, and Schwarzentruber proposed Ceteris Paribus Logic that describes the properties of modality K_X when X is a set of *Boolean* variables [21]. Baltag and van Benthem considered a modality similar but not identical to K_X [6]; we compare our work with theirs in Section 6. We previously proposed an axiomatisation of the interplay between modality K_X and a data-informed coalition power modality [24]. In particular, we observed that K_X is an S5 modality. We also introduced and axiomatised trust-based belief modality $\mathsf{B}^T_X\varphi$ that stands for "under the assumption of trustworthiness of dataset T, dataset X informs the belief in statement φ" [25]. The expression $\mathsf{B}^\varnothing_X\varphi$ is equivalent to $\mathsf{K}_X\varphi$.

1.2 Information Dynamics: Revelations and Concealments

One of the main contributions of this article is a sound and complete logical system that combines the data-informed knowledge modality K_X with modalities representing *epistemic events*. We consider two such modalities. The first of them represents public announcements of data. We call such announcements *revelations*. This modality was first proposed by van Eijck, Gattinger and Wang under the name "public inspection" [14]. For any dataset X and any formula φ, we write $[X]\varphi$ if formula φ becomes true after a public revelation of the values of all variables in dataset X. For example, consider world w_2 in which no values have been publicly announced yet, see Fig. 1. In this setting, after a public revelation of the value of variable y, knowing just the value of x in the current world informs the knowledge of the fact that proposition p is true in the current world:

$$w_2, \varnothing \Vdash [y]\mathsf{K}_x p.$$

This statement follows from statement (2). Baltag and van Benthem have observed that modality $[X]$ can be eliminated from the syntax of their logical system describing K_X-like modality [6]. In Section 6, we show that their result does not apply to our system. Axiomatisation of revelation modality $[X]$ is not considered in [24] and is only discussed as a possible future work in [25].

In this article, we propose a complementary modality of public *concealment*. We write $[X]^c\varphi$ if formula φ becomes true after the values of all variables in dataset X are removed from public access (concealed). For example, as it follows from statement (1),

$$w_2, \{y\} \Vdash [y]^c\neg\mathsf{K}_x p.$$

A real-world example of concealment is the removal of information (data) about Mr Gonzalez's foreclosure from the Google search results. We further analyse this example in Section 3.

Public data revelation modality $[X]\varphi$ has its roots in the Public Announcement Logic (PAL) [13, Chapter 4]. PAL extends the language of the epistemic logic with a public announcement modality $[\varphi]\psi$ which means "if truthful statement φ is publicly announced, then statement ψ will become true". Multiple extensions of Public Announcement Logic are suggested. Wáng and Ågotnes add to it the distributed knowledge modality [37].

Ågotnes, Balbiani, van Ditmarsch, and Seban propose a group announcement modality $\langle G\rangle\varphi$ that means "group G can announce certain facts, individually known to the members of the group, after which statement φ will be true" [1]. Although modality $\langle G\rangle\varphi$ states that φ will become true after an announcement by group G, it does not require φ to remain true after further announcements are made by agents outside of group G. However, such requirement is imposed by modality $\langle[G]\rangle\varphi$ introduced by Galimullin and Alechina [17]. Plaza introduces expression $\mathsf{Kv}_a(x)$ meaning that agent a knows the value of variable x [29, 30]. Wang and Fan propose a logical system that combines the public announcement modality with knowing data expressions [35]. In [36], they extend the language by the conditional knowledge expression $\mathsf{Kv}_a(\varphi, x)$ that stands for "agent a knows the value of x assuming φ is true".

Operation "forgetting of variables" has been considered in [27, 32] without interpreting it as a modality. The terms "revelation" and "concealment" that we use in this article were proposed by van der Hoek, Iliev, and Wooldridge in their work on *private* revelations and concealments [23]. In that paper, they consider modalities $[r(p, a)]\varphi$ and $[c(p, a)]\varphi$. The first (second) of them stands for "formula φ is true after the value of atomic proposition p is privately revealed to (concealed from) agent a". Our public revelation event is significantly different from their private revelation because public revelation creates common knowledge of the revealed values. Given how private concealments are defined in [23], our public concealments could be expressed through theirs by concealing the same data set from each agent. Naumov and Tao assumed that there is a cost associated with the concealment of each data variable [28]. They proposed a "hiding" modality $\mathsf{H}^q_A\varphi$ that means that coalition A can hide from public knowledge that statement φ is true by concealing some dataset at cost q. Unlike our modality $[X]^c\varphi$, their modality $\mathsf{H}^q_A\varphi$ does not explicitly mention the dataset being concealed.

1.3 Functional Dependency

One of the advantages of our data-centric approach to epistemology is that it allows extensions by data-specific operators that do not have agent-based equivalents. An important example of such an operator is Armstrong's functional dependency relation $X \rhd Y$ between two datasets [4]. Informally, $X \rhd Y$ means that the values of the variables in dataset X functionally determine the values of the variables in dataset Y. We also say that dataset X *informs* dataset Y. There are two ways in which "determine" could be interpreted: local and global. Recall that in the epistemic model depicted in Fig. 1, variable x has two values: one in world w_1 and another in worlds w_2, w_3, and w_4. Thus, in world w_1, the value of variable x uniquely determines the value of variable y (the one that y has in world w_1). At the same time, the value of x in world w_2 does not uniquely determine the value of y because worlds w_2 and w_4 have the same value of x, but different values of y. We refer to this as *local* dependency and write it as

$$
\begin{aligned}
&w_1, \varnothing \Vdash x \rhd y\\
&w_2, \varnothing \Vdash \neg(x \rhd y).
\end{aligned}
$$

Global dependency is local dependency in every world of the model. As we show in Section 2, global dependency is expressible through local dependency and data-informed knowledge modality.

Armstrong gave a sound and complete axiomatisation of this relation [4]. His axioms became known in database literature as Armstrong's axioms [18, p. 81]. Beeri, Fagin, and Howard [9] have suggested a variation of Armstrong's axioms that describe properties of multi-valued dependence. Baltag proposed a logical system for expression $X \rhd_a Y$, that stands for "agent a knows how to compute dataset Y based on dataset X" [5]. A connection between Armstrong axioms and strategies in imperfect information setting is discussed in [12].

Dependency $x \rhd y$ between single variables x and y could be expressed in the dynamic epistemic logic of "knowing the value" [14]. Namely, $x \rhd y$ is equivalent to $[x]\mathsf{Kv}(y)$, where $\mathsf{Kv}(y)$ stands for "variable y is publicly known". More generally, if X and Y are finite sets $\{x_1, \dots, x_m\}$ and $\{y_1, \dots, y_n\}$ respectively, then $X \rhd Y$ is equivalent to $[x_1]\dots[x_m]\bigwedge_{i\le n}\mathsf{Kv}(y_i)$. Another approach to dependency is proposed in Dependence Logic [34].

The functional dependency relation $\rhd$ is present in the logical system [24] that describes the interplay between K_X and data-informed strategic power. Baltag and van Benthem include it into the logical system that describes the properties of their K_X-like modality [6]. The language of Logic of Revelation and Concealment [23] does not include functional dependency, perhaps because it only deals with Boolean variables. This relation is also only discussed in the future work section of the paper on trust-based beliefs [25].

1.4 Contribution and Outline

As discussed in the previous subsections, data-informed knowledge, revelations, concealments, and functional dependencies have been studied before. We summarise these works in Fig. 2. The original contribution of this work is describing the interplay between these notions through a complete logical system and a model checking algorithm. The system contains new and non-trivial axioms and inference rules not present in the existing literature. The proof of completeness significantly modifies the existing proof techniques from the literature.

The article is organised as follows. In the section that follows, we define the syntax and give the formal semantics of our system. Then, we show how our opening Gonzalez example can be formalised in our setting. Section 4 shows that revelation and concealment modalities are not definable through each other. In Section 5, we list the axioms and inference rules. Then, in Section 6, we give a detailed comparison between our logical system and the Simple Logic of Functional Dependence [6]. We prove the soundness of our axioms and rules in Section 7. In Section 8, we show its completeness, and in Section 9, we discuss and analyse a model checking algorithm for this system. Section 10 discusses the applications of our system to reasoning about causality. Section 11 concludes.

2 Syntax and Semantics

In this section, we introduce the class of formal models and use it to give a semantics of our logical system. Throughout most of the article (until Section 9) we assume a fixed finite set of data variables V and a fixed set of atomic propositions.

	[23]	[14]	[6]	[25]	[24]	current
data-informed knowledge			✓	✓	✓	✓
revelation	✓	✓	✓			✓
concealment	✓					✓
functional dependency		✓	✓		✓	✓

Fig. 2 Concepts covered in closely related literature

Intuitively, we think about data variables as having "values". However, as we have seen in our introductory example depicted in Fig. 1, for us, it is only important to know in which worlds the values of any given variable are different. For this reason and to keep the presentation more succinct, we formally model data variables as equivalence relations on the set of epistemic worlds. Informally, two worlds are $\sim_x$-equivalent if the value of data variable x in both worlds is the same.

Definition 1 A triple $(W, \{\sim_x\}_{x\in V}, \pi)$ is called an epistemic model, when

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

In the epistemic model depicted in Fig. 1, atomic proposition p is true in worlds w_2 and w_3 and is false in worlds w_1 and w_4; the value of this proposition does not depend on the revelations made in the world. In other words, in this model, atomic proposition p captures a *property of the world*. In Definition 1, we have chosen a more general approach where atomic propositions could express a *property of the world and the public revelations made in this world*. For example, an atomic proposition can be the statement "it has already been publicly revealed that Gonzalez owned the house". To do this, we define $\pi(p)$ as a set of pairs (w, U), where w is a world and U is the set of all variables that have been publicly revealed. We further discuss this after Definition 2 and in Section 6.

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

$$\varphi := p \mid \neg\varphi \mid \varphi \to \varphi \mid \mathsf{K}_X\varphi \mid [X]\varphi \mid [X]^c\varphi \mid X \rhd Y,$$

where p is an atomic proposition and $X, Y \subseteq V$ are datasets. We read formula $\mathsf{K}_X\varphi$ as "dataset X informs the knowledge of statement φ", formula $[X]\varphi$ as "formula φ is true after a public revelation of dataset X", formula $[X]^c\varphi$ as "formula φ is true after a public concealment of dataset X", and $X \rhd Y$ as "the values of the variables in dataset X determine the values of the variables in dataset Y". We assume that conjunction, disjunction, and biconditional are defined in the standard way.

The definition below specifies the formal semantics of this language with respect to the epistemic models. We write $w \sim_X u$ if $w \sim_x u$ for each data variable $x \in X$.

Usually, the satisfaction relation is defined in modal logics as a relation between a world and a formula. In our case, it is a ternary relation between a world w, a dataset U, and a formula φ. Informally, it means that formula φ is true after dataset U has been publicly revealed in world w.

Definition 2 For any world $w \in W$, any dataset $U \subseteq V$, and any formula $\varphi \in \Phi$, satisfaction relation $w, U \Vdash \varphi$ is defined as follows

1. $w, U \Vdash p$ if $(w, U) \in \pi(p)$,
2. $w, U \Vdash \neg\varphi$ if $w, U \nVdash \varphi$,
3. $w, U \Vdash \varphi \to \psi$ if $w, U \nVdash \varphi$ or $w, U \Vdash \psi$,
4. $w, U \Vdash \mathsf{K}_X\varphi$, if $v, U \Vdash \varphi$ for each world $v \in W$ such that $w \sim_{U\cup X} v$,

5. $w, U \Vdash [X]\varphi$ if $w, U \cup X \Vdash \varphi$,
6. $w, U \Vdash [X]^c\varphi$ if $w, U \setminus X \Vdash \varphi$.
7. $w, U \Vdash X \rhd Y$ if $w \sim_Y v$ for each world $v \in W$ such that $w \sim_{U \cup X} v$.

Traditionally, in modal logic, the satisfaction relation is defined as a relation between a world and a formula. In this case, each modal formula expresses a property of worlds. In our case, the satisfaction relation is a ternary relation between a world, a dataset, and a formula. This means that modal formulae in our system express the properties definable in terms of the world and the public revelations made in this world. To make our logical system coherent (and closed with respect to substitution), we assume that atomic propositions capture the properties of the type. This is the reason for defining $\pi(p)$ to be a subset of $W \times \mathcal{P}(V)$, see Definition 1.

Modality K_X captures the knowledge informed by dataset X given that dataset U has already been publicly revealed. In other words, the knowledge comes from both datasets X and U. This explains the use of relation $\sim_{X \cup U}$ in item 4 of the above definition. Baltag and van Benthem considered a similar modality for a set of arbitrary variable [6]. We compare our approaches in Section 6.

Informally, $X \rhd Y$ means that the values of the variables in dataset X functionally determine the values of the variables in dataset Y. The term "determine" can be interpreted in two ways: globally (in each world) and locally (in the current world). Item 7 of Definition 2 gives $X \rhd Y$ the local interpretation. Recall that V is the set of all data variables. It is easy to see that $w, U \Vdash [V]^c\mathsf{K}_\varnothing[X]\varphi$ means that $w', X \Vdash \varphi$ for each world w' of the model. Thus, the global interpretation could be expressed as $[V]^c\mathsf{K}_\varnothing(X \rhd Y)$.

