Observing interventions: a logic for thinking about experiments

Barbero, F.; Schulz, K.; Velázquez-Quesada, F.R.; Xie, K.

DOI
10.1093/logcom/exac011
Publication date
2023
Document Version
Final published version
Published in
Journal of Logic and Computation

Observing interventions: A logic for thinking about experiments

FAUSTO BARBERO*, Department of Philosophy, History and Art Studies University of Helsinki PL 24 (Unioninkatu 40) 00014 Helsinki, Finland.*
E-mail: fausto.barbero@helsinki.fi

KATRIN SCHULZ*, Institute for Logic, Language and Computation (ILLC) Faculty of Science University of Amsterdam Science Park 107 1098 XG Amsterdam The Netherlands.*
E-mail: K.Schulz@uva.nl

FERNANDO R. VELÁZQUEZ-QUESADA*, Department of Information Science and Media Studies, Universitetet i Bergen. Bergen, 5020, Norway.*
E-mail: Fernando.VelazquezQuesada@uib.no

KAIBO XIE*, Department of Philosophy, Tsinghua University, 100084, Beijing, China.*
E-mail: xkb@mail.tsinghua.edu.cn

Abstract

This paper makes a first step towards a logic of learning from experiments. For this, we investigate formal frameworks for modeling the interaction of causal and (qualitative) epistemic reasoning. Crucial for our approach is the idea that the notion of an *intervention* can be used as a formal expression of a (real or hypothetical) experiment (Pearl, 2009, *Causality. Models, Reasoning, and Inference*, 2nd edn. Cambridge University Press, Cambridge; Woodward, 2003, *Making Things Happen*, vol. 114 of Oxford Studies in the Philosophy of Science. Oxford University Press). In a first step we extend a causal model (Briggs, 2012, *Philosophical Studies*, 160, 139–166; Galles and Pearl, 1998, An axiomatic characterisation of causal counterfactuals. *Foundations of Science*, 3, 151–182; Halpern, 2000, Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12, 317–337; Pearl, 2009, *Causality. Models, Reasoning, and Inference*, 2nd edn. Cambridge University Press, Cambridge) with a simple Hintikka-style representation of the epistemic state of an agent. In the resulting setting, one can talk about the knowledge of an agent and information update. The resulting logic can model reasoning about thought experiments. However, it is unable to account for learning from experiments, which is clearly brought out by the fact that it validates the principle of *no learning* for interventions. Therefore, in a second step, we implement a more complex notion of knowledge (Nozick, 1981, *Philosophical Explanations*. Harvard University Press, Cambridge, Massachusetts) that allows an agent to observe (measure) certain variables when an experiment is carried out. This extended system does allow for learning from experiments. For all the proposed logics, we provide a sound and complete axiomatization.

Keywords: Causal reasoning, epistemic reasoning, interventions, counterfactuals, dynamic epistemic operators, experiments

1 Introduction

In recent years a lot of effort has been put in the development of formal models of causal reasoning. A central motivation for this is the importance of causal reasoning for AI. Making computers take

Vol. 33, No. 6,
Advance Access Publication on 1 April 2022 https://doi.org/10.1093/logcom/exac011

into account causal information is currently one of the central challenges of AI research [11, 37]. There has also been tremendous progress in this direction after the earlier groundbreaking work in [36] and [40]. Advanced formal and computational tools have been developed for modelling causal reasoning and learning causal information, with applications in many different scientific areas.

However, there is one aspect of causal reasoning that has not gotten enough attention yet: the interaction between causal and epistemic reasoning. Even though the standard logical approach to causal reasoning [22, 23, 36] can model epistemic uncertainty,[1] it does not come with an object language that can make statements about the epistemic state of some agent. There are recent proposals adding probabilistic expressions to the object language [e.g. 30], but very little has been done on combining causal and qualitative epistemic reasoning.[2] However, this kind of reasoning occurs frequently in our daily life. Consider, for instance, the following situation.

EXAMPLE 1.1
Sarah got a new flashlight for her birthday. The flashlight emits light only if the button is pushed and the batteries are charged. At the moment the button is not pressed and the light is off. Sarah hasn't tried the flashlight yet. So, she doesn't know whether the battery is full or not. Let us assume that, in fact, the battery is empty.

In such cases, we want to be able to infer that Sarah is not sure whether, had the button been pushed, the flashlight would have emitted light. A logic accounting for such inferences needs a language with statements involving epistemic attitudes towards causal inferences. This will be the topic of this paper.

We are interested here in one particular situation where the interaction between causal and epistemic is essential: experimentation. Our goal is to develop a logical framework that allows us to model and reason about experiments. Experiments play a central role for our survival, because they facilitate learning. We do not only learn by observing; we actively and purposefully interact with the world in order to understand our surroundings better. For instance, in the example given above, Sarah could experiment with the button of the flashlight (press it) in order to find out whether the battery is charged. Studying the laws and conditions that govern learning from experiments will help us to see the possibilities, but also the limits of the experimental method. In this paper we set out to make a first step towards a logical model for this type of reasoning.

The state of the art on causal and epistemic reasoning provides us with all the necessary means to embark on this project. On the one hand, we have well-developed systems of causal reasoning that center on the concept of intervention—which is meant to capture the notion of a (hypothetical) experiment. On the other hand, we have a rich literature on epistemic and dynamic epistemic logics designed explicitly to model an agent's knowledge and the way it changes. The only step missing is an integration of these two logical systems. This is what we set out to do here. More concretely, we will combine the standard approach to causal reasoning [22, 23, 36] with tools from Epistemic Logic [*EL*; 16, 26][3] and Dynamic Epistemic Logic [*DEL*, 4, 43, 44].[4]

[1] For this, add a probability distribution over the exogenous variables of the model. Uncertainty is then restricted to the value of variables. All causal dependencies are deterministic.

[2] See [6] for an exception, although the epistemic element is not made fully explicit in the language considered in that paper.

[3] Other options might be more powerful/expressive [e.g. probabilistic tools, as in 15, and their dynamics extensions, as in 3, 31, 32], yet, *EL* is enough for our purposes.

[4] Different from other options [e.g. epistemic/doxastic temporal logic; 34], *DEL* allows us to represent different epistemic actions (interventions, observations), and also to study the way they interact with one another (see, e.g. the axioms in Table 3).

We will start with building a very simple extension of the standard system of causal reasoning to a semantic framework that can interpret modal statements about the knowledge of an agent.This system will be extended with the dynamic epistemic operator of public announcement. In the context of experimentation, a public announcement can be naturally interpreted as an *observation*.[5] The resulting logic will be able to model reasoning about hypothetical experiments (thought experiments). In particular, it will account for the first inference described in the context of Example 1.1: Sarah is not sure whether, had the button been pushed, the flashlight would have emitted light.