Note that the proposed logical system, in a sense, subsumes the standard epistemic logic of agent-based distributed knowledge. Indeed, the semantics of our data variables is defined through indistinguishability relations between worlds. To model an agent's knowledge, one can introduce a variable that represents "all that the agent knows". The indistinguishability relation of such a variable is exactly the indistinguishability relation of the agent. If X is a set of such variables corresponding to different agents, then the modality K_X represents the distributed knowledge of these agents. Under such an interpretation, operator $X \rhd Y$ means that group X distributively knows at least as much as group Y. Modality $[X]$ corresponds to a public announcement of *everything* that the group of agents X knows. The meaning of the modality $[X]^c$ is perhaps less intuitive.

3 Gonzalez Example

An epistemic model capturing our introductory example is depicted in Fig. 3. For the sake of this example, we assume that there are only two agents, Don Quixote (DQ) and Mario Costeja Gonzalez (MG), and four epistemic worlds: w_1, w_2, w_3, and w_4. In Fig. 3, the financial advisor (either DQ or MG) is marked with the euro sign. The former owner of the foreclosed house is marked with a key sign. The actual world is w_1 where MG is the financial advisor and the former owner.

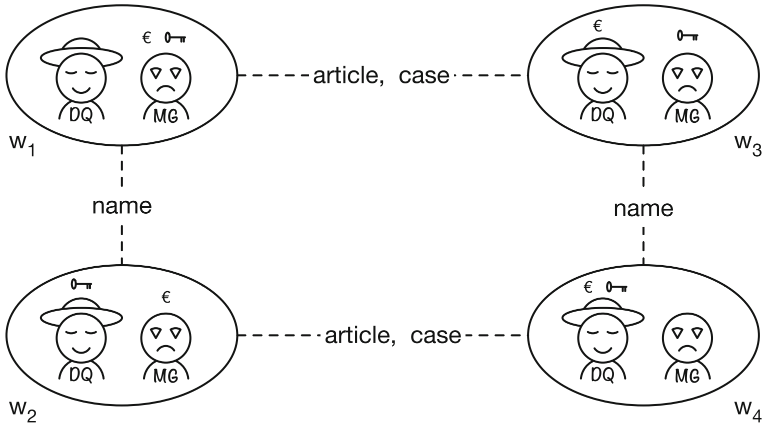

Fig. 3 The right to be forgotten epistemic model

We consider three data variables, "name", "article", and "case". Data variable "name" is the name of the financial advisor. Note that MG is the advisor in worlds w_1 and w_2. Thus, these two worlds are not distinguishable by this data variable. We denote this by a dashed line between worlds w_1 and w_2 labelled with the data variable "name". Data variable "article" is the content of the original article about the foreclosure, which contains the name of the former owner. Because MG is the former owner in worlds w_1 and w_3, the content of the article cannot be used to distinguish these two worlds. Data variable "case" is the content of the legal case brought up by the former owner to protect his privacy. The case also contains the name of the owner. Because MG is the owner in worlds w_1 and w_3, the content of the legal case also cannot be used to distinguish these two worlds.

Note that $w_1, \{\text{article}\} \Vdash$ "the advisor had foreclosure", see Fig. 3. Thus, by item 4 of Definition 2,

$$w_1, \{\text{article}\} \Vdash \mathsf{K}_{\text{name}}(\text{“the advisor had foreclosure”}) \tag{3}$$

because there is only one world, namely w_1 itself, which is indistinguishability by the dataset {article, name} from world w_1. In other words, once the content of the article had been revealed by Google, everyone who knew the name of the financial advisor knew that he had a foreclosure.

At the same time,

$$w_1, \varnothing \nVdash \mathsf{K}_{\text{name}}(\text{“the advisor had foreclosure”}) \tag{4}$$

by item 4 of Definition 2, because $w_2, \varnothing \nVdash$ "the advisor had foreclosure" and $w_1 \sim_{\text{name}} w_2$, see Fig. 3. Thus, by item 2 and item 6 of Definition 2,

$$w_1, \{\text{article}\} \Vdash [\text{article}]^c \neg \mathsf{K}_{\text{name}}(\text{“the advisor had foreclosure”}).$$

In other words, after the article was concealed by Google following the court order, the name of the advisor no longer informed the knowledge that his house had been foreclosed. Similarly,

$$w_1, \{\text{article}\} \Vdash [\text{article}]^c[\text{case}]\mathsf{K}_{\text{name}}(\text{“the advisor had foreclosure”}).$$

To say it the other way, after the case information became public, the name of the advisor again informed the knowledge that his house had been foreclosed.

4 Undefinability

In this section, we prove that modalities [ ] and $[\ ]^c$ are not definable through each other. Of course, the *operation* "concealment" cannot be defined through *operation* "revelation" because no matter what you reveal, you never conceal anything. However, this argument does not apply to *statements* containing these operations. Indeed, as an operation, subtraction cannot be defined through addition, but any statement that uses subtraction is equivalent to a statement that uses only addition. For example, the statement $x - y = z$ is equivalent to the statement $x = y + z$. In this section, we show that the situation is different for modalities [ ] and $[\ ]^c$. Each of these modalities can express something not expressible through the other modality.

Instead of a more traditional *bisimulation* technique for proving undefinability, we use a recently proposed *truth sets algebra technique* [26]. Unlike bisimulation, the new technique is using a single model to prove an undefinability result. Without loss of generality, assume that set of data variables V contains a single data variable x and the set of atomic propositions contains a single atomic proposition p.

Definition 3 For any given epistemic model, the truth set $[\![\varphi]\!]$ of a formula $\varphi \in \Phi$ is the set $\{(w, U) \mid w, U \Vdash \varphi\}$.

Definition 4 Formulae $\varphi, \psi \in \Phi$ are semantically equivalent if $[\![\varphi]\!] = [\![\psi]\!]$ for any epistemic model.

Note that Definition 4 can be rephrased as follows: formulae $\varphi, \psi \in \Phi$ are semantically equivalent when $w, U \Vdash \varphi$ iff $w, U \Vdash \psi$ for any world w of any epistemic model and any dataset $U \subseteq V$.

4.1 Undefinability of modality $[\ ]^c$ through modalities [ ] and K

To prove our first undefinability result, consider an epistemic model consisting of two worlds, w_1 and w_2, distinguishable by variable x. That is, $w_1 \nsim_x w_2$. Also, let $\pi(p) = \{(w_1, \varnothing), (w_1, \{x\}), (w_2, \{x\})\}$.

To make the proof more readable, we visualise the truth sets of various formulae using 2×2 diagrams as shown in Fig. 4. The rows of the diagram are labelled by worlds and the columns of the diagrams are labelled by datasets. In our model, there are only two worlds, w_1 and w_2, and two datasets, $\varnothing$ and $\{x\}$. If pair (w, U) belongs

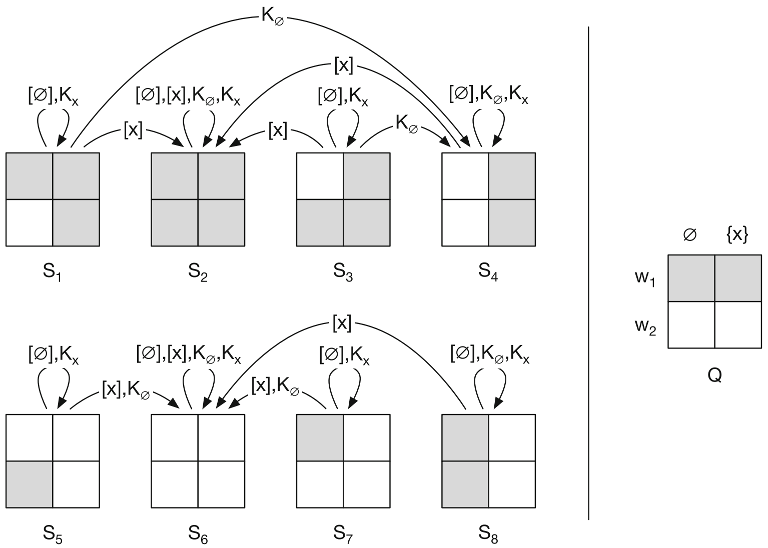

Fig. 4 Towards the proof of Theorem 1

to a truth set, then we grey out the square of the diagram that corresponds to this pair. In this proof, we will focus on the truth sets $S_1, \dots, S_8, Q$ visualised in Fig. 4.

Lemma 1 $[\![[\varnothing]\,\varphi]\!], [\![[x]\,\varphi]\!] \in \{S_1, \dots, S_8\}$ *for any formula* $\varphi \in \Phi$ *such that* $[\![\varphi]\!] \in \{S_1, \dots, S_8\}$.

Proof Suppose that $[\![\varphi]\!] = S_1$. Thus,

$$w_1, \varnothing \Vdash \varphi, \qquad\qquad w_1, \{x\} \Vdash \varphi, \tag{5}$$

$$w_2, \varnothing \nVdash \varphi, \qquad\qquad w_2, \{x\} \Vdash \varphi, \tag{6}$$

see truth set S_1 in Fig. 4. Hence, by item 5 of Definition 2,

$$w_1, \varnothing \Vdash [\varnothing]\varphi, \qquad\qquad w_1, \{x\} \Vdash [\varnothing]\varphi,$$

$$w_2, \varnothing \nVdash [\varnothing]\varphi, \qquad\qquad w_2, \{x\} \Vdash [\varnothing]\varphi.$$

Then, $[\![[\varnothing]\varphi]\!] = S_1$, see again truth set S_1 in Fig. 4. Thus, for any formula $\varphi \in \Phi$, if $[\![\varphi]\!] = S_1$, then $[\![[\varnothing]\,\varphi]\!] = S_1$. In Fig. 4, we visualise this result by the directed loop arrow from truth set S_1 back to truth set S_1 labelled with modality $[\varnothing]$.

Note also that parts $w_1, \{x\} \Vdash \varphi$ and $w_2, \{x\} \Vdash \varphi$ of statements (5) and (6), respectively, by item 5 of Definition 2, imply that

$$w_1, U \Vdash [x]\varphi \quad \text{and} \quad w_2, U \Vdash [x]\varphi.$$

for any dataset $U \in \{\varnothing, \{x\}\}$. Then, $[\![[x]\varphi]\!] = S_2$, see truth set S_2 in Fig. 4. In Fig. 4, we visualise this result by the directed arrow from truth set S_1 to truth set S_2 labelled with modality $[x]$.

Therefore, if $[\![\varphi]\!] = S_1$, then $[\![[\varnothing]\varphi]\!], [\![[x]\varphi]\!] \in \{S_1, \dots, S_8\}$ for any formula $\varphi \in \Phi$. The other seven cases are similar. We show the corresponding labelled directed arrows in Fig. 4. □

The proof of the next lemma is similar to the proof of Lemma 1, but it uses item 4 of Definition 2 instead of item 5.

Lemma 2 $[\![\mathsf{K}_\varnothing\varphi]\!], [\![\mathsf{K}_x\varphi]\!] \in \{S_1, \dots, S_8\}$ *for any formula* $\varphi \in \Phi$ *such that* $[\![\varphi]\!] \in \{S_1, \dots, S_8\}$.

For any set A of pairs (w, U), where w is a world and $U \subseteq V$ is a dataset, by $\overline{A}$, the complement of A, we denote the set of all pairs (w, U) that do not belong to set A.

Lemma 3 *The set* $\{S_1, \dots, S_8\}$ *is closed with respect to complement and union.*

Proof The statement of the lemma could be easily verified by considering all possible cases. For example, $\overline{S_1} = S_5$ and $S_5 \cup S_7 = S_8$, see Fig. 4. □

Lemma 4 $[\![\varphi]\!] \in \{S_1, \dots, S_8\}$ *for any formula* $\varphi \in \Phi$ *that does not use modality* $[\ ]^c$.

Proof We prove the lemma by induction on the structural complexity of formula φ. If φ is the atomic proposition p, then, by Definition 3 and Definition 2,

$$[\![\varphi]\!] = [\![p]\!] = \pi(p) = \{(w_1, \varnothing), (w_1, \{x\}), (w_2, \{x\})\} = S_1 \in \{S_1, \dots, S_8\}.$$

Next, suppose that formula φ has the form $X \rhd Y$, where $X, Y \subseteq \{x\}$. Observe that, $[\![\varnothing \rhd \varnothing]\!] = [\![x \rhd \varnothing]\!] = [\![x \rhd x]\!] = S_2$ by Definition 3 and Definition 2. Additionally, $[\![\varnothing \rhd x]\!] = S_4$ by the assumption $w_1 \nsim_x w_2$ and the same Definition 3 and Definition 2. Thus, $\varphi \in \{S_1, \dots, S_8\}$.

If formula φ is a negation or an implication, then the statement of the lemma follows from Lemma 3 and the induction hypothesis because $[\![\neg\psi]\!] = \overline{[\![\psi]\!]}$ and $[\![\psi \to \chi]\!] = \overline{[\![\psi]\!]} \cup [\![\chi]\!]$.

Finally, if formula φ has either the form $[X]\psi$ or $\mathsf{K}_X\psi$, then the statement of the lemma follows from the induction hypothesis and either Lemma 1 or Lemma 2, respectively. □

Lemma 5 $[\![[x]^c p]\!] \notin \{S_1, \dots, S_8\}$.