However, this system cannot model learning from interventions. This is reflected in the fact that it validates the so called rule of *No-Learning*: if the agent knows ϕ after an intervention, then before the intervention she knew that intervening would make ϕ true; in symbols,

$$\mathbf{NL}: \quad [\vec{X}=\vec{x}]K\phi \to K[\vec{X}=\vec{x}]\phi.$$

For overcoming this limitation, we will introduce the concept of observables. Observables describe the variables the agent can observe (the experimenter measures). The extended system can account for the inference that, after testing the flashlight, Sarah would know that the batteries need to be charged.

We will proceed as follows. In Section 2 we introduce what we take here to be the basic system of causal reasoning: a slightly generalized form of the logic of causal reasoning introduced by Judea Pearl and collaborators and further developed in [14, 22, 36]. In Section 3 we introduce the notion of an *epistemic causal model* and extend the language with (i) a knowledge operator and (ii) public announcement. Section 4 discusses the shortcomings of this first extension and motivates the new system that will be introduced in Section 5. We finish the paper with the conclusions and an outlook on future work in Section 6. For all systems discussed in the paper, we present a sound and strongly complete axiomatization. For space reasons, the proofs of the main results have been omitted here, but they can be found in this manuscript's full version [10].

2 Causal models

We start by introducing a basic system of causal reasoning. This logic goes back to [18] [based on the work of structural causal models in 35] and was further developed in, among others, [14, 22, 36]. We slightly generalize the framework [along the lines of 14] by allowing for interventions on exogenous variables.[6]

The structure. The starting point is a formal representation of causal dependencies. This is done through *causal models*, which represent the causal relationships between a finite set of variables. The variables are sorted into the set $\mathcal{U}$ of *exogenous* variables (those whose value is causally independent from the value of every other variable in the system) and the set $\mathcal{V}$ of *endogenous* variables (those whose value is completely determined by the value of other variables in the system). Note that $\mathcal{U} \cap \mathcal{V} = \emptyset$. Each variable $X \in \mathcal{U} \cup \mathcal{V}$ is assigned a non-empty and finite range $\mathcal{R}(X)$,

[5] Cf. the *selective implication* of [6, 7].

[6] This choice might be confusing for some readers. Exogenous variables are often seen as variables out of our control; theoretical entities that explain the (causal) behavior of those variables we actually want to model. In this paper we take the position that this is a conceptual choice that should not depend on formal restrictions. Thus, as far as possible exogenous variables will be treated the same way as endogenous variables.

which contains the possible values the variable can take. All this information is provided by a *signature*.

DEFINITION 2.1 (Signature).
Throughout this text, let $\mathscr{S} = \langle \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{R}} \rangle$ be a *finite* signature where

- $\boldsymbol{\mathcal{U}} = \{U_1, \ldots, U_m\}$ is the finite set of *exogenous* variables,
- $\boldsymbol{\mathcal{V}} = \{V_1, \ldots, V_n\}$ is the finite set of *endogenous* variables, and
- $\boldsymbol{\mathcal{R}}(X)$ is the finite non-empty range of X, for X a variable in $\boldsymbol{\mathcal{U}} \cup \boldsymbol{\mathcal{V}}$.[7]

Define the set $\boldsymbol{\mathcal{W}} := \boldsymbol{\mathcal{U}} \cup \boldsymbol{\mathcal{V}}$ and, for $X \in \boldsymbol{\mathcal{W}}$, the abbreviation $\boldsymbol{\mathcal{W}}_{-X} := \boldsymbol{\mathcal{W}} \setminus \{X\}$. For simplicity, tuples of variables will be sometimes manipulated by set operations. To do this properly, we assume a canonical order over $\boldsymbol{\mathcal{W}}$ (say, $\langle U_1, \ldots, U_m, V_1, \ldots, V_n \rangle$), which yields a one-to-one correspondence between a subset of $\boldsymbol{\mathcal{W}}$ and the tuple containing its elements in the canonical order.

A causal model is formally defined as follows.

DEFINITION 2.2 (Causal model).
A *causal model* is a triple $\langle \mathscr{S}, \mathcal{F}, \mathcal{A} \rangle$ where

- $\mathscr{S} = \langle \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{R}} \rangle$ is the model's signature,
- $\mathcal{F} = \{F_{V_j} \mid V_j \in \boldsymbol{\mathcal{V}}\}$ contains, for each endogenous variable $V_j \in \boldsymbol{\mathcal{V}}$, a map

$$F_{V_j} : \boldsymbol{\mathcal{R}}(U_1, \ldots, U_m, V_1, \ldots, V_{j-1}, V_{j+1}, \ldots, V_n) \to \boldsymbol{\mathcal{R}}(V_j).$$

  Each F_V is V's *structural function*.
- $\mathcal{A}$ is a *valuation* function, assigning to every $X \in \boldsymbol{\mathcal{W}}$ a value $\mathcal{A}(X) \in \boldsymbol{\mathcal{R}}(X)$.[8] The valuation must comply with the exogenous variables' structural functions: for each endogenous variable $V_j \in \boldsymbol{\mathcal{V}}$ we have

$$\mathcal{A}(V_j) = F_{V_j}\left(\mathcal{A}(U_1, \ldots, U_m, V_1, \ldots, V_{j-1}, V_{j+1}, \ldots, V_n)\right).$$

In a causal model $\langle \mathscr{S}, \mathcal{F}, \mathcal{A} \rangle$, the functions in $\mathcal{F}$ describe the causal relationship between the variables. Using these functional dependencies, we can define what it means for a variable to directly causally affect another variable.[9]

DEFINITION 2.3 (Causal dependency).
Let $\mathcal{F}$ be a set of structural functions for $\boldsymbol{\mathcal{V}}$. Given $V_j \in \boldsymbol{\mathcal{V}}$, rename the variables in $\boldsymbol{\mathcal{W}}_{-V_j}$ as $X_1, \ldots, X_{m+n-1}$.

Under $\mathcal{F}$, an endogenous variable $V_j \in \boldsymbol{\mathcal{V}}$ is *directly causally affected* by a variable $X_i \in \boldsymbol{\mathcal{W}}_{-V_j}$ (in symbols, $X_i \hookrightarrow_{\mathcal{F}} V_j$) if and only if there is a tuple

$$(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{m+n-1}) \in \boldsymbol{\mathcal{R}}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{m+n-1})$$

[7] Given $(X_1, \ldots, X_k) \in (\boldsymbol{\mathcal{U}} \cup \boldsymbol{\mathcal{V}})^k$, abbreviate $\boldsymbol{\mathcal{R}}(X_1) \times \cdots \times \boldsymbol{\mathcal{R}}(X_k)$ as $\boldsymbol{\mathcal{R}}(X_1, \ldots, X_k)$.

[8] Given $(X_1, \ldots, X_k) \in (\boldsymbol{\mathcal{U}} \cup \boldsymbol{\mathcal{V}})^k$, abbreviate $(\mathcal{A}(X_1), \ldots \mathcal{A}(X_k))$ as $\mathcal{A}(X_1, \ldots, X_k)$.