Proof By the choice of the epistemic model, item 6 of Definition 2, and Definition 3, the set $[\![[x]^c p]\!]$ is equal to set Q depicted on the right-most diagram in Fig. 4. □

The next theorem follows from the two lemmas before and Definition 4.

Theorem 1 (undefinability) *Formula* $[x]^c p$ *is not semantically equivalent to any formula in language* Φ *that does not use modality* $[\ ]^c$.

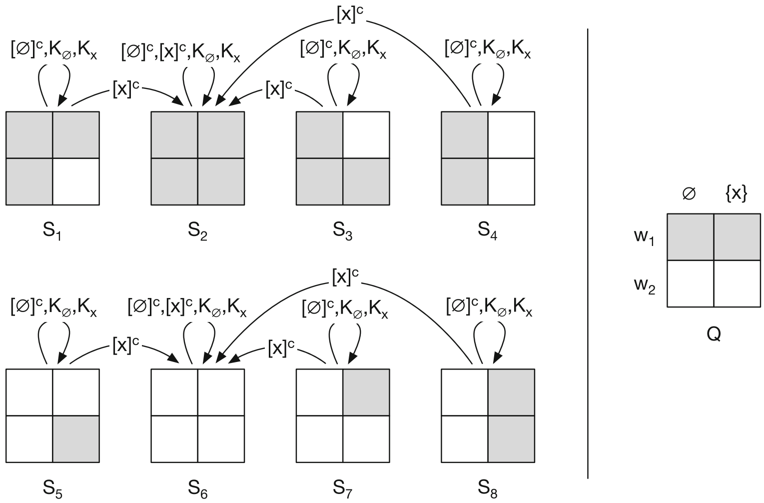

Fig. 5 Towards the proof of Theorem 2

4.2 Undefinability of [ ] through modalities $[\ ]^c$ and K

To prove that revelation modality cannot be defined through concealment modality, we consider a model like the one used in Section 4.1, except $\pi(p) = \{(w_1, \varnothing), (w_1, \{x\}), (w_2, \varnothing)\}$ and use truth sets as shown in Fig. 5. The diagram labelled as Q in Fig. 5 visualises the truth set of formula $[x]p$.

The proof of the following theorem is similar to the proof for Theorem 1, but we use Fig. 5 instead of Fig. 4.

Theorem 2 (undefinability) *Formula $[x]p$ is not semantically equivalent to any formula in language Φ that does not use modality* [ ].

5 Axioms

In addition to propositional tautologies in language Φ, our logical system contains the following axioms.

1. Reflexivity: $X \rhd Y$, where $Y \subseteq X$,
2. Augmentation: $X \rhd Y \to (X \cup Z \rhd Y \cup Z)$,
3. Transitivity: $X \rhd Y \to (Y \rhd Z \to X \rhd Z)$,
4. Truth: $\mathsf{K}_X\varphi \to \varphi$,
5. Negative Introspection: $\neg\mathsf{K}_X\varphi \to \mathsf{K}_X\neg\mathsf{K}_X\varphi$,
6. Distributivity: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, where $\Box \in \{\mathsf{K}_X, [X], [X]^c\}$,
7. Monotonicity: $X \rhd Y \to (\mathsf{K}_Y\varphi \to \mathsf{K}_X\varphi)$,
8. Introspection of Dependency: $X \rhd Y \to \mathsf{K}_X(X \rhd Y)$,
9. Partial Revelation: $(X \cup Y) \rhd Z \leftrightarrow [X](Y \rhd Z)$,
10. Empty Dataset: $\Box\varphi \leftrightarrow \varphi$, where $\Box \in \{[\varnothing], [\varnothing]^c\}$,

11. Negation: $\neg\Box\varphi \leftrightarrow \Box\neg\varphi$, where $\Box \in \{[X], [X]^c\}$,
12. Composition: $[X][Y]\varphi \leftrightarrow [X \cup Y]\varphi$ and $[X]^c[Y]^c\varphi \leftrightarrow [X \cup Y]^c\varphi$,
13. Inversion: $[X][X]^c\varphi \leftrightarrow [X]^c\varphi$ and $[X]^c[X]\varphi \leftrightarrow [X]\varphi$,
14. Commutativity: $\Box_X\Box_Y\varphi \leftrightarrow \Box_Y\Box_X\varphi$, if sets X and Y are disjoint and $\Box_X \in \{[X], [X]^c\}$, $\Box_Y \in \{[Y], [Y]^c\}$,
15. Epistemic Commutativity: $\mathsf{K}_{X\cup Y}[Y]\varphi \leftrightarrow [Y]\mathsf{K}_X\varphi$,
$\mathsf{K}_X[Y]^c\varphi \to [Y]^c\mathsf{K}_{X\cup Y}\varphi$, and $[Y]^c\mathsf{K}_X\varphi \to \mathsf{K}_X[Y]^c\varphi$,
16. Powerset: $\bigwedge_{U\subseteq V}[V]^c[U]\varphi \to \varphi$.

The Reflexivity, Augmentation, and Transitivity axioms are the standard axioms of functional dependency [4]. The Truth and the Negative Introspection axioms as well as the K_X part of the Distributivity axiom are the standard principles of S5 epistemic logic [15] restated for data-informed knowledge. The Distributivity axioms for modalities $[X]$ and $[X]^c$ are straightforward.

The Monotonicity axiom states that if, in the current world, the values of variables in dataset X determine those in dataset Y and the values of the variables in dataset Y inform the knowledge of φ, then those in dataset X also inform the knowledge of φ.

Recall that $X \rhd Y$ denotes *local* functional dependency. In other words, it means that the values of the variables in dataset X *in the current world* determine the values of the variables in the dataset Y. In general, $X \rhd Y$ does not imply that X determines Y in each world. However, it does imply that X determines Y in each world where all variables in dataset X have the same values as in the current world. Hence, $X \rhd Y$ implies that X determines Y in each X-indistinguishability world. This property is captured in the Introspection of Dependency axiom.

The Partial Revelation axiom has two parts. The first of them states that if the knowledge of the values of the variables in dataset $X \cup Y$ is enough to determine the values of the variables in dataset Z, then after a public revelation of dataset X, the knowledge of the values of the variables in dataset Y alone is enough to determine the values of the variables in dataset Z. The second part states the converse.

The Empty Dataset axiom states that a public revelation or concealment of an empty dataset does not affect the validity of any formula.

The Negation axiom states that a formula does not hold after a revelation (concealment) if and only if the negation of the formula holds after the revelation (concealment). It captures the fact that revelations and concealments are deterministic epistemic actions.

The Composition axiom states that consecutive revelations (concealment) of two datasets are equivalent to a single revelation (concealment) of the union of these datasets.

Intuitively, revelation and concealment are "opposite" operations. However, they are not opposite from the algebraic point of view. The connection between them in the modal language is captured by the Inversion axiom. In algebraic terms, these operations form an *inverse semigroup* [11, p.40] with respect to composition.

The Commutativity axiom states that revelations and concealments of *disjoint* datasets could be done in any order. The Epistemic Commutativity axiom describes the laws of commutativity between the knowledge modality and the public revelation and concealment modalities.

The Powerset axiom captures the fact that V is the set of all data variables in an epistemic model. Informally, it states that concealing the whole set V and then revealing a dataset U results in φ being true no matter what dataset $U \subseteq V$ is chosen, then φ should have been true to start with. This axiom is true because the set of currently revealed variables is one of such datasets U.

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

$$\frac{\varphi, \varphi \to \psi}{\psi} \qquad \frac{\varphi}{\mathsf{K}_X \varphi} \qquad \frac{\varphi}{[X]\varphi} \qquad \frac{\varphi}{[X]^c \varphi}$$

and the Concealed Monotonicity:

$$\frac{[X]^c\varphi \to [X]^c\psi}{[X]^c\mathsf{K}_Y\varphi \to [X]^c\mathsf{K}_Y\psi}$$

inference rules.

The Concealed Monotonicity inference rule describes an aspect of the interplay between modalities K_X and $[X]^c$. To the best of our knowledge, this rule is not derivable from the Epistemic Commutativity axiom.

In addition to the unary relation $\vdash \varphi$, we also consider a binary relation $F \vdash \varphi$ between a set of formulae $F \subseteq \Phi$ and a formula $\varphi \in \Phi$. Statement $F \vdash \varphi$ is true if formula φ is provable from the *theorems* of our logical system and the set of additional assumption F using the Modus Ponens inference rule *only*. It is easy to see that $\varnothing \vdash \varphi$ iff $\vdash \varphi$.

6 Comparison to the Simple Logic of Functional Dependence

As one can easily observe from Fig. 2, Baltag and van Benthem's Simple Logic of Functional Dependence [6] is the closest logical system to the one presented in the current article. The fundamental difference between the two works is that our system contains concealment modality $[X]^c$ while their system does not. However, there are also multiple technical choices made by them and us differently, which lead to a significant difference in semantics, the axioms, and the properties of these two systems. In this section, we discuss some of these choices and their consequences.

At the beginning of Section 2, we observed that there is no need to have explicit values assigned to data variables in different worlds. Thus, following [24, 25], we represent data variables as equivalence relations on the worlds. Driven perhaps by the same minimalist's desire, Baltag and van Benthem decided to keep the values of the variables in the semantics, but to eliminate the worlds. Assuming that the set of all data variables V is $\{x_1, \ldots, x_n\}$, the satisfaction relation in their system is defined as a relation $v_1, \ldots, v_n \Vdash \varphi$ between values $v_1, \ldots, v_n$ of the variables $x_1, \ldots, x_n$ and a formula φ. In essence, in their setting, a tuple of values $v_1, \ldots, v_n$ plays the role of a world. To avoid trivialisation, they assume that not all combinations of values, generally speaking, are possible. It's an elegant approach, but this approach eliminates the possibility of having multiple "worlds" with the same values of all data variables. As a result, the data-informed knowledge modality K_X in their system has some non-S5 properties. For example, the formula $\varphi \to \mathsf{K}_V\varphi$ is universally true under their semantics. In our system, modality K_X is an S5-modality.

In the Simple Logic of Functional Dependence, each formula can be transformed into an equivalent form that does not contain public revelation modality $[X]$. A similar property holds for the original Public Announcement Logic (PAL) [13, Chapter 4], but it is not true for our logical system. To understand why such transformation does not work for our system, let's first recall how this transformation works for formulae in [6]. The transformation is based on the following equivalences valid in their logic:

$$[X]p \equiv p, \tag{7}$$

$$[X](Y \rhd Z) \equiv (X \cup Y) \rhd Z, \tag{8}$$

$$[X]\neg\varphi \equiv \neg[X]\varphi, \tag{9}$$

$$[X](\varphi \to \psi) \equiv [X]\varphi \to [X]\psi, \tag{10}$$

$$[X]\mathsf{K}_Y\varphi \equiv \mathsf{K}_{X\cup Y}[X]\varphi. \tag{11}$$

Equivalences (9) through (11) could be used to "push" the revelation modality $[X]$ to the atomic level and equivalences (7) and (8) to eliminate this modality on the atomic level.

There are two issues that prevent applying this technique to our system. First, equivalence (7) is not valid in our system. Indeed, recall from our discussion after Definition 1 that atomic propositions in our system capture properties of a world *and the public revelations made in this world*. Thus, the validity of an atomic proposition can change after a public revelation. This is not true in [6], where atomic propositions capture properties of just worlds (or, to be exact, tuples of values $v_1, \ldots, v_n$).

In classical PAL [13, Chapter 4], atomic propositions also capture properties of worlds. So, a PAL traditionalist might argue that this is a problem of our own making. We should just make the atomic propositions to be about worlds, and this will ensure that the transformation works in our system. This is not true because of the second issue: the transformation does not work for the public concealment modality $[X]^c$. Namely, to the best of our knowledge, an equivalence-like statement (11) does not exist for the concealment modality. There are two distinct versions of the Epistemic Commutativity axiom for concealment: $\mathsf{K}_X[Y]^c\varphi \to [Y]^c\mathsf{K}_{X\cup Y}\varphi$ and $[Y]^c\mathsf{K}_X\varphi \to \mathsf{K}_X[Y]^c\varphi$, but neither of them is an equivalence statement. The lack of such an equivalence prevents the elimination of the concealment modality in our system. It also highlights the significant difference between [6] and the present work.

7 Soundness

In this section, we show the soundness of our logical system, which is stated as Theorem 3 at the end of this section. The soundness of the Truth, Negative Introspection, Monotonicity, Distributivity, Empty Dataset, Composition, and Commutativity axioms is straightforward. Below we prove the soundness of the remaining axioms as separate lemmas.

Lemma 6 $w, U \nVdash [X]\varphi$ *iff* $w, U \Vdash [X]\neg\varphi$.

Proof By item 5 of Definition 2, the statement $w, U \nVdash [X]\varphi$ is equivalent to $w, U \cup X \nVdash \varphi$. The latter statement is equivalent to $w, U \cup X \Vdash \neg\varphi$ by item 2 of Definition 2. In turn, the statement $w, U \cup X \Vdash \neg\varphi$ is equivalent to $w, U \Vdash [X]\neg\varphi$ again by item 5 of Definition 2. □

The proof of the next lemma is similar to the proof of the previous one except that it uses item 6 of Definition 2 instead of item 5.