[9] This notion of a *direct cause*, adopted from [18], is related to the notion *direct effect* of a variable to another, as discussed in [36] (for Causal Bayes Nets). The notion used here differs from Halpern's notion of *affect* [22]. This difference shows up in the axiomatization: axiom (Table 1) has the same function as in [22] (making the canonical model recursive), but does so in a slightly different way.

and there are $x_i' \neq x_i'' \in \boldsymbol{\mathcal{R}}(X_i)$ such that

$$F_{V_j}(x_1, \ldots, x_i', \ldots, x_{m+n-1}) \neq F_{V_j}(x_1, \ldots, x_i'', \ldots, x_{m+n-1}).$$

When $X_i \hookrightarrow_{\mathcal{F}} V_j$, we will also say that X_i is a *causal parent* of V_j. The pair $\langle \boldsymbol{\mathcal{W}}, \hookrightarrow_{\mathcal{F}} \rangle$ is called the *causal graph* induced by $\mathcal{F}$. The relation $\hookrightarrow_{\mathcal{F}}^{+}$ is the transitive closure of $\hookrightarrow_{\mathcal{F}}$.

As it is common in the literature, we restrict ourselves to causal models in which circular causal dependencies do not occur.[10]

DEFINITION 2.4 (Recursive causal model).
A set of structural functions $\mathcal{F}$ is *recursive* if and only if $\hookrightarrow_{\mathcal{F}}^{+}$ is a strict partial order (i.e. an asymmetric and transitive relation).[11] A causal model $\langle \mathscr{S}, \mathcal{F}, \mathcal{A} \rangle$ is *recursive* if and only if $\mathcal{F}$ is recursive. Here, a recursive causal model will be called simply a *causal model*.

An important feature of recursive causal models is that, once the value of all exogenous values is fixed, the value of all endogenous variables is uniquely determined [see, e.g. 22].

Interventions. The most important concept in this approach to causal reasoning is that of *intervention*: an action that changes the values of the system's variables. Before defining it formally, we introduce the notion of assignment.

DEFINITION 2.5 (Assignment, subassignment).
Given $k \in \mathbb{N}$ and a set of variables $\{X_1, \ldots, X_k\} \subseteq \boldsymbol{\mathcal{W}}$, an *assignment* on $\mathscr{S}$ is a function that allocates a value in $\boldsymbol{\mathcal{R}}(X_i)$ to each variable X_i. For simplicity, an assignment will be denoted as $\vec{X} = \vec{x}$, with $\vec{X} = (X_1, \ldots, X_k)$ the assignment's domain (thus, with $X_i \neq X_j$ for $i \neq j$) and $\vec{x} \in \boldsymbol{\mathcal{R}}(\vec{X})$ the tuple containing a value for each variable in $\vec{X}$.[12] An assignment $\vec{X}' = \vec{x}'$ is a *subassignment* of $\vec{X} = \vec{x}$ if and only if $\vec{X}' \subseteq \vec{X}$ and $\vec{x}'$ is the restriction of $\vec{x}$ to the values indicated for variables in $\vec{X}'$.

Given an assignment $\vec{X} = \vec{x}$, an intervention that sets each variable in $\vec{X}$ to its respective value in $\vec{x}$ can be defined as an operation that maps a given causal model M to a new model $M_{\vec{X} = \vec{x}}$.

DEFINITION 2.6 (Intervention).
Let $M = \langle \mathscr{S}, \mathcal{F}, \mathcal{A} \rangle$ be a causal model with $\vec{X} = \vec{x}$ an assignment on $\mathscr{S}$. In the causal model $M_{\vec{X} = \vec{x}} = \langle \mathscr{S}, \mathcal{F}_{\vec{X} = \vec{x}}, \mathcal{A}^{\mathcal{F}}_{\vec{X} = \vec{x}} \rangle$, which results from an intervention setting the values of variables in $\vec{X}$ to $\vec{x}$,

- the functions in $\mathcal{F}_{\vec{X} = \vec{x}} = \{F'_V \mid V \in \boldsymbol{\mathcal{V}}\}$ are such that (i) for each V *not in* $\vec{X}$, the function F'_V is exactly as F_V, and (ii) for each $V = X_i \in \vec{X}$, the function F'_{X_i} is a *constant* function returning the value $x_i \in \vec{x}$ regardless of the values of all other variables.
- $\mathcal{A}^{\mathcal{F}}_{\vec{X} = \vec{x}}$ is the unique valuation (uniqueness will be proved below) where (i) the value of each exogenous variable in $\vec{X}$ is the one in $\vec{x}$, (ii) the value of each exogenous variable not in $\vec{X}$ is as in $\mathcal{A}$ and (iii) each endogenous variable complies with its *new* structural function (that in

[10] This is because only acyclic relations are thought to have a causal interpretation [see, e.g. 41]. Counterfactuals behave differently if cyclic dependencies are allowed [see 22].
[11] Alternatively (but equivalently), $\mathcal{F}$ is *recursive* if and only if $\hookrightarrow_{\mathcal{F}}$ does not contain cycles.
[12] Note how taking $k = 0$ yields the *empty assignment*.

$\mathcal{F}_{\vec{X}=\vec{x}}$). In other words, $\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}(Y)$ is the unique valuation satisfying the following equations:

$$\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}(Y) = \begin{cases} x_i & \text{if } Y \in \boldsymbol{\mathcal{U}} \text{ and } Y = X_i \in \vec{X} \\ \mathcal{A}(Y) & \text{if } Y \in \boldsymbol{\mathcal{U}} \text{ and } Y \notin \vec{X} \\ F'_Y(\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}(\boldsymbol{\mathcal{W}}_{-Y})) & \text{if } Y \in \boldsymbol{\mathcal{V}}. \end{cases}$$

Note that $\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}$ can be equivalently defined as the (again, unique) valuation where (i) the value of each $X_i \in \vec{X}$ is the indicated $x_i \in \vec{x}$, (ii) the value of each exogenous variable not in $\vec{X}$ is exactly as in $\mathcal{A}$ and (iii) the value of each endogenous variable not in $\vec{X}$ complies with its structural function (in $\mathcal{F}$ or in $\mathcal{F}_{\vec{X}=\vec{x}}$, as the functions for endogenous variables not in $\vec{X}$ remain the same). In this reformulation it becomes more transparent that an intervention with empty assignment does not affect the given causal model.

An important observation with respect to Definition 2.6 is that intervention preserves (recursive) causal models.

PROPOSITION 2.7
Let $\vec{X}=\vec{x}$ be an assignment. If M is a (recursive) causal model, then so is $M_{\vec{X}=\vec{x}}$.

PROOF. For showing that $M_{\vec{X}=\vec{x}}$ is a causal model, one needs to show that its valuation $\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}$ complies with the structural functions $\mathcal{F}_{\vec{X}=\vec{x}}$. This is given directly by the third case of the definition of $\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}$. For showing that $M_{\vec{X}=\vec{x}}$ is recursive, simply notice that the intervention only removes causal dependencies, and thus no circular dependencies can be created. □

Lastly, we prove that indeed there is a unique valuation satisfying the constraints imposed by the definition of intervention.

COROLLARY 2.8
In $M_{\vec{X}=\vec{x}}$, the valuation $\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}$ is uniquely determined.