Lemma 7 $w, U \nVdash [X]^c\varphi$ *iff* $w, U \Vdash [X]^c\neg\varphi$.

Lemma 8 $w, U \Vdash [X][X]^c\varphi$ *iff* $w, U \Vdash [X]^c\varphi$.

Proof The statement $w, U \Vdash [X][X]^c\varphi$ is equivalent to $w, U \cup X \Vdash [X]^c\varphi$ by item 5 of Definition 2. The latter statement is equivalent to $w, (U \cup X) \setminus X \Vdash \varphi$ by item 6 of Definition 2. In turn, the statement $w, (U \cup X) \setminus X \Vdash \varphi$ is equivalent to $w, U \setminus X \Vdash \varphi$ because $(U \cup X) \setminus X \equiv U \setminus X$. Finally, statement $w, U \setminus X \Vdash \varphi$ is equivalent to $w, U \Vdash [X]^c\varphi$ by item 6 of Definition 2. □

The proof of the next lemma is similar to the proof of the previous one except that it uses the set equivalence $(U \setminus X) \cup X \equiv U \cup X$ instead of $(U \cup X) \setminus X \equiv U \setminus X$.

Lemma 9 $w, U \Vdash [X]^c[X]\varphi$ *iff* $w, U \Vdash [X]\varphi$.

Lemma 10 $w, U \Vdash \mathsf{K}_{X\cup Y}[Y]\varphi$ *iff* $w, U \Vdash [Y]\mathsf{K}_X\varphi$.

Proof $(\Rightarrow)$: Suppose $w, U \nVdash [Y]\mathsf{K}_X\varphi$. Thus, it follows that $w, U \cup Y \nVdash \mathsf{K}_X\varphi$ by item 5 of Definition 2. Hence, by item 4 of Definition 2, there exists a world $v \in W$ such that $w \sim_{U\cup Y\cup X} v$ and $v, U \cup Y \nVdash \varphi$. Then, $v, U \nVdash [Y]\varphi$ by item 5 of Definition 2. Therefore, $w, U \nVdash \mathsf{K}_{X\cup Y}[Y]\varphi$ by item 4 of Definition 2.

$(\Leftarrow)$: Assume that $w, U \nVdash \mathsf{K}_{X\cup Y}[Y]\varphi$. Thus, by item 4 of Definition 2, there exists a world $v \in W$ such that $w \sim_{U\cup X\cup Y} v$ and $v, U \nVdash [Y]\varphi$. Hence, $v, U \cup Y \nVdash \varphi$ by item 5 of Definition 2. Then, $w, U \cup Y \nVdash \mathsf{K}_X\varphi$ by item 4 of Definition 2. Therefore, $w, U \nVdash [Y]\mathsf{K}_X\varphi$ again by item 5 of Definition 2. □

Lemma 11 *If* $w, U \Vdash \mathsf{K}_X[Y]^c\varphi$*, then* $w, U \Vdash [Y]^c\mathsf{K}_{X\cup Y}\varphi$.

Proof Let $w, U \nVdash [Y]^c\mathsf{K}_{X\cup Y}\varphi$. Thus, $w, U \setminus Y \nVdash \mathsf{K}_{X\cup Y}\varphi$ by item 6 of Definition 2. Hence, by item 4 of Definition 2, there exists a world $v \in W$ such that $w \sim_{(U\setminus Y)\cup X\cup Y} v$ and $v, U \setminus Y \nVdash \varphi$. Then, $v, U \nVdash [Y]^c\varphi$ by item 6 of Definition 2. Note that $U \cup X \subseteq (U \setminus Y) \cup X \cup Y$. Thus, the statement $w \sim_{(U\setminus Y)\cup X\cup Y} v$ implies that $w \sim_{U\cup X} v$. Hence, $w, U \nVdash \mathsf{K}_X[Y]^c\varphi$ by item 4 of Definition 2. □

Lemma 12 *If* $w, U \Vdash [Y]^c\mathsf{K}_X\varphi$*, then* $w, U \Vdash \mathsf{K}_X[Y]^c\varphi$.

Proof Suppose that $w, U \nVdash \mathsf{K}_X[Y]^c\varphi$. Thus, by item 4 of Definition 2, there is a world $v \in W$ such that $w \sim_{U\cup X} v$ and $v, U \nVdash [Y]^c\varphi$. Then, $w \sim_{(U\setminus Y)\cup X} v$ because $(U \setminus Y) \cup X \subseteq U \cup X$. Also, $v, U \setminus Y \nVdash \varphi$ by item 6 of Definition 2. Hence, $w, U\setminus Y \nVdash \mathsf{K}_X\varphi$ by item 4 of Definition 2. Therefore, $w, U \nVdash [Y]^c\mathsf{K}_X\varphi$ by Definition 2. □

Lemma 13 *If $w, U \Vdash [V]^c[U']\varphi$ for each dataset $U' \subseteq V$, then $w, U \Vdash \varphi$.*

Proof The assumption of the lemma implies, in particular, that $w, U \Vdash [V]^c[U]\varphi$. Thus, $w, U \setminus V \Vdash [U]\varphi$ by item 6 of Definition 2. Hence, $w, \varnothing \Vdash [U]\varphi$ because $U \subseteq V$. Therefore, $w, U \Vdash \varphi$ by item 5 of Definition 2. □

The soundness of the Modus Ponens and of the Necessitation axioms is straightforward. In the next lemma, we prove the soundness of the Concealed Monotonicity inference rule.

Lemma 14 *If $w, U \Vdash [X]^c\varphi \to [X]^c\psi$ for each world $w \in W$ of each epistemic model and each dataset $U \subseteq V$, then*

$$w, U \Vdash [X]^c\mathsf{K}_Y\varphi \to [X]^c\mathsf{K}_Y\psi$$

for each world $w \in W$ of each epistemic model and any dataset $U \subseteq V$.

Proof Consider any world $w \in W$ of an epistemic model and any dataset $U \subseteq V$. Suppose that

$$w, U \Vdash [X]^c\mathsf{K}_Y\varphi. \tag{12}$$

It suffices to show that $w, U \Vdash [X]^c\mathsf{K}_Y\psi$.

Suppose that $w, U \nVdash [X]^c\mathsf{K}_Y\psi$. Thus, $w, U \setminus X \nVdash \mathsf{K}_Y\psi$ by item 6 of Definition 2. Hence, by item 4 of Definition 2, there exists a world $v \in W$ such that

$$w \sim_{(U\setminus X)\cup Y} v \tag{13}$$

and $v, U \setminus X \nVdash \psi$. Then, $v, U \nVdash [X]^c\psi$ by item 6 of Definition 2. Thus, $v, U \nVdash [X]^c\varphi$ by the assumption of the lemma. Hence, $v, U \setminus X \nVdash \varphi$ by item 6 of Definition 2. Then, $w, U \setminus X \nVdash \mathsf{K}_Y\varphi$ by the assumption (13) and item 4 of Definition 2. Therefore, $w, U \nVdash [X]^c\mathsf{K}_Y\varphi$ by item 6 of Definition 2, which contradicts assumption (12). □

The next strong soundness theorem follows from the lemmas above.

Theorem 3 (strong soundness) *For any set of formulae $F \subseteq \Phi$, any formula φ, any world w of an epistemic model, and any dataset $U \subseteq V$, if $w, U \Vdash f$ for each formula $f \in F$ and $F \vdash \varphi$, then $w, U \Vdash \varphi$.*

8 Completeness

In this section, we prove the completeness of our logical system using the canonical model construction. The proof is divided into five subsections. First, we state several auxiliary lemmas. Next, we define a dataset closure operation and prove its basic properties. After that, we specify a canonical model $M(F_0)$ for any fixed maximal consistent set of formulae F_0. Then, we establish several properties of this model required for the proof of completeness. As usual in modal logic, the proof of completeness relies on a "truth" lemma. In the fifth subsection, we prove this lemma and derive the completeness theorem from it. We highlight the novel parts of the completeness proof as they are being introduced.

8.1 Auxiliary Lemmas

In this subsection, we state auxiliary results that are used in the proof of completeness. The first four statements are well-known properties whose proofs we omit.

Lemma 15 (Deduction) *If $F, \varphi \vdash \psi$, then $F \vdash \varphi \to \psi$.*

Lemma 16 $\vdash \mathsf{K}_X\varphi \to \mathsf{K}_X\mathsf{K}_X\varphi$.

Lemma 17 *For any formulae $\varphi_1, \dots, \varphi_n, \psi \in \Phi$, if $\varphi_1, \dots, \varphi_n \vdash \psi$, then $\mathsf{K}_Y\varphi_1, .., \mathsf{K}_Y\varphi_n \vdash \mathsf{K}_Y\psi$, $[X]\varphi_1, .., [X]\varphi_n \vdash [X]\psi$, and $[X]^c\varphi_1, .., [X]^c\varphi_n \vdash [X]^c\psi$.*

Lemma 18 (Lindenbaum) *Any consistent set of formulae can be extended to a maximal consistent set of formulae.*

Lemma 19 *For any datasets $X, Y \subseteq V$, and any formulae $\varphi_1, \dots, \varphi_n, \psi \in \Phi$, if $[X]^c\varphi_1, \dots, [X]^c\varphi_n \vdash [X]^c\psi$, then $[X]^c\mathsf{K}_Y\varphi_1, \dots, [X]^c\mathsf{K}_Y\varphi_n \vdash [X]^c\mathsf{K}_Y\psi$.*

Proof By propositional reasoning, $\bigwedge_i \varphi_i \vdash \varphi_j$ for each $j \le n$. Thus, by Lemma 17,

$$[X]^c \bigwedge_i \varphi_i \vdash [X]^c\varphi_j$$

for each $j \le n$. Hence, by the assumption $[X]^c\varphi_1, \dots, [X]^c\varphi_n \vdash [X]^c\psi$ of the lemma, $[X]^c \bigwedge_i \varphi_i \vdash [X]^c\psi$. Then, $\vdash [X]^c \bigwedge_i \varphi_i \to [X]^c\psi$ by Lemma 15. Thus, by the Concealed Monotonicity inference rule,

$$\vdash [X]^c\mathsf{K}_Y \bigwedge_i \varphi_i \to [X]^c\mathsf{K}_Y\psi. \tag{14}$$

At the same time, $\varphi_1, \dots, \varphi_n \vdash \bigwedge_i \varphi_i$ by the laws of propositional reasoning. Hence, $\mathsf{K}_Y\varphi_1, \dots, \mathsf{K}_Y\varphi_n \vdash \mathsf{K}_Y \bigwedge_i \varphi_i$ by Lemma 17. Then, again by Lemma 17, it follows that $[X]^c\mathsf{K}_Y\varphi_1, \dots, [X]^c\mathsf{K}_Y\varphi_n \vdash [X]^c\mathsf{K}_Y \bigwedge_i \varphi_i$. Therefore,

$$[X]^c\mathsf{K}_Y\varphi_1, \dots, [X]^c\mathsf{K}_Y\varphi_n \vdash [X]^c\mathsf{K}_Y\psi,$$

by statement (14) and the Modus Ponens inference rule. □

The next lemma follows from the Necessitation inference rule and the Distributivity axiom in the standard way.

Lemma 20 *For any formulae* $\varphi, \psi \in \Phi$ *and any* $X \subseteq V$*, if* $\vdash \varphi \to \psi$*, then* $\vdash [X]\varphi \to [X]\psi$ *and* $\vdash [X]^c\varphi \to [X]^c\psi$.

Corollary 1 *For any formulae* $\varphi, \psi \in \Phi$ *and any* $X \subseteq V$*, if* $\vdash \varphi \leftrightarrow \psi$*, then* $\vdash [X]\varphi \leftrightarrow [X]\psi$*,* $\vdash [X]^c\varphi \leftrightarrow [X]^c\psi$*, and* $\vdash [X]^c[Y]\varphi \leftrightarrow [X]^c[Y]\psi$.

Lemma 21 $\vdash [X](\varphi \to \psi) \leftrightarrow ([X]\varphi \to [X]\psi)$,
$\vdash [X]^c(\varphi \to \psi) \leftrightarrow ([X]^c\varphi \to [X]^c\psi)$*, and*
$\vdash [X]^c[Y](\varphi \to \psi) \leftrightarrow ([X]^c[Y]\varphi \to [X]^c[Y]\psi)$.

Proof Note that the formulae $\neg\varphi \to (\varphi \to \psi)$ and $\psi \to (\varphi \to \psi)$ are propositional tautologies. Hence, by the Necessitation inference rule,

$$\vdash [X](\neg\varphi \to (\varphi \to \psi)) \quad \text{and} \quad \vdash [X](\psi \to (\varphi \to \psi)).$$

Thus, $\vdash [X]\neg\varphi \to [X](\varphi \to \psi)$ and $\vdash [X]\psi \to [X](\varphi \to \psi)$ by the Distributivity axiom and the Modus Ponens. Then, $\vdash ([X]\neg\varphi \vee [X]\psi) \to [X](\varphi \to \psi)$ by the laws of propositional reasoning. Hence, by the Negation axiom and propositional reasoning, $\vdash (\neg[X]\varphi \vee [X]\psi) \to [X](\varphi \to \psi)$. Thus, again by propositional reasoning, $\vdash ([X]\varphi \to [X]\psi) \to [X](\varphi \to \psi)$. Therefore, by the Distributivity axiom and even more propositional reasoning,

$$\vdash [X](\varphi \to \psi) \leftrightarrow ([X]\varphi \to [X]\psi).$$

The proof of the second part of the lemma is similar. The third part follows from the first two parts by Corollary 1. □

Lemma 22 $\vdash [V]^c[U \setminus X]\psi \leftrightarrow [V]^c[U][X]^c\psi$.

Proof Note that $V = V \cup X$ because $X \subseteq V$. Hence, the statement

$$[V]^c[U \setminus X]\psi \leftrightarrow [V \cup X]^c[U \setminus X]\psi \tag{15}$$

is a tautology. At the same time, by the Composition axiom,

$$\vdash [V \cup X]^c[U \setminus X]\psi \leftrightarrow [V]^c[X]^c[U \setminus X]\psi. \tag{16}$$

Also, by the Commutativity axiom, Corollary 1, and because sets X and $U \setminus X$ are disjoint,

$$\vdash [V]^c[X]^c[U \setminus X]\psi \leftrightarrow [V]^c[U \setminus X][X]^c\psi. \tag{17}$$

In addition, because $X = (U \cap X) \cup (X \setminus U)$, the statement

$$[V]^c[U \setminus X][X]^c\psi \leftrightarrow [V]^c[U \setminus X][(U \cap X) \cup (X \setminus U)]^c\psi \tag{18}$$

is also a tautology. At the same time, by the Composition axiom and Corollary 1,

$$\vdash [V]^c[U \setminus X][(U \cap X) \cup (X \setminus U)]^c\psi \leftrightarrow [V]^c[U \setminus X][U \cap X]^c[X \setminus U]^c\psi. \tag{19}$$

Also, by the Inversion axiom and Corollary 1,

$$\vdash [V]^c[U \setminus X][U \cap X]^c[X \setminus U]^c\psi \\ \leftrightarrow [V]^c[U \setminus X][U \cap X][U \cap X]^c[X \setminus U]^c\psi. \quad (20)$$

In addition, because $(U \setminus X) \cup (U \cap X) = U$, by the Composition axiom and Corollary 1,

$$\vdash [V]^c[U \setminus X][U \cap X][U \cap X]^c[X \setminus U]^c\psi \\ \leftrightarrow [V]^c[U][U \cap X]^c[X \setminus U]^c\psi. \quad (21)$$

Note also that $(U \cap X) \cup (X \setminus U) = X$. Thus, by the Composition axiom and Corollary 1,

$$\vdash [V]^c[U][U \cap X]^c[X \setminus U]^c\psi \leftrightarrow [V]^c[U][X]^c\psi. \quad (22)$$

Finally, observe that the statement of the lemma follows from statements (15), (16), (17), (18), (19), (20), (21), and (22) by the laws of propositional reasoning. ☐

8.2 Dataset Closure

An important idea used in our proof of completeness is "dataset closure". Informally, for each set of formulae F and each dataset X, by closure X^*_F we denote the set of all data variables about which set F can prove that they are informed by set X. This notion goes back to "saturated" sets in Armstrong's article on functional dependency [4, Section 6]. However, the prefix $[V]^c$ and set F in the definition below are original to the current article. Closures are used in Definition 7 of the next subsection to specify the labels of the edges of a tree.

Definition 5 $X^*_F = \{x \in V \mid [V]^c(X \rhd x) \in F\}$ for any dataset $X \subseteq V$ and any maximal consistent set of formulae $F \subseteq \Phi$.

In other words, the closure X^*_F is the set of all data variables that, according to set F, are functionally determined by dataset X. Intuitively, such a set must include variables from the dataset X itself. Next, we formally prove this.

Lemma 23 $X \subseteq X^*_F$.

Proof Consider any data variable $x \in X$. Thus, $\vdash X \rhd x$ by the Reflexivity axiom. Hence, $\vdash [V]^c(X \rhd x)$ by the Necessitation inference rule. Then, $[V]^c(X \rhd x) \in F$ because F is a maximal consistent set of formulae. Therefore, $x \in X^*_F$ by Definition 5. ☐

Note that $[V]^c(X \rhd x) \in F$ for each data variable $x \in X^*_F$ by Definition 5. The next lemma shows that all such variables x could be brought together on the right-hand-side of $\rhd$ expression.

Lemma 24 $F \vdash [V]^c(X \rhd X^*_F)$.

Proof The set X^*_F is finite by Definition 5 and the assumption in Section 2 that set V is finite. Let X^*_F be set $\{x_1, \dots, x_n\}$. Note that $F \vdash [V]^c(X \rhd x_i)$ for each $i \le n$ by Definition 5. We prove by induction for each integer k such that $0 \le k \le n$ that $F \vdash [V]^c(X \rhd x_1, \dots, x_k)$.

Base Case: $F \vdash X \rhd \varnothing$ by the Reflexivity axiom. Hence, $F \vdash [V]^c(X \rhd \varnothing)$ by the Necessitation inference rule.

Induction Step: By the Augmentation axiom,

$$\vdash X \rhd x_1, \dots, x_k \to X \cup \{x_{k+1}\} \rhd x_1, \dots, x_k, x_{k+1}.$$

and

$$\vdash X \rhd x_{k+1} \to X \rhd X \cup \{x_{k+1}\}.$$

Thus, by the Transitivity axiom and the laws of propositional reasoning,

$$\vdash X \rhd x_{k+1} \to (X \rhd x_1, \dots, x_k \to X \rhd x_1, \dots, x_k, x_{k+1}).$$

Hence, by the Necessitation inference rule,

$$\vdash [V]^c(X \rhd x_{k+1} \to (X \rhd x_1, \dots, x_k \to X \rhd x_1, \dots, x_{k+1})).$$

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

$$\vdash [V]^c(X \rhd x_{k+1}) \to [V]^c(X \rhd x_1, \dots, x_k \to X \rhd x_1, \dots, x_k, x_{k+1}).$$

Recall that $x_{k+1} \in X^*_F$. Thus, $F \vdash [V]^c(X \rhd x_{k+1})$ by Definition 5. Hence, by the Modus Ponens inference rule,

$$F \vdash [V]^c(X \rhd x_1, \dots, x_k \to X \rhd x_1, \dots, x_k, x_{k+1}).$$

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

$$F \vdash [V]^c(X \rhd x_1, \dots, x_k) \to [V]^c(X \rhd x_1, \dots, x_k, x_{k+1}).$$

Note that $F \vdash [V]^c(X \rhd x_1, \dots, x_k)$ by the induction hypothesis. Therefore, by the Modus Ponens inference rule, $F \vdash [V]^c(X \rhd x_1, \dots, x_k, x_{k+1})$. $\square$

8.3 Canonical Model

In this subsection, for any maximal consistent set of formulae $F_0 \subseteq \Phi$, we define the canonical epistemic model $M(F_0) = (W, \{\sim_x\}_{x \in V}, \pi)$. We define this model using the tree construction that has been previously used in completeness proofs involving distributed knowledge [16]. The addition of the prefix $[V]^c$ in item 3 of Definition 6 is original to the current article.

Definition 6 Set of worlds W is the set of all sequences $F_0, X_1, \ldots, X_n, F_n$ where for each i such that $1 \leq i \leq n$,

1. set $F_i \subseteq \Phi$ is a maximal consistent set of formulae,
2. set $X_i \subseteq V$ is a dataset,
3. $\{[V]^c\varphi \mid [V]^c\mathsf{K}_{X_i}\varphi \in F_{i-1}\} \subseteq F_i$.

If $w = F_0, X_1, \ldots, X_n, F_n$ and $u = F_0, X_1, \ldots, X_n, F_n, X_{n+1}, F_{n+1}$, then we say that worlds w and u are *adjacent*. Note that this adjacency relation forms an (undirected) tree structure on set W. By $hd(w)$ we denote the set of formulae F_n.

Definition 7 For any worlds

$$w = F_0, X_1, \ldots, X_n, F_n \in W$$
$$u = F_0, X_1, \ldots, X_n, F_n, X_{n+1}, F_{n+1} \in W,$$

the undirected edge (w, u) is ***labelled*** with a variable $x \in V$ if $x \in (X_{n+1})^*_{F_n}$.

It is convenient to intuitively visualise the worlds of the canonical model as paths in the "Tree of Knowledge", whose fragment is depicted in Fig. 6. For example, on that figure, the world F_0, X_1, F_1 is adjacent to the world F_0, X_1, F_1, X_5, F_5 and the edge between them is labelled with each variable in the dataset $(X_5)^*_{F_1}$.

Definition 8 $w \sim_x u$ if each edge of the unique simple path between nodes w and u is labelled with data variable x.

Definition 9 $\pi(p) = \{(w, U) \mid [V]^c[U]p \in hd(w)\}$.

8.4 Canonical Model Properties

In this subsection, we establish the core properties of the canonical model that will be used later in the proof of completeness. The first of these properties is captured in the lemma below. It describes how formulae can be transferred between adjacent nodes of the tree.

Lemma 25 $[V]^c\mathsf{K}_Y\varphi \in F_{n-1}$ *iff* $[V]^c\mathsf{K}_Y\varphi \in F_n$ *for any world* $F_0, \ldots, X_n, F_n$, *any dataset* $Y \subseteq (X_n)^*_{F_{n-1}}$, *and any formula* $\varphi \in \Phi$, *where* $n \geq 1$.

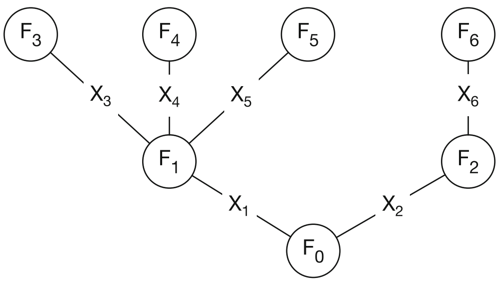

Fig. 6 Fragment of the tree of knowledge

Proof (⇒) Note that $\vdash \mathsf{K}_Y\varphi \to \mathsf{K}_Y\mathsf{K}_Y\varphi$ by Lemma 16. Also,

$$\vdash \mathsf{K}_Y\mathsf{K}_Y\varphi \to \mathsf{K}_{(X_n)^*_{F_{n-1}}}\mathsf{K}_Y\varphi$$

by the Monotonicity axiom and the assumption $Y \subseteq (X_n)^*_{F_{n-1}}$ of the lemma. Thus, by the laws of propositional reasoning,

$$\vdash \mathsf{K}_Y\varphi \to \mathsf{K}_{(X_n)^*_{F_{n-1}}}\mathsf{K}_Y\varphi.$$

Hence,

$$\vdash X_n \rhd (X_n)^*_{F_{n-1}} \to (\mathsf{K}_Y\varphi \to \mathsf{K}_{X_n}\mathsf{K}_Y\varphi)$$

by the Monotonicity axiom and propositional reasoning. Then, by the Necessitation inference rule,

$$\vdash [V]^c(X_n \rhd (X_n)^*_{F_{n-1}} \to (\mathsf{K}_Y\varphi \to \mathsf{K}_{X_n}\mathsf{K}_Y\varphi)).$$

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

$$\vdash [V]^c(X_n \rhd (X_n)^*_{F_{n-1}}) \to [V]^c(\mathsf{K}_Y\varphi \to \mathsf{K}_{X_n}\mathsf{K}_Y\varphi).$$

Hence, by Lemma 24 and the Modus Ponens inference rule,

$$F_{n-1} \vdash [V]^c(\mathsf{K}_Y\varphi \to \mathsf{K}_{X_n}\mathsf{K}_Y\varphi).$$

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

$$F_{n-1} \vdash [V]^c\mathsf{K}_Y\varphi \to [V]^c\mathsf{K}_{X_n}\mathsf{K}_Y\varphi.$$

Thus, by the assumption $[V]^c\mathsf{K}_Y\varphi \in F_{n-1}$ of the case (⇒) and the Modus Ponens rule, $F_{n-1} \vdash [V]^c\mathsf{K}_{X_n}\mathsf{K}_Y\varphi$. Hence, $[V]^c\mathsf{K}_{X_n}\mathsf{K}_Y\varphi \in F_{n-1}$ because set F_{n-1} is maximal. Therefore, $[V]^c\mathsf{K}_Y\varphi \in F_n$ by item 3 of Definition 6.