PROOF. In $M_{\vec{X}=\vec{x}}$, the value of every exogenous variable U is fixed, either from $\vec{x}$ (if U occurs in $\vec{X}$) or from $\mathcal{A}$ (otherwise). Then, by $\mathcal{F}_{\vec{X}=\vec{x}}$' recursiveness (Proposition 2.7), each endogenous variable gets a unique value. □

The language. Now, everything is in place for introducing the object language.

DEFINITION 2.9 (Language).
Formulas φ of the language based on the signature $\mathscr{S}$ are given by

$$\varphi ::= Y{=}y \mid \neg\varphi \mid \varphi \wedge \varphi \mid [\vec{X}{=}\vec{x}]\varphi$$

for $Y \in \boldsymbol{\mathcal{W}}$, $y \in \boldsymbol{\mathcal{R}}(Y)$ and $\vec{X}=\vec{x}$ an assignment on $\mathscr{S}$. The expression $[\vec{X}=\vec{x}]\varphi$ should be read as a counterfactual conditional: 'if the variables in $\vec{X}$ were set to the values $\vec{x}$, respectively, then φ would be the case'.[13] The assignment $\vec{X}=\vec{x}$ is called the counterfactual's *antecedent*, and φ is called its *consequent*. Other Boolean operators ($\vee, \rightarrow, \leftrightarrow$) are defined as usual.

[13] Note: [ ]φ expresses that φ holds after an intervention with the empty assignment.

For the semantic interpretation, let $\langle \mathscr{S}, \mathcal{F}, \mathcal{A} \rangle$ be a causal model. Boolean operators are evaluated as usual; for the rest,

$$\langle \mathscr{S}, \mathcal{F}, \mathcal{A} \rangle \models Y{=}y \quad \text{iff} \quad \mathcal{A}(Y) = y,$$

$$\langle \mathscr{S}, \mathcal{F}, \mathcal{A} \rangle \models [\vec{X}{=}\,\vec{x}]\varphi \quad \text{iff} \quad \langle \mathscr{S}, \mathcal{F}_{\vec{X}=\,\vec{x}}, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\,\vec{x}} \rangle \models \varphi.$$

A formula $\varphi \in \mathcal{L}_C$ is valid w.r.t *recursive* causal models (notation: $\models \varphi$) if and only if $M \models \varphi$ for every recursive causal model M.

The language $\mathcal{L}_C$ has more restrictions than that in [14]: while here a counterfactual's antecedent $\vec{X}{=}\,\vec{x}$ is effectively only a conjunction of requirements (set X_1 to x_1 *and* so on), in [14] the antecedent might contain disjunctions. Yet, $\mathcal{L}_C$ has less restrictions than the languages in [22], as it allows complex formulas in a counterfactual's consequent.[14] Note also how, in contrast to most literature on causal models, $\mathcal{L}_C$ can talk about the values of exogenous variables, and also allows interventions on them.[15]

Note also how $\mathcal{L}_C$ is expressive enough to characterize syntactically the notion of direct causal effect $\hookrightarrow$. Indeed, recall (Definition 2.3) that $X \hookrightarrow_{\mathcal{F}} V$ holds in a given $\langle \mathscr{S}, \mathcal{F}, \mathcal{A} \rangle$ if and only if there are values $\vec{z}$ for the variables in $\vec{Z} = \boldsymbol{\mathcal{W}} \setminus \{X, V\}$ and two different values x_1, x_2 for X such that the value V gets from F_V by setting $(\vec{Z}, X)$ to $(\vec{z}, x_1)$ is different from the value it gets from F_V by setting the same variables to $(\vec{z}, x_2)$. This is expressed by the formula

$$\bigvee_{\substack{\vec{z} \in \boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{W}} \setminus \{X, V\}), \\ \{x_1, x_2\} \subseteq \boldsymbol{\mathcal{R}}(X),\ x_1 \neq x_2, \\ \{v_1, v_2\} \subseteq \boldsymbol{\mathcal{R}}(V),\ v_1 \neq v_2}} [\vec{Z}{=}\,\vec{z}, X{=}x_1](V{=}v_1) \ \wedge\ [\vec{Z}{=}\,\vec{z}, X{=}x_2](V{=}v_2),$$

which is abbreviated as $X \rightsquigarrow V$ [cf. with the syntactic definition of causal dependency in 22]. Thus, for any causal model $\langle \mathscr{S}, \mathcal{F}, \mathcal{A} \rangle$ and any variables $X \in \boldsymbol{\mathcal{W}}$ and $V \in \boldsymbol{\mathcal{V}}$,

$$\langle \mathscr{S}, \mathcal{F}, \mathcal{A} \rangle \models X \rightsquigarrow V \qquad \text{if and only if} \qquad X \hookrightarrow_{\mathcal{F}} V$$

Axiom system. As stated, is different from previous languages used for causal models. Thus, it is worthwhile to provide its axiom system. The system, whose axioms and rules appear in Table 1, uses as a parameter the signature $\mathscr{S}$ of both the language and the intended class of models (in some cases relying on the signature's finiteness). Yet, note that no axiom refers to a particular model.

Note the role played by the signature's *finiteness*, i.e. by the fact that there are only a finite number of variables, each one of them having a finite range. Just as in [22], axiom makes use of the second condition to indicate that a variable gets assigned a value in its range and axiom makes use of both to characterize recursive models (via the abbreviation $X \rightsquigarrow V$).[16]

THEOREM 2.10 (Axiom system for $\mathcal{L}_C$).
Recall that $\mathscr{S} = \langle \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{R}} \rangle$ is a finite signature. The axiom system $\mathcal{L}_C$, whose axioms and rules are

[14] Thus, $\mathcal{L}_C$ has also less restrictions than the languages in [17, 18].

[15] Some interventionist approaches to counterfactuals, such as [14] and [7], also allow interventions on exogenous variables.

[16] Still, [22, Page 323] also mentions how to deal with infinite variables with infinite range: in such cases, axiom is required only for variables with finite range, and axiom can be replaced *'by all its instances'*. (This somehow vague suggestion is turned into a concrete rule in [14], although that paper only addresses finite signatures.) The rest of the setting remains unchanged. An alternative approach for dealing with infinite variables and infinite ranges is discussed in [29].

TABLE 1. Axiom system L_C, for $\mathcal{L}_C$-formulas valid in recursive causal models.