(⇐) Suppose that $[V]^c\mathsf{K}_Y\varphi \notin F_{n-1}$. Note that formula $\neg\mathsf{K}_Y\varphi \to \mathsf{K}_Y\neg\mathsf{K}_Y\varphi$ is an instance of the Negative Introspection axiom. Also,

$$\vdash \mathsf{K}_Y\neg\mathsf{K}_Y\varphi \to \mathsf{K}_{(X_n)^*_{F_{n-1}}}\neg\mathsf{K}_Y\varphi$$

by the Monotonicity axiom and the assumption $Y \subseteq (X_n)^*_{F_{n-1}}$ of the lemma. Thus, by the laws of propositional reasoning,

$$\vdash \neg\mathsf{K}_Y\varphi \to \mathsf{K}_{(X_n)^*_{F_{n-1}}}\neg\mathsf{K}_Y\varphi.$$

Hence,

$$\vdash X_n \rhd (X_n)^*_{F_{n-1}} \to (\neg\mathsf{K}_Y\varphi \to \mathsf{K}_{X_n}\neg\mathsf{K}_Y\varphi)$$

by the Monotonicity axiom and propositional reasoning. Then, by the Necessitation inference rule,

$$\vdash [V]^c(X_n \rhd (X_n)^*_{F_{n-1}} \to (\neg \mathsf{K}_Y\varphi \to \mathsf{K}_{X_n}\neg \mathsf{K}_Y\varphi)).$$

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

$$\vdash [V]^c(X_n \rhd (X_n)^*_{F_{n-1}}) \to [V]^c(\neg \mathsf{K}_Y\varphi \to \mathsf{K}_{X_n}\neg \mathsf{K}_Y\varphi).$$

Hence, by Lemma 24 and the Modus Ponens inference rule,

$$F_{n-1} \vdash [V]^c(\neg \mathsf{K}_Y\varphi \to \mathsf{K}_{X_n}\neg \mathsf{K}_Y\varphi).$$

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

$$F_{n-1} \vdash [V]^c\neg \mathsf{K}_Y\varphi \to [V]^c\mathsf{K}_{X_n}\neg \mathsf{K}_Y\varphi. \tag{23}$$

At the same time, the assumption $[V]^c\mathsf{K}_Y\varphi \notin F_{n-1}$ of the case ($\Leftarrow$) implies that $\neg[V]^c\mathsf{K}_Y\varphi \in F_{n-1}$ because set F_{n-1} is maximal. Hence, $F_{n-1} \vdash [V]^c\neg \mathsf{K}_Y\varphi$ by the Negation axiom and propositional reasoning. Thus, by statement (23) and the Modus Ponens rule $F_{n-1} \vdash [V]^c\mathsf{K}_{X_n}\neg \mathsf{K}_Y\varphi$. Then, because set F_{n-1} is maximal. $[V]^c\mathsf{K}_{X_n}\neg \mathsf{K}_Y\varphi \in F_{n-1}$. Hence, $[V]^c\neg \mathsf{K}_Y\varphi \in F_n$ by item 3 of Definition 6. Then, $F_n \vdash \neg[V]^c\mathsf{K}_Y\varphi$ by the Deterministity axiom and propositional reasoning. Therefore, $[V]^c\mathsf{K}_Y\varphi \notin F_n$ because set F_n is consistent. □

Recall that there is a unique simple path between any two nodes of a tree. Thus, by Definition 8, statement $w \sim_X u$ implies that each edge of the unique path between nodes w and u is labelled with each variable in set X. Then, the above lemma can be generalised from a property of two adjacent nodes to two arbitrary nodes as follows.

Corollary 2 *$[V]^c\mathsf{K}_Y\varphi \in hd(w)$ iff $[V]^c\mathsf{K}_Y\varphi \in hd(u)$ for any worlds $w, u \in W$ such that $w \sim_Y u$.*

The next two lemmas are used during the induction step of the proof of the "truth lemma" (Lemma 30).

Lemma 26 *For any world $w \in W$ and any formula $\neg[V]^c[U]\mathsf{K}_X\varphi \in hd(w)$, there is a world $u \in W$ such that $w \sim_{U\cup X} u$ and $\neg[V]^c[U]\varphi \in hd(u)$.*

Proof First, we show that the set of formulae

$$G = \{\neg[V]^c[U]\varphi\} \cup \{[V]^c\psi \mid [V]^c\mathsf{K}_{U\cup X}\psi \in hd(w)\} \tag{24}$$

is consistent. Towards a contradiction, assume the opposite. Then, there are formulae

$$[V]^c\mathsf{K}_{U\cup X}\psi_1, \dots, [V]^c\mathsf{K}_{U\cup X}\psi_n \in hd(w), \tag{25}$$

where $[V]^c\psi_1, \dots, [V]^c\psi_n \vdash [V]^c[U]\varphi$. Thus, by Lemma 19,

$$[V]^c\mathsf{K}_{U\cup X}\psi_1, \dots, [V]^c\mathsf{K}_{U\cup X}\psi_n \vdash [V]^c\mathsf{K}_{U\cup X}[U]\varphi.$$

Hence, by assumption (25),

$$hd(w) \vdash [V]^c\mathsf{K}_{U\cup X}[U]\varphi. \tag{26}$$

At the same time, $\vdash [V]^c\mathsf{K}_{U\cup X}[U]\varphi \to [V]^c[U]\mathsf{K}_X\varphi$ by the Epistemic Commutativity axiom and Lemma 20. Thus, $hd(w) \vdash [V]^c[U]\mathsf{K}_X\varphi$ by statement (26) and the Modus Ponens rule, which contradicts the assumption $\neg[V]^c[U]\mathsf{K}_X\varphi \in hd(w)$ of the lemma and the consistency of the set $hd(w)$. Thus, set G is consistent.

By Lemma 18, set G can be extended to a maximal consistent set of formulae G'. Suppose that epistemic world w is the sequence $F_0, X_1, \dots, X_n, F_n$. Define world u to be the sequence $F_0, X_1, \dots, X_n, F_n, U \cup X, G'$. Note that $u \in W$ by Definition 6, statement (24), and the assumption $G \subseteq G'$.

Observe that $U \cup X \subseteq (U \cup X)^*_{F_n}$ by Lemma 23. Hence, $w \sim_{U\cup X} u$ by Definition 8 and the choice of sequence u. Finally, $\neg[V]^c[U]\varphi \in G \subseteq G' = hd(u)$ by statement (24). □

Lemma 27 *If* $[V]^c[U]\mathsf{K}_X\varphi \in hd(w)$ *and* $w \sim_{U\cup X} u$*, then* $[V]^c[U]\varphi \in hd(u)$.

Proof Observe that the formula $\mathsf{K}_{U\cup X}[U]\varphi \leftrightarrow [U]\mathsf{K}_X\varphi$ is an instance of the Epistemic Commutativity axiom. Thus, $\vdash [V]^c\mathsf{K}_{U\cup X}[U]\varphi \leftrightarrow [V]^c[U]\mathsf{K}_X\varphi$ by Corollary 1.

Then, $hd(w) \vdash [V]^c\mathsf{K}_{U\cup X}[U]\varphi$ by the assumption $[V]^c[U]\mathsf{K}_X\varphi \in hd(w)$ of the lemma. Hence, because set $hd(w)$ is maximal, $[V]^c\mathsf{K}_{U\cup X}[U]\varphi \in hd(w)$. Then, by Lemma 25 and the assumption $w \sim_{U\cup X} u$ of the lemma,

$$[V]^c\mathsf{K}_{U\cup X}[U]\varphi \in hd(u). \tag{27}$$

Note that $\mathsf{K}_{U\cup X}[U]\varphi \to [U]\varphi$ is an instance of the Truth axiom. Thus, by Lemma 20,

$$\vdash [V]^c\mathsf{K}_{U\cup X}[U]\varphi \to [V]^c[U]\varphi.$$

Then, $hd(u) \vdash [V]^c[U]\varphi$ by statement (27) and the Modus Ponens inference rule. Therefore, $[V]^c[U]\varphi \in hd(u)$ because set $hd(u)$ is maximal. □

Lemma 28 *If* $[V]^c[U](X \rhd Y) \in hd(w)$ *and* $w \sim_{U\cup X} w'$*, then* $w \sim_Y w'$.

Proof We prove the lemma by induction on the length of the simple path between vertices w and w'. If $w = w'$, then, vacuously, each edge along the simple path between vertices w and w' is labelled with each data variable. Hence, $w \sim_Y w'$ by Definition 8.

Suppose that $w \neq w'$. Consider the unique simple path between vertices w and w'. By the assumption $w \sim_{U\cup X} w'$ of the lemma and Definition 8, each edge along this path is labelled with each data variable in set $U \cup X$. Because $w \neq w'$, there must exist a vertex $u \in W$ on the unique simple path between w and w' such that vertices

u and w' are adjacent. Thus, each edge along the simple path between vertices w and u is labelled with each data variable in set $U \cup X$. Hence, by Definition 8,

$$w \sim_{U\cup X} u. \tag{28}$$

Claim The edge between vertices u and w' is labelled with each data variable in set Y.

Proof of Claim We consider the following two cases separately, see Fig. 7:

Case I: $u = F_0, X_1, F_1, \ldots, F_{n-1}$ and $w' = F_0, X_1, F_1, \ldots, X_n, F_n$. Consider any data variable $y \in Y$. By Definition 7, it suffices to show that $y \in (X_n)^*_{hd(u)}$. Note that $\vdash Y \rhd \{y\}$ by the Reflexivity axiom. Then, by the Necessitation inference rule applied twice,

$$\vdash [V]^c[U](Y \rhd \{y\}). \tag{29}$$

At the same time, by the Introspection of Dependency axiom,

$$\vdash X \rhd Y \to \mathsf{K}_X(X \rhd Y).$$

Thus, by Lemma 20 applied twice,

$$\vdash [V]^c[U](X \rhd Y) \to [V]^c[U]\mathsf{K}_X(X \rhd Y).$$

Note that $[V]^c[U](X \rhd Y) \in hd(w)$ by the assumption of the lemma. Thus, by the Modus Ponens inference rule,

$$hd(w) \vdash [V]^c[U]\mathsf{K}_X(X \rhd Y).$$

Hence, $[V]^c[U]\mathsf{K}_X(X \rhd Y) \in hd(w)$ because set $hd(w)$ is maximal. Then, by Lemma 27 and statement (28),

$$[V]^c[U](X \rhd Y) \in hd(u). \tag{30}$$

Recall that u is a vertex on the simple path connecting vertices w and w' and all edges along this path are labelled with variables from set X. Hence, $X \subseteq (X_n)^*_{hd(u)}$

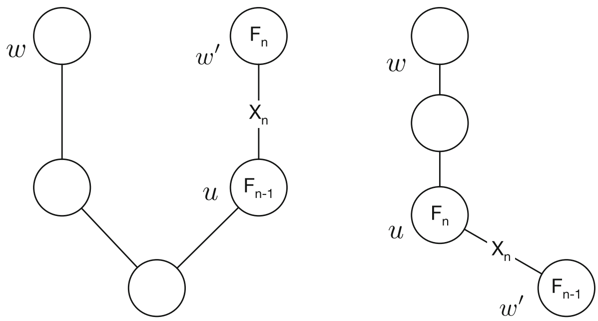

Fig. 7 Case I (left) and Case II (right)

by Definition 7. Then, $\vdash (X_n)^*_{hd(u)} \rhd X$ by the Reflexivity axiom. Thus, by the Necessitation inference rule applied twice,

$$\vdash [V]^c[U]\left((X_n)^*_{hd(u)} \rhd X\right). \tag{31}$$

Finally, note that the following two formulae are instances of the Transitivity axiom:

$$(X_n)^*_{hd(u)} \rhd X \to \left(X \rhd Y \to (X_n)^*_{hd(u)} \rhd Y\right),$$
$$(X_n)^*_{hd(u)} \rhd Y \to \left(Y \rhd \{y\} \to (X_n)^*_{hd(u)} \rhd \{y\}\right).$$

Thus, by the laws of propositional reasoning,

$$(X_n)^*_{hd(u)} \rhd X,\ \ X \rhd Y,\ \ Y \rhd \{y\} \vdash (X_n)^*_{hd(u)} \rhd \{y\}.$$

Then, by Lemma 19 applied twice,

$$[V]^c[U]\left((X_n)^*_{hd(u)} \rhd X\right), [V]^c[U](X \rhd Y),\ [V]^c[U](Y \rhd \{y\})$$
$$\vdash [V]^c[U]\left((X_n)^*_{hd(u)} \rhd \{y\}\right).$$

Hence, by statements (31), (30), and (29),

$$hd(u) \vdash [V]^c[U]\left((X_n)^*_{hd(u)} \rhd \{y\}\right).$$

Therefore, $y \in (X_n)^*_{hd(u)}$ by Definition 5.

Case II: $w' = F_0, X_1, F_1, \dots, F_{n-1}$ and $u = F_0, X_1, F_1, \dots, X_n, F_n$. This case is similar to the previous one, except that it uses the set $hd(w')$ instead of the set $hd(u)$ everywhere in the proof. □

To finish the proof of the lemma, note that the simple path between vertices w and u is shorter than the simple path between vertices w and w'. Hence, $w \sim_Y u$, by the induction hypothesis. Also, $u \sim_Y w'$ by Claim 8.4 and Definition 8. Therefore, $w \sim_Y w'$ because relation $\sim_Y$ is transitive. □

Lemma 29 *If $[V]^c[U](X \rhd Y) \notin hd(w)$, then there is a world $w' \in W$ such that $w \sim_{U\cup X} w'$ and $w \nsim_Y w'$.*

Proof Let state w be sequence $F_0, X_1, \dots, X_{n-1}, F_{n-1}, X_n, F_n$. Consider sequence

$$w' = F_0, X_1, \dots, X_{n-1}, F_{n-1}, X_n, F_n, U \cup X, F_n.$$

To prove that $w' \in W$, consider any formula $[V]^c\mathsf{K}_{U\cup X}\varphi \in F_n$. By item 3 of Definition 6, it suffices to show that $[V]^c\varphi \in F_n$. Indeed, $\vdash \mathsf{K}_{U\cup X}\varphi \to \varphi$ by the Truth axiom. Thus, $\vdash [V]^c\mathsf{K}_{U\cup X}\varphi \to [V]^c\varphi$ by Lemma 20. Hence, the assumption

$[V]^c \mathsf{K}_{U\cup X}\varphi \in F_n$ implies $F_n \vdash [V]^c\varphi$ by the Modus Ponens inference rule. Therefore, $[V]^c\varphi \in F_n$ because set F_n is maximal.

To prove $w \sim_{U\cup X} w'$, note that $(U \cup X) \subseteq (U \cup X)^*_{F_n}$ by Lemma 23. Thus, by Definition 7, the edge between vertices w and w' is labelled with each data variable in set $U \cup X$. Therefore, $w \sim_{U\cup X} w'$ by Definition 8.

Finally, we show that $w \nsim_Y w'$. By Definition 8, it suffices to prove that the simple path between vertices w and w' is *not* labelled by at least one variable from set Y. Then, by Definition 7, it suffices to show that $Y \nsubseteq (U \cup X)^*_{F_n}$. Suppose the opposite. Thus, $\vdash (U \cup X)^*_{F_n} \rhd Y$ by the Reflexivity axiom. Hence, by the Necessitation inference rule,

$$\vdash [V]^c \left((U \cup X)^*_{F_n} \rhd Y\right). \tag{32}$$

Note that the following two formulae are instances of the Transitivity and the Partial Revelation axioms, respectively:

$$(U \cup X) \rhd (U \cup X)^*_{F_n} \to \left((U \cup X)^*_{F_n} \rhd Y \to (U \cup X) \rhd Y\right),$$

$$(U \cup X) \rhd Y \to [U](X \rhd Y).$$

Thus, by the laws of propositional reasoning,

$$(U \cup X) \rhd (U \cup X)^*_{F_n},\ (U \cup X)^*_{F_n} \rhd Y \vdash [U](X \rhd Y).$$

Then, by Lemma 19,

$$[V]^c \left((U \cup X) \rhd (U \cup X)^*_{F_n}\right),\ [V]^c \left((U \cup X)^*_{F_n} \rhd Y\right) \vdash [V]^c[U](X \rhd Y).$$

Thus, by statement (32),

$$[V]^c \left((U \cup X) \rhd (U \cup X)^*_{F_n}\right) \vdash [V]^c[U](X \rhd Y).$$

Hence, $F_n \vdash [V]^c[U](X \rhd Y)$ by Lemma 24. Thus, because set F_n is consistent,

$$[V]^c[U](X \rhd Y) \in F_n = hd(w),$$

which contradicts the assumption $[V]^c[U](X \rhd Y) \notin hd(w)$ of the lemma. $\square$

8.5 Final Steps

Next is the “truth lemma” for our construction. The prefix $[V]^c[U]$ distinguishes this lemma from the standard truth lemma in modal logic.

Lemma 30 *$w, U \Vdash \varphi$ iff $[V]^c[U]\varphi \in hd(w)$, for each world $w \in W$, each dataset $U \subseteq V$, and each formula $\varphi \in \Phi$.*

Proof We prove this statement by induction on the structural complexity of formula φ. If formula φ is an atomic proposition, then the statement of the lemma follows from item 1 of Definition 2 and Definition 9.

Suppose that formula φ has the form $X \rhd Y$.

$(\Rightarrow)$: Assume that $[V]^c[U](X \rhd Y) \notin hd(w)$. Thus, by Lemma 29, there is a world $w' \in W$ such that $w \sim_{U \cup X} w'$ and $w \nsim_Y w'$. Therefore, $w, U \nVdash X \rhd Y$ by item 7 of Definition 2.

$(\Leftarrow)$: Assume that $[V]^c[U](X \rhd Y) \in hd(w)$. Then, $w \sim_Y w'$ for any world $w' \in W$ such that $w \sim_{U \cup X} w'$ by Lemma 28. Therefore, $w, U \Vdash X \rhd Y$ by item 7 of Definition 2.

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

By Definition 2, the statement $w, U \Vdash \neg\psi$ is equivalent to the statement $w, U \nVdash \psi$. By the induction hypothesis, the last statement is equivalent to $[V]^c[U]\psi \notin hd(w)$. The statement $[V]^c[U]\psi \notin hd(w)$ is equivalent to the statement $hd(w) \vdash \neg[V]^c[U]\psi$ because $hd(w)$ is a maximal consistent set of formulae. The latter statement is equivalent to the statement $hd(w) \vdash [V]^c\neg[U]\psi$ by the Negation axiom. The statement $hd(w) \vdash [V]^c\neg[U]\psi$ is equivalent to the statement $hd(w) \vdash [V]^c[U]\neg\psi$ by the Negation axiom and Corollary 1. Finally, because $hd(w)$ is a maximal consistent set of formulae, the last statement is equivalent to $[V]^c[U]\neg\psi \in hd(w)$.

Suppose formula φ has the form $\psi \to \chi$. By Definition 2, the statement $w, U \Vdash \psi \to \chi$ is equivalent to the *disjunction* of the following statements: $w, U \nVdash \psi$ and $w, U \Vdash \chi$. Note that, by the induction hypothesis, the disjunction of those two statements is equivalent to the *disjunction* of the statement $[V]^c[U]\psi \notin hd(w)$ and the statement $[V]^c[U]\chi \in hd(w)$. The latter is equivalent to $hd(w) \vdash \neg[V]^c[U]\psi \vee [V]^c[U]\chi$ because $hd(w)$ is a maximal consistent set of formulae. In turn, the last statement is equivalent to $hd(w) \vdash [V]^c[U]\psi \to [V]^c[U]\chi$ by propositional reasoning. Note that the previous statement is equivalent to $hd(w) \vdash [V]^c[U](\psi \to \chi)$ by Lemma 21 and propositional reasoning. Finally, the last statement is equivalent to $[V]^c[U](\psi \to \chi) \in hd(w)$ because $hd(w)$ is a maximal consistent set of formulae.

Suppose that formula φ has the form $\mathsf{K}_X\psi$.

$(\Rightarrow)$: Assume that $[V]^c[U]\mathsf{K}_X\psi \notin hd(w)$. Thus, $\neg[V]^c[U]\mathsf{K}_X\psi \in hd(w)$ because set $hd(w)$ is maximal. Hence, by Lemma 26, there is a world $u \in W$ such that $w \sim_{U \cup X} u$ and $\neg[V]^c[U]\psi \in hd(u)$. Then, $[V]^c[U]\psi \notin hd(u)$ because set $hd(u)$ is consistent. Thus, $u, U \nVdash \psi$ by the induction hypothesis. Therefore, $w, U \nVdash \mathsf{K}_X\psi$ by item 4 of Definition 2 and statement $w \sim_{U \cup X} u$.

$(\Leftarrow)$: Assume that $[V]^c[U]\mathsf{K}_X\psi \in hd(w)$. Thus, by Lemma 27, for each epistemic world $u \in W$, if $w \sim_{U \cup X} u$, then $[V]^c[U]\psi \in hd(u)$. Hence, by the induction hypothesis, for each world $u \in W$, if $w \sim_{U \cup X} u$, then $u, U \Vdash \psi$. Therefore, $w, U \Vdash \mathsf{K}_X\psi$ by item 4 of Definition 2.

Suppose that formula φ has the form $[X]\psi$.

By item 5 of Definition 2, the statement $w, U \Vdash [X]\psi$ is equivalent to the statement $w, U \cup X \Vdash \psi$. In turn, $w, U \cup X \Vdash \psi$ iff $[V]^c[U \cup X]\psi \in hd(w)$ by the induction hypothesis. The statement $[V]^c[U \cup X]\psi \in hd(w)$ is equivalent to the statement $[V]^c[U][X]\psi \in hd(w)$ by the Composition axiom, Corollary 1, and the maximality of the set $hd(w)$.

Finally, suppose formula φ has the form $[X]^c\psi$.

By item 6 of Definition 2, the statement $w, U \Vdash [X]^c\psi$ is equivalent to the statement $w, U \setminus X \Vdash \psi$. The latter is equivalent to the statement $[V]^c[U \setminus X]\psi \in hd(w)$ by the induction hypothesis. In turn, by Lemma 22 and the maximality of the set $hd(w)$, the last statement is equivalent to $[V]^c[U][X]^c\psi \in hd(w)$. □

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

Theorem 4 *If $\nvdash \varphi$, then there is a world w of an epistemic model, and a dataset $U \subseteq V$ such that $w, U \nVdash \varphi$.*

Proof Suppose that $\nvdash \varphi$. Thus, it follows by the Powerset axiom that there is a dataset $U \subseteq V$ such that $\nvdash [V]^c[U]\varphi$. Thus, the single-element set $\{\neg[V]^c[U]\varphi\}$ is consistent. Then, by Lemma 18, there is a maximal consistent set of formulae $F_0 \subseteq \Phi$ such that $\{\neg[V]^c[U]\varphi\} \subseteq F_0$. Consider the canonical model $M(F_0)$. Let world w be the single-element sequence F_0. Then,

$$\neg[V]^c[U]\varphi \in \{\neg[V]^c[U]\varphi\} \subseteq F_0 = hd(w).$$

Hence, $[V]^c[U]\varphi \notin hd(w)$ because set $hd(w)$ is consistent. Therefore, $w, U \nVdash \varphi$ by Lemma 30. □

Because of $[V]^c[U]$ prefix in the proof above, this proof cannot be easily generalised to a proof of the *strong* completeness. The strong completeness of our system remains an open problem.

9 Model Checking

Recall that in the beginning of Section 2 we have assumed that the finite set V is fixed throughout the article. In such a setting, by a model checking problem, we mean deciding if $w_0, U_0 \Vdash \varphi_0$ is true for the given world w_0 of an epistemic model, dataset $U_0 \subseteq V$, and formula $\varphi_0 \in \Phi$.

There is a straightforward dynamic programming algorithm for solving this model checking problem. The algorithm fills in an array $sat[w, U, \varphi]$ for each world $w \in W$, each dataset $U \subseteq V$, and each subformula φ of the original formula φ_0. It ensures that the value of $sat[w, U, \varphi]$ is true iff $w, U \Vdash \varphi$. The size of the array sat is a polynomial function of the input because the set V is fixed and, thus, has a constant number of subsets U. Thus, the straightforward implementation of this algorithm has a polynomial execution time.

One can also consider a more natural form of model checking problem where a finite set V is given as a part of the input. In this case, there are exponentially many subsets of V. As a result, a straightforward implementation of the above algorithm has exponential execution time.

In the rest of this section, we present a polynomial-time model checking algorithm for the more general model checking problem when V is given as a part of the input. Informally, to address the issue of the exponential number of subsets of V, we compute the value of $sat[w, U, \varphi]$ only for the pairs (U, φ) which are needed to evaluate the value $sat[w_0, U_0, \varphi_0]$. Formally, our version of the model checking algorithm first invokes a helper function H that returns the list of all pairs (U_1, φ_1), ..., (U_n, φ_n) required for the computation. Then, it creates a two-dimensional array $sat[w, i]$, where $1 \leq i \leq n$. Finally, the algorithm fills the array while ensuring that $sat[w, i]$ has the Boolean value true iff $w, U_i \Vdash \varphi_i$.

The recursive code for the function $H(U, \varphi)$ is given in Fig. 8. In that code, the expression $[(U', \psi)]$ denotes a list whose only element is the pair (U', ψ). Note that the execution time of this algorithm is a linear function of the size of the formula φ.

Figure 9 presents the code for the model checking algorithm that uses the helper function $H(U, \varphi)$. Observe that the execution time of this algorithm is polynomial in terms of the size of the array $sat[w, i]$. The size of the array, as discussed earlier, is a polynomial function of the input.

Another commonly asked algorithmic question about logical systems is the decidability of the set of all its theorems. Usually, the decidability is shown by using filtration to prove completeness with respect to the set of finite models. Unfortunately, we do not know how to apply filtration to the tree construction in our proof of completeness. As a result, the decidability of our logical system remains an open question.