P	$\vdash_{\mathsf{L}_C} \varphi$ for φ an instance of a propositional tautology
MP	From φ_1 and $\varphi_1 \to \varphi_2$ infer φ_2
$\mathbf{A_1}$	$\vdash_{\mathsf{L}_C} [\vec{X}{=}\vec{x}]Y{=}y \to \neg[\vec{X}{=}\vec{x}]Y{=}y'$ for $y, y' \in \mathcal{R}(Y)$ with $y \neq y'$
$\mathbf{A_2}$	$\vdash_{\mathsf{L}_C} \bigvee_{y \in \mathcal{R}(Y)} [\vec{X}{=}\vec{x}]Y{=}y$
$\mathbf{A_3}$	$\vdash_{\mathsf{L}_C} \big([\vec{X}{=}\vec{x}](Y{=}y) \wedge [\vec{X}{=}\vec{x}](Z{=}z)\big) \to [\vec{X}{=}\vec{x}, Y{=}y](Z{=}z)$
$\mathbf{A_4}$	$\vdash_{\mathsf{L}_C} [\vec{X}{=}\vec{x}, Y{=}y](Y{=}y)$
$\mathbf{A_5}$	$\vdash_{\mathsf{L}_C} \big([\vec{X}{=}\vec{x}, Y{=}y](Z{=}z) \wedge [\vec{X}{=}\vec{x}, Z{=}z](Y{=}y)\big) \to [\vec{X}{=}\vec{x}](Z{=}z)$ for $Y \neq Z$
$\mathbf{A_6}$	$\vdash_{\mathsf{L}_C} (X_0 \rightsquigarrow X_1 \wedge \cdots \wedge X_{k-1} \rightsquigarrow X_k) \to \neg(X_k \rightsquigarrow X_0)$
$\mathbf{A_7}$	$\vdash_{\mathsf{L}_C} [\,]U{=}u \leftrightarrow [\vec{X}{=}\vec{x}]U{=}u$ for $U \in \mathcal{U}$ with $U \notin \vec{X}$
$\mathbf{A_{[]}}$	$\vdash_{\mathsf{L}_C} Y{=}y \leftrightarrow [\,]Y{=}y$
$\mathbf{A_\neg}$	$\vdash_{\mathsf{L}_C} [\vec{X}{=}\vec{x}]\neg\varphi \leftrightarrow \neg[\vec{X}{=}\vec{x}]\varphi$
$\mathbf{A_\wedge}$	$\vdash_{\mathsf{L}_C} [\vec{X}{=}\vec{x}](\varphi_1 \wedge \varphi_2) \leftrightarrow ([\vec{X}{=}\vec{x}]\varphi_1 \wedge [\vec{X}{=}\vec{x}]\varphi_2)$
$\mathbf{A_{[][]}}$	$\vdash_{\mathsf{L}_C} [\vec{X}{=}\vec{x}][\vec{Y}{=}\vec{y}]\varphi \leftrightarrow [\vec{X}'{=}\vec{x}', \vec{Y}{=}\vec{y}]\varphi$ with $\vec{X}'{=}\vec{x}'$ the subassignment of $\vec{X}{=}\vec{x}$ for $\vec{X}' := \vec{X} \setminus \vec{Y}$

shown in Table 1, is sound and strongly complete for the language $\mathcal{L}_C$ based on $\mathscr{S}$ with respect to recursive causal models for $\mathscr{S}$.

PROOF. See Appendix A.1. □

3 Epistemic causal models

Section 2 introduced the logic of causal reasoning that we will work with. To this logic we will now add an epistemic component. We first extend the causal model with a representation of the epistemic state of an agent. It is assumed that, while the agent knows the causal laws, she might not know the value of some variables. This uncertainty can be represented in the form of a set of valuations $\mathcal{T}$, which represents the alternatives the agent considers possible.[17] To model the assumption that the agent knows the causal laws we require that all possible valuations comply with the same underlying set of structural functions.

DEFINITION 3.1 (Epistemic causal model).
An *epistemic (note: recursive) causal model* $\mathcal{E}$ is a tuple $\langle \mathscr{S}, \mathcal{F}, \mathcal{T} \rangle$ where $\mathscr{S} = \langle \mathcal{U}, \mathcal{V}, \mathcal{R} \rangle$ is a signature, $\mathcal{F}$ is a (note: recursive) set of structural functions for $\mathcal{V}$, and $\mathcal{T}$ is a non-empty set of valuation functions for $\mathcal{U} \cup \mathcal{V}$, each one of them complying with $\mathcal{F}$.[18]

[17] The choice of the name $\mathcal{T}$ is made for analogy with causal team semantics [6, 7], where a set of valuations is called a 'team'. The way we use the 'team' here is purely modal: the formulas take truth values at single valuations, and $\mathcal{T}$ is the set of 'worlds' accessible from this valuation. In team semantics this local perspective is absent.

[18] A remark. In standard *EL*, an epistemic possibility is a world, and each world gets assigned an atomic valuation; hence, the same atomic valuation might be assigned to two different worlds in the agent's epistemic range. In our setting, an epistemic possibility is a valuation; hence, the same valuation cannot appear twice in the set $\mathcal{T}$. Because of this, and given the finiteness of $\mathcal{W}$, each epistemic possibility (valuation) $\mathcal{A}$ can be characterized by the conjunction $\bigwedge_{X \in \mathcal{W}} X = \mathcal{A}(X)$; still, not every such formula characterizes an epistemic possibility, as not all valuations need to be considered possible. (cf. the *nominals* in hybrid logic [1; 13]: in named models, each epistemic possibility (world) is characterized by a nominal, and every nominal names a unique epistemic possibility.)

Using this new notion of an epistemic causal model we can now formalize the context of Example 1.1 from the introduction. We define an epistemic causal model E $= \langle \mathscr{S}, \mathcal{F}, \mathcal{T} \rangle$ whose signature $\mathscr{S}$ has three variables: the exogenous variables P for pushing the button and B for the batteries being full, and the endogenous variable L for the flashlight emitting light. All three variables can take two values, 0 or 1. The set of functions $\mathcal{F}$ contains only one element: the function mapping L to 1 iff $P = 1$ and $B = 1$. It seems natural to assume that the agent is aware of the state of the button (not pressed) and the lamp (off), so the set $\mathcal{T}$ contains the valuation $\mathcal{A}_1$ that maps B to 0, P to 0 and L to 0, and the valuation $\mathcal{A}_2$ that maps B to 1, P to 0 and L to 0. Note how $\mathcal{T}$ cannot contain the valuation $B = 0$, $P = 1$ and $L = 1$, for instance, because this valuation does not comply with the causal law in $\mathcal{F}$. This observation highlights an important assumption we made above: there is no uncertainty about the causal dependencies. Investigating the consequences of lifting this restriction is left for future research.

The next step is to extend the notion of intervention to epistemic causal models, see Definition 3.2.

DEFINITION 3.2 (Intervention on epistemic causal models).
Let $\mathcal{E} = \langle \mathscr{S}, \mathcal{F}, \mathcal{T} \rangle$ be an epistemic causal model; let $\vec{X} = \vec{x}$ be an assignment on $\mathscr{S}$. The epistemic causal model $\mathcal{E}_{\vec{X}=\vec{x}} = \langle \mathscr{S}, \mathcal{F}_{\vec{X}=\vec{x}}, \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}} \rangle$, resulting from an intervention setting the values of variables in $\vec{X}$ to $\vec{x}$, is such that

- $\mathcal{F}_{\vec{X}=\vec{x}}$ is defined from $\mathcal{F}$ just as in Definition 2.6,
- $\mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}} := \{\mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}} \mid \mathcal{B} \in \mathcal{T}\}$ (see Definition 2.6).

In this definition, $\langle \mathscr{S}, \mathcal{F}_{\vec{X}=\vec{x}}, \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}} \rangle$ is indeed an epistemic causal model: $\mathcal{F}_{\vec{X}=\vec{x}}$ is recursive and all valuations in $\mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}}$ comply with it. With this we can now calculate the effects of an intervention $P = 1$ on the epistemic causal model E defined above. According to Definition 3.2, an intervention on an epistemic causal model amounts to intervening on each of the valuations contained in the epistemic state. Thus, for our concrete example, we need to calculate the effects of an intervention with $P = 1$ on the valuations $\mathcal{A}_1$ and $\mathcal{A}_2$ that make up the epistemic state $\mathcal{T}$. It is not difficult to see that the new epistemic state $\mathcal{T}^{\mathcal{F}}_{P=1}$ will contain the valuation $\mathcal{A}^{\mathcal{F}}_{1,P=1}$ that maps B to 0, P to 1 and L to 0 and the valuation $\mathcal{A}^{\mathcal{F}}_{2,P=1}$ that maps B to 1, P to 1 and L to 1.

An important assumption underlying this definition of intervention is that the agent has full epistemic access to the effect of the intervention on the model. In particular, she knows that the intervention takes place (in the counterfactual scenario considered). This is reasonable if one thinks of the agent whose epistemic state is being modelled as the one engaging in counterfactual thinking. It is less plausible in connection to counterfactual thinking about the knowledge states of other agents. But this is something that we can leave for now, as we are not considering epistemic causal models for multiple agents in this paper.

Based on these changes on the semantic side, we can extend the object language with a modality for talking about the epistemic state of the agent.

DEFINITION 3.3 (Language $\mathcal{L}_{KC}$).
Formulas ξ of the language based on $\mathscr{S}$ are given by

$$\xi ::= Y{=}y \mid \neg\xi \mid \xi \wedge \xi \mid K\xi \mid [\vec{X}{=}\vec{x}]\xi$$

for $Y \in \boldsymbol{\mathcal{W}}$, $y \in \boldsymbol{\mathcal{R}}(Y)$ and $\vec{X} = \vec{x}$ an assignment on $\mathscr{S}$. Formulas of the form $K\xi$ are read as 'the agent knows ξ'.

The semantics for this epistemic language is straightforward.

DEFINITION 3.4
Formulas in are evaluated in $\mathcal{L}_{KC}$ a pair $(\mathcal{E}, \mathcal{A})$ with $\mathcal{E} = \langle \mathscr{S}, \mathcal{F}, \mathcal{T} \rangle$ an epistemic causal model and $\mathcal{A} \in \mathcal{T}$. The semantic interpretation for Boolean operators is the usual; for the rest,

$$
\begin{array}{lll}
(\mathcal{E}, \mathcal{A}) \models Y{=}y & \text{iff} & \mathcal{A}(Y) = y, \\
(\mathcal{E}, \mathcal{A}) \models [\vec{X}{=}\vec{x}]\xi & \text{iff} & (\mathcal{E} : \vec{X}{=}\vec{x}, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \xi, \\
(\mathcal{E}, \mathcal{A}) \models K\xi & \text{iff} & (\mathcal{E}, \mathcal{B}) \models \xi \text{ for every } \mathcal{B} \in \mathcal{T}.
\end{array}
$$

To illustrate this definition, we go back to the epistemic model E introduced for Example 1.1. For evaluating a concrete formula with respect to this model we need to select, next to E, a valuation representing the actual world. In the example this is valuation $\mathcal{A}_1$: in the actual world, the batteries are empty. We can calculate that the counterfactual $[P{=}1]L{=}0$ comes out as true given E and $\mathcal{A}_1$, just as in the non-epistemic approach (Section 2). But because we now also have a representation of the epistemic state of some agent, we can additionally consider epistemic attitudes the agent has towards this counterfactual. For instance, we can check that $K([P{=}1]L{=}0)$ is not true given E and $\mathcal{A}_1$: the agent cannot predict the outcome of pressing the button in this situation—this was one of the inferences we discussed in the introduction. This sentence is only true in case the formula $[P{=}1]L{=}0$ holds in both $(\mathrm{E}, \mathcal{A}_1)$ and $(\mathrm{E}, \mathcal{A}_2)$ (where $\mathcal{A}_1, \mathcal{A}_2$ are the two elements of $\mathcal{T}$). However, while in $\mathcal{A}^{\mathcal{F}}_{2,P=1}$ the flashlight is emitting light (in this possibility the battery is charged), this is not the case in $\mathcal{A}^{\mathcal{F}}_{1,P=1}$. Thus, $(\mathrm{E}_{P=1}, \mathcal{A}^{\mathcal{F}}_{2,P=1}) \not\models L{=}0$ and, hence, $(\mathrm{E}, \mathcal{A}_1) \not\models K([P{=}1]L{=}0)$. Thus, the agent cannot predict whether pressing the button of the flashlight will turn on the light, just as intended in this case.

Axiom system. To get a clearer idea of the laws governing the interaction between knowledge and causality in the proposed system, we can look for an axiom system characterizing the formulas in $\mathcal{L}_{KC}$ that are valid in epistemic (recursive) causal models. A sound and complete axiom system for the proposed logic is given by the axiom system L_{KC}, which extends L_C (Table 1) with the axioms and rules from Table 2; thus, once again, it is relative to a given signature $\mathscr{S}$. Among the new axioms, those in the *epistemic* part form the standard modal S5 axiomatization for truthful knowledge with positive and negative introspection [see, e.g. 16]. Then, axiom indicates that what the agent will know after an intervention ($[\vec{X}{=}\ \vec{x}]K\xi$) is exactly what she knows now about the effects of the intervention ($K[\vec{X}{=}\ \vec{x}]\xi$). Although maybe novel in the literature on causal models, the axiom is simply an instance of the more general pattern of interaction between knowledge and a deterministic action without precondition. Axiom indicates that the agent knows how each endogenous variable $Y \in \boldsymbol{\mathcal{V}}$ is affected when *all other variables* are intervened; it can be understood as stating that the agent knows the causal laws.

THEOREM 3.5 (Axiom system for $\mathcal{L}_{KC}$).
Recall that $\mathscr{S} = \langle \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{R}} \rangle$ is a finite signature. The axiom system L_{KC}, extending L_C (Table 1) with the axioms and rules in Table 2, is sound and strongly complete for the language $\mathcal{L}_{KC}$ based on $\mathscr{S}$ with respect to epistemic (recursive) causal models for $\mathscr{S}$.

PROOF. See Appendix A.2. □

Adding public announcement So far the extension of the logic of causal inference that we introduced is completely static from the epistemic point of view: there is no action that can

TABLE 2. Additional axioms and rules for axiom system L_{KC}, for $\mathcal{L}_{KC}$-formulas valid in epistemic (recursive) causal models.