10 Causality by Epistemic Events

In this section, we discuss how causality by epistemic events can be captured in our logical system. Various formal approaches to capturing causality have been proposed before [2, 7, 8, 10, 19, 20, 31, 33, 38]. Here, we focus on *counterfactual* definition of causality: an event is a cause of φ if statement φ would not have been true without the event. A variation of counterfactual causality in a formal setting has been studied in [22].

procedure $\mathrm{H}(U, \varphi)$
  switch φ **do**
    case φ is an atomic proposition
      return $[(U, \varphi)]$
    case φ has the form $\neg\psi$
      return $\mathrm{H}(U, \psi) + [(U, \varphi)]$
    case φ has the form $\psi \to \chi$
      return $\mathrm{H}(U, \psi) + \mathrm{H}(U, \chi) + [(U, \varphi)]$
    case φ has the form $\mathsf{K}_X\psi$
      return $\mathrm{H}(U, \psi) + [(U, \varphi)]$
    case φ has the form $[X]\psi$
      return $\mathrm{H}(U \cup X, \psi) + [(U, \varphi)]$
    case φ has the form $[X]^c\psi$
      return $\mathrm{H}(U \setminus X, \psi) + [(U, \varphi)]$
    case φ has the form $X \rhd Y$
      return $[(U, \varphi)]$
end procedure

Fig. 8 Helper function

```
for (U, φ) ∈ H(U_0, φ_0) do
    for w ∈ W do
        switch φ do
            case φ is an atomic proposition
                if (w, U) ∈ π(φ) then
                    sat[w, (U, φ)] ← true
                else
                    sat[w, (U, φ)] ← false
                end if
            case φ has the form ¬ψ
                sat[w, (U, φ)] ← ¬sat[w, (U, ψ)]
            case φ has the form ψ → χ
               sat[w,(U, φ)] ← ¬sat[w,(U, ψ)] ∨ sat[w, (U, χ)]
            case φ has the form K_X ψ
                answer ← true
                for w′ ∈ W do
                    if w ∼_{U∪X} w′ and ¬sat[w′, (U, ψ)] then
                        answer ← false
                        break
                    end if
                end for
                sat[w, (U, φ)] ← answer
            case φ has the form [X]ψ
                sat[w, (U, φ)] ← sat[w, (U ∪ X, ψ)]
            case φ has the form [X]^c ψ
                sat[w, (U, φ)] ← sat[w, (U \ X, ψ)]
            case φ has the form X ▷ Y
                answer ← true
                for w′ ∈ W do
                    if w ∼_{U∪X} w′ and w ≁_Y w′ then
                        answer ← false
                        break
                    end if
                end for
                sat[w, (U, φ)] ← answer
    end for
end for
```

Fig. 9 Model checking algorithm. To improve the readability of the code, if (U, φ) is (U_i, φ_i) for some $i \leq n$, then we write $sat[w, (U, \varphi)]$ instead of $sat[w, i]$

In the case of causality by epistemic events, we can distinguish *forward* and *backward* forms of causality that we denote by modalities $\mathsf{F}_x\varphi$ and $\mathsf{B}_x\varphi$, respectively. Modality $\mathsf{F}_x\varphi$ means that "statement φ is true, and it would not have been true in the same situation if data variable x had been publicly announced". Modality $\mathsf{B}_x\varphi$ means that "statement φ is true, and it would not have been true in the same situation if data variable x had not been publicly announced". Both of these two modalities can be expressed in our language:

$$\mathsf{F}_x\varphi \equiv \varphi \wedge [x]\neg\varphi, \tag{33}$$

$$\mathsf{B}_x\varphi \equiv \varphi \wedge [x]^c\neg\varphi. \tag{34}$$

In other words, backward causality B_x captures the effect of the revelation of the value of x, while forward causality captures the effect of keeping the value of x concealed.

It is perhaps surprising that the opposite is also true. Our revelation and concealment modalities can be expressed through F_x and B_x:

$$[x]\varphi \equiv (\neg\varphi \wedge \mathsf{F}_x\neg\varphi) \vee (\varphi \wedge \neg\mathsf{F}_x\varphi),$$
$$[x]^c\varphi \equiv (\neg\varphi \wedge \mathsf{B}_x\neg\varphi) \vee (\varphi \wedge \neg\mathsf{B}_x\varphi).$$

Going back to our introductory example, the epistemic event of the public revelation of the content of the article is a backward cause of the fact that everyone who knew the name of the financial advisor knew that he had a foreclosure. Formally, using statements (3) and (4),

$$w_1, \{\text{article}\} \Vdash \mathsf{B}_{\text{article}}\mathsf{K}_{\text{name}}(\text{“the advisor had foreclosure”}).$$

Of course, the subsequent concealment of the article by Google is a *forward* cause for the lack of such knowledge:

$$w_1, \{\text{article}\} \Vdash [\text{article}]^c\mathsf{F}_{\text{article}}\neg\mathsf{K}_{\text{name}}(\text{“the advisor had foreclosure”}).$$

Modalities F and B can be generalised from single variables to datasets. To do this, we need to add the minimality condition to formulae (33) and (34):

$$\mathsf{F}_X\varphi \equiv \varphi \wedge [X]\neg\varphi \wedge \bigwedge_{Y \subsetneq X}[Y]\varphi,$$
$$\mathsf{B}_X\varphi \equiv \varphi \wedge [X]^c\neg\varphi \wedge \bigwedge_{Y \subsetneq X}[Y]^c\varphi.$$

11 Conclusion

Following several other recent works, in this article, we have advocated an approach to reasoning about knowledge that treats datasets, not agents, as the source of knowledge. Our main technical result is a sound and complete logical system that describes the interplay between data-informed knowledge, public concealment, public revelation, and functional dependency operators. We also prove the undefinability of the revelation and the concealment modalities through each other. Finally, we describe and analyse a model checking algorithm for this system.

Although variations of these modalities have been studied before, original to this article is the description of the interplay between them captured in the Monotonicity, Introspection of Dependency, Inversion, Commutativity, Epistemic Commutativity, and Powerset axioms as well as in a non-trivial Concealed Monotonicity inference rule. To show that our axiomatisation of this interplay is complete, we significantly modified several known constructions. First, in Definition 5, we modified the dataset closure operator $*$, whose origins could be traced to [4], by adding parameter F

and including prefix $[V]^c$. Second, in Definition 6, we modified the previously used tree construction [16] by including the prefix $[V]^c$ in item 3. Third, in Definition 7, we redefined tree labelling to use closure operator $(\cdot)^*_{F_n}$. Finally, we propose a non-standard (for modal logic) version of the truth lemma (Lemma 30) that includes the prefix $[V]^c[U]$ in front of formula φ.

A possible direction for future work is an extension of this system by other modalities and relations dealing with datasets. One of them is the conditional functional dependency relation $X \rhd_\varphi Y$ which means that dataset X informs dataset Y conditionally upon statement φ being true. It is inspired by Wang and Fan's "conditionally knowing value" expression defined as "agent a knows the value of data variable x assuming condition φ is true"[36].

Author Contributions All authors contributed equally to this work. Their names are listed in alphabetical order.

Funding Junli Jiang acknowledges the support of the "Innovation Research 2035 Pilot Plan of Southwest University" (NO. SWUPilotPlan018).

Availability of data and materials This work does not use any datasets or materials.

Declarations

Ethical Approval This work does not deal with human beings and as such it requires no ethical approvals.

Competing interests The authors declare that they have no potential conflict of interest in relation to the study in this article.

References

1. Ågotnes, T., Balbiani, P., van Ditmarsch, H., & Seban, P. (2010). Group announcement logic. *Journal of Applied Logic, 8*(1), 62–81. https://doi.org/10.1016/j.jal.2008.12.002
2. Alur, R., Peled, D., Penczek, W. (1995). Model-checking of causality properties. In: Proceedings of Tenth Annual IEEE Symposium on Logic in Computer Science, pp. 90–100. IEEE.
3. Andreassen, A.M. (2017). Nine lines that changed history – at least on the internet. EU Observer, 12. https://euobserver.com/digital/138958.
4. Armstrong, W.W. (1974). Dependency structures of data base relationships. In: Information Processing 74 (Proc. IFIP Congress, S tockholm, 1974), pp. 580–583. North-Holland, Amsterdam.
5. Baltag, A. (2016). To know is to know the value of a variable. In: L.D. Beklemishev, S. Demri, A. Maté (eds.) Advances in Modal Logic 11, proceedings of the 11th conference on Advances in Modal Logic, Budapest, Hungary, August 30 – September 2, 2016, pp. 135–155. College Publications.
6. Baltag, A., & van Benthem, J. (2021). A simple logic of functional dependence. *Journal of Philosophical Logic, 50*, 1–67.

7. Batusov, V., Soutchanski, M. (2018). Situation calculus semantics for actual causality. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)
8. Beckers, S. (2021). The counterfactual NESS definition of causation. *Proceedings of the AAAI Conference on Artificial Intelligence, 35*, 6210–6217.
9. Beeri, C., Fagin, R., Howard, J.H. (1977). A complete axiomatization for functional and multivalued dependencies in database relations. In: SIGMOD'77: Proceedings of the 1977 ACM SIGMOD international conference on Management of data, pp. 47–61. ACM, New York, NY, USA. https://doi.org/10.1145/509404.509414.
10. Berreby, F., Bourgne, G., Ganascia, J.G. (2018). Event-based and scenario-based causality for computational ethics. In: AAMAS 2018-17th International Conference on Autonomous Agents and Multiagent Systems, pp. 147–155. International Foundation for Autonomous Agents and Multiagent Systems.
11. Clifford, A., Preston, G. (1967). The Algebraic Theory of Semigroups, Volume II, vol. 2. American Mathematical Soc.
12. Deuser, K., Naumov, P. (2018). Armstrong's axioms and navigation strategies. In: Proceedings of Thirty-Second AAAI Conference on Artificial Intelligence.
13. van Ditmarsch, H., van der Hoek, W., & Kooi, B. (2007). Dynamic Epistemic Logic. *Springer*. https://doi.org/10.1007/978-1-4020-5839-4
14. van Eijck, J., Gattinger, M., Wang, Y. (2017). Knowing values and public inspection. In: Indian Conference on Logic and Its Applications, pp. 77–90. Springer
15. Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). Reasoning about knowledge. *MIT Press, Cambridge, MA*. https://doi.org/10.7551/mitpress/5803.001.0001
16. Fagin, R., Halpern, J. Y., & Vardi, M. Y. (1992). What can machines know? on the properties of knowledge in distributed systems. *Journal of the ACM (JACM), 39*(2), 328–376.
17. Galimullin, R., Alechina, N (2017) Coalition and group announcement logic. In: Proceedings Sixteenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK) 2017, Liverpool, UK, 24-26 July 2017, pp. 207–220.
18. Garcia-Molina, H., Ullman, J., & Widom, J. (2009). *Database Systems: The Complete Book* (2nd ed.). Prentice-Hall.
19. Giordano, L., Martelli, A., & Schwind, C. (2000). Ramification and causality in a modal action logic. *Journal of Logic and Computation, 10*(5), 625–662.
20. Giordano, L., & Schwind, C. (2004). Conditional logic of actions and causation. *Artificial Intelligence, 157*(1–2), 239–279.
21. Grossi, D., Lorini, E., & Schwarzentruber, F. (2015). The ceteris paribus structure of logics of game forms. *Journal of Artificial Intelligence Research, 53*, 91–126.
22. Halpern, J. Y. (2016). *Actual causality*. MIT Press.
23. van der Hoek, W., Iliev, P., & Wooldridge, M. J. (2012). A logic of revelation and concealment. In W. van der Hoek, L. Padgham, V. Conitzer, & M. Winikoff (Eds.), *International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2012, Valencia, Spain, June 4–8, 2012 (3 Volumes)* (pp. 1115–1122). IFAAMAS.
24. Jiang, J., Naumov, P. (2022). Data-informed knowledge and strategies. Artificial Intelligence, 309, 103727. https://doi.org/10.1016/j.artint.2022.103727.
25. Jiang, J., Naumov, P. (2022). In data we trust: The logic of trust-based beliefs. In: the 31st International Joint Conference on Artificial Intelligence (IJCAI-22).
26. Knight, S., Naumov, P., Shi, Q., Suntharraj, V. (2022). Truth set algebra: A new way to prove undefinability. arXiv:2208.04422.
27. Lin, F., Reiter, R. (1994). Forget it. In: Working Notes of AAAI Fall Symposium on Relevance, pp. 154–159.
28. Naumov, P., & Tao, J. (2017). Price of privacy. *Journal of Applied Logic, 20*, 32–48. https://doi.org/10.1016/j.jal.2016.11.035
29. Plaza, J. (1989). Logics of public communications. In: Proceedings 4th International Symposium on Methodologies for Intelligent Systems: Poster session program, pp. 201–216. Oak Ridge National Laboratory.
30. Plaza, J. (2007). Logics of public communications. *Synthese, 158*(2), 165–179.
31. Sarmiento, C., Bourgne, G., Inoue, K., Ganascia, J.G. (2022). Action languages based actual causality in decision making contexts. In: PRIMA 2022: Principles and Practice of Multi-Agent Systems: 24th International Conference, Valencia, Spain, November 16–18, 2022, Proceedings, pp. 243–259. Springer.

32. Su, K., Sattar, A., Lv, G., & Zhang, Y. (2009). Variable forgetting in reasoning about knowledge. *Journal of Artificial Intelligence Research, 35*, 677–716.
33. Turner, H. (1999). A logic of universal causation. *Artificial Intelligence, 113*(1–2), 87–123.
34. Väänänen, J. (2007). Dependence logic: A new approach to independence friendly logic, vol. 70. Cambridge University Press.
35. Wang, Y., Fan, J. (2013). Knowing that, knowing what, and public communication: Public announcement logic with kv operators. In: F. Rossi (ed.) IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013, pp. 1147–1154. IJCAI/AAAI.
36. Wang, Y., & Fan, J. (2014). Conditionally knowing what. In R. Goré, B. P. Kooi, & A. Kurucz (Eds.), *Advances in Modal Logic 10, invited and contributed papers from the tenth conference on Advances in Modal Logic, held in Groningen, The Netherlands, August 5–8, 2014* (pp. 569–587). College Publications.
37. Wáng, Y. N., & Ågotnes, T. (2013). Public announcement logic with distributed knowledge: expressivity, completeness and complexity. *Synthese, 190*(1), 135–162.
38. Wright, R. W. (2011). The NESS account of natural causation: A response to criticisms. In R. Goldberg (Ed.), *Perspectives on Causation* (pp. 285–322). London, UK: Hart Publishing.