Epistemic:			
K	$\vdash_{\mathsf{L}_{KC}} K(\xi_1 \to \xi_2) \to (K\xi_1 \to K\xi_2)$	**T**	$\vdash_{\mathsf{L}_{KC}} K\xi \to \xi$
$\mathbf{N_K}$	From $\vdash_{\mathsf{L}_{KC}} \xi$ derive $\vdash_{\mathsf{L}_{KC}} K\xi$	**4**	$\vdash_{\mathsf{L}_{KC}} K\xi \to KK\xi$
5	$\vdash_{\mathsf{L}_{KC}} \neg K\xi \to K\neg K\xi$		
Epistemic+Intervention:			
CM	$\vdash_{\mathsf{L}_{KC}} [\vec{X}=\vec{x}]K\xi \leftrightarrow K[\vec{X}=\vec{x}]\xi$		
KL	$\vdash_{\mathsf{L}_{KC}} [\vec{X}=\vec{x}]Y=y \to K[\vec{X}=\vec{x}]Y=y$ for $Y \in \boldsymbol{\mathcal{V}}$ and $\vec{X} = \boldsymbol{\mathcal{W}} \setminus \{Y\}$		

induce a change in the information state of an agent. But this can be easily amended. For instance, we can 'import' the well-known action of *(truthful) public announcements* [20, 39], which in this single-agent setting can be understood as an act of observing/learning. This action corresponds, semantically, to a model operation that removes the epistemic possibilities in which the announced/observed/learnt formula does not hold. We first add public announcements to the syntax of the language and then provide a semantics for this extended language.

DEFINITION 3.6 (Language $\mathcal{L}_{PAKC}$).
Formulas χ of the language $\mathcal{L}_{PAKC}$ based on $\mathscr{S}$ are given by

$$\chi ::= Y{=}y \mid \neg\chi \mid \chi \wedge \chi \mid K\chi \mid [\chi!]\chi \mid [\vec{X}{=}\vec{x}]\chi$$

for $Y \in \boldsymbol{\mathcal{W}}$, $y \in \boldsymbol{\mathcal{R}}(Y)$ and $\vec{X}{=}\vec{x}$ an assignment on $\mathscr{S}$. Formulas of the form $[\alpha!]\chi$ are read as 'after α is observed, χ is the case'.

DEFINITION 3.7 (Semantics with announcements).
Formulas in $\mathcal{L}_{PAKC}$ are evaluated in a pair $(\mathcal{E}, \mathcal{A})$ with $\mathcal{E} = \langle \mathscr{S}, \mathcal{F}, \mathcal{T} \rangle$ an epistemic causal model and $\mathcal{A} \in \mathcal{T}$. Operators already in $\mathcal{L}_{KC}$ are evaluated as before (Definition 3.4); for announcements,

$$(\mathcal{E}, \mathcal{A}) \models [\alpha!]\chi \qquad \text{iff} \qquad (\mathcal{E}, \mathcal{A}) \models \alpha \text{ implies } (\mathcal{E}^{\alpha}, \mathcal{A}) \models \chi,$$

where $\mathcal{E}^{\alpha} = \langle \mathscr{S}, \mathcal{F}, \mathcal{T}^{\alpha} \rangle$ is such that $\mathcal{T}^{\alpha} := \{\mathcal{B} \in \mathcal{T} \mid (\mathcal{E}, \mathcal{B}) \models \alpha\}$.

Note how $\mathcal{E}^{\alpha}$ is indeed an epistemic causal model: since $\mathcal{E}$ is an epistemic causal model, $\mathcal{F}$ is recursive and all valuations in $\mathcal{T}^{\alpha}$ comply with it.

Let us go back to Example 1.1 to illustrate the working of these definitions. Above we modelled the scenario using an epistemic causal model E and a designated assignment $\mathcal{A}_1$ representing the actual world. Let us assume that in the given context Sarah is told that the counterfactual $[P{=}1]L{=}0$ is true (public announcement). In this case the model correctly predicts that with this new information Sarah can conclude that the batteries are empty, i.e. $(\mathrm{E}, \mathcal{A}_1) \models [[P{=}1]L = 0!]K(B{=}0)$. Notice first that $(\mathrm{E}, \mathcal{A}_1) \models [P{=}1]L{=}0$. So, we only need to check if $(\mathrm{E}^{[P=1]L=0}, \mathcal{A}_1) \models K(B{=}0)$, i.e. whether for all assignments $\mathcal{A} \in \mathcal{T}^{[P=1]L=0}$, $\mathcal{A}(B) = 0$. Recall that $\mathcal{T} = \{\mathcal{A}_1, \mathcal{A}_2\}$ where $\mathcal{A}_1$ describes the scenario in which the battery is empty, the button not pressed and the light is off, and $\mathcal{A}_2$ describes the scenario where the battery is charged, the button not pressed and the light is off. Only with respect to the first assignment $\mathcal{A}_1$ the counterfactual $[P{=}1]L{=}0$ is true. This means,

TABLE 3. Additional axioms and rules for axiom system L_{PAKC}, for $\mathcal{L}_{PAKC}$-formulas valid in epistemic (recursive) causal models.

$\mathbf{N}_{=}$	From $\vdash_{\mathsf{L}_{PAKC}} \chi$ derive $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}{=}\vec{x}]\chi$		
$\mathbf{!}_{=}$	$\vdash_{\mathsf{L}_{PAKC}} [\alpha!][\vec{X}{=}\vec{x}]Y{=}y \leftrightarrow (\alpha \to [\vec{X}{=}\vec{x}]Y{=}y)$		
$\mathbf{!}_{\neg}$	$\vdash_{\mathsf{L}_{PAKC}} [\alpha!]\neg\chi \leftrightarrow (\alpha \to \neg[\alpha!]\chi)$	$\mathbf{!}_{\wedge}$	$\vdash_{\mathsf{L}_{PAKC}} [\alpha!](\chi \wedge \chi) \leftrightarrow ([\alpha!]\chi \wedge [\alpha!]\chi)$
$\mathbf{!}_{\mathbf{K}}$	$\vdash_{\mathsf{L}_{PAKC}} [\alpha!]K\chi \leftrightarrow (\alpha \to K(\alpha \to [\alpha!]\chi))$	$\mathbf{!}_{!}$	$\vdash_{\mathsf{L}_{PAKC}} [\alpha_1!][\alpha_2!]\chi \leftrightarrow [\alpha_1 \wedge [\alpha_1!]\alpha_2!]\chi$
$\mathbf{K}_{!}$	$\vdash_{\mathsf{L}_{PAKC}} [\alpha!](\chi_1 \to \chi_2) \to ([\alpha!]\chi_1 \to [\alpha!]\chi_2)$	$\mathbf{N}_{!}$	From $\vdash_{\mathsf{L}_{PAKC}} \chi$ derive $\vdash_{\mathsf{L}_{PAKC}} [\alpha!]\chi$
$_{!}\mathbf{RE}$	From $\vdash_{\mathsf{L}_{PAKC}} \alpha_1 \leftrightarrow \alpha_2$ derive $\vdash_{\mathsf{L}_{PAKC}} [\alpha_1!]\chi \leftrightarrow [\alpha_2!]\chi$		
$=_{!}$	$\vdash_{\mathsf{L}_{PAKC}} [\vec{X}{=}\vec{x}][\alpha!]\chi \leftrightarrow [[\vec{X}{=}\vec{x}]\alpha!][\vec{X}{=}\vec{x}]\chi$		

following Definition 3.6, that $\mathrm{E}^{[P=1]E=0} = \{\mathcal{A}_1\}$. Furthermore, we have $\mathcal{A}_1(B) = 0$. Therefore, following Definition 3.4, $(\mathrm{E}, \mathcal{A}_1) \models [[P{=}1]L{=}0!]K(B{=}0)$, just as intended.[19]

The languages $\mathcal{L}_{KC}$ and $\mathcal{L}_{PAKC}$ introduced in this section have many similarities to the $\mathcal{CO}$ languages proposed in [6]; these also extend the basic causal language with an operator ('selective implication'), which plays the role of an observation or announcement operator. However, the $\mathcal{CO}$ languages are not given a modal interpretation: they are interpreted in a variant of *team semantics* [28, 42], which makes a comparison of the two languages a difficult task. In [9] we have shown that, in the special case of *recursive* models of *finite* signature, there is a truth-preserving translation from $\mathcal{CO}$ to $\mathcal{L}_{PAKC}$.[20] The translation nicely brings out the point that team semantics interprets many formulas as preceded by a K operator, even though this operator is not explicitly present in $\mathcal{CO}$.

Axiom system. We also provide a sound and complete axiomatization for this extension with public announcements. The axiom system $\mathcal{L}_{PAKC}$ (relative to the signature $\mathscr{S}$) extends $\mathcal{L}_{KC}$ (Tables 1 and 2) with the axioms and rules from Table 3. In the latter table, the rule in the first block is simply the necessitation rule for interventions; this rule was admissible in but it is not anymore in the presence of announcement operators. The axioms and rules in the second block are well-known from public announcement logic; they essentially define a translation that eliminates the observation operator $[\alpha!]$. [See 46 for a discussion on their role in the completeness proof.] The lone axiom in the third block, indicating a form of commutativity between interventions and observations, is what makes the translation work for our language, which makes free use of the intervention operators.[21]

THEOREM 3.8 (Axiom system for $\mathcal{L}_{PAKC}$).
Recall that $\mathscr{S} = \langle \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{R}} \rangle$ is a finite signature. The axiom system L_{PAKC}, extending Tables 1 and 2 with the axioms and rules on Table 3, is sound and strongly complete for the language $\mathcal{L}_{PAKC}$ based on $\mathscr{S}$ with respect to epistemic (recursive) causal models for $\mathscr{S}$.

PROOF. See Appendix A.3. □

[19] This example illustrates an interesting aspect of the system introduced in this section: it allows us to model a restricted form of update with counterfactual information. The only kind of information an agent can learn from counterfactuals is information about the value of so far unknown variables. In case the counterfactual that is announced is inconsistent with the given causal dependencies and the setting of independent variables, the update breaks down.

[20] The translation works also if $\mathcal{CO}$ is extended with *dependence atoms* [42].

[21] Recall that, in contrast with the early literature on interventionist counterfactuals [18, 22], here we allow nesting in the consequents of counterfactuals.

4 Limitations of the system: the 'no learning' assumption

Section 3 introduced an epistemic extension for the standard structural equational model of causal inference. We also saw that this extension can successfully account for some intuitive valid inferences concerning the interaction of causal and epistemic reasoning. However, when introducing Example 1.1, we also discussed inferences that this extension cannot yet account for. Consider, for instance, the consequences of pressing the button. Such an action would increase Sarah's knowledge about the world: given that we assumed that she has epistemic access to the button and the light, she would be able to observe that the light stays off and conclude that the battery is empty. This is an example for reasoning about actual experiments, the kind of reasoning that this paper attempts to model. We can see clearly here how this experiment can increase the knowledge of our agent.

However, the logic introduced in Section 3 cannot account for this type of reasoning; in other words, the formal counterpart of the inference we just described, $[P{=}1]K(B{=}1)$, is not valid in the model E from the example. The underlying reason for this is axiom of $\mathsf{L}_{KC}/\mathsf{L}_{PAKC}$ (see Table 2 on Page 28). More specifically the left-to-right direction of this rule is causing trouble here.

$$\textbf{NL}\quad [\vec{X}{=}\vec{x}]K\phi \rightarrow K[\vec{X}{=}\vec{x}]\phi.$$

In the literature, the principle is known as the 'no learning' property. In the language of dynamic epistemic logic, the property of 'no learning' can be characterized by the formula $\bigcirc_{\pi} K\phi \rightarrow K\bigcirc_{\pi}\phi$, where $\bigcirc_{\pi}\phi$ stands for 'after executing an action π, ϕ will be the case'. In our case the relevant action is an intervention. Now, notice that by contraposition and by using axiom we can derive from this sentence the principle $\neg K[\vec{X}{=}\vec{x}]\phi \rightarrow [\vec{X}{=}\vec{x}]\neg K\phi$: if an agent cannot predict whether an intervention results in a certain outcome, then the agent will also not know this outcome after the intervention is actually performed. Intuitively, this means that an agent is unable to gain any new knowledge from doing an intervention.[22] The example above points out that should not be valid for the type of action that we are considering here. The entire point of experimentation is to provide us with new information. Thus, we need to look for a variation of the proposed logic that doesn't validate.

How can we enable learning from experimentation in our logic? For this we need to change how an intervention affects the knowledge state of an agent. The way the system is set up at the moment will warrant that the agent knows about the action itself (the intervention), but not necessarily about its consequences. These are computed on the epistemic state of the agent completely independently from what happens in the actual world. The intervention affects the knowledge state of the agent as if she *hypothetically considers* the intervention at the same time it is taking place. This is why there is no learning. Just *thinking* about an experiment will teach you nothing about the outside world. You need to *observe* the effects.

To make learning possible, we need to make sure that the value of those variables the agent has epistemic access to (the variables measured in the experiment) is updated when change occurs. We will call these variables *observables*. Thus, we need to make sure that, when an intervention occurs, the agent learns the value the observables take in the changed world. This, together with other knowledge the agent might have, determines the inferences she can draw.

Another way to put it is that we need to implement *knowing the value of a variable* differently. In the system introduced in Section 3 we implement a basic Hintikka-style notion of knowledge: knowledge is simply true belief, which is known to be problematic (e.g. the Gettier cases). One very attractive alternative comes from Nozick [33]. He proposes that an agent knows ψ if the sentence is

[22] See also [25]: 'no learning' expresses that uncertainty cannot be reduced when an action is executed.

true and if one's belief in it was acquired by a method such that if the sentence ψ were not true, the method would not lead one to believe ψ, and such that if ψ were true, one would believe ψ. In our case this method is observation (or whatever method the experimenter uses to measure the value of the observable variables). Reliance on such a method ensures that change in the world will translate directly into change in the epistemic state of the agent, which is exactly what we are after here.

5 Semantics with observables

In the previous section we saw that the framework introduced in Section 3 does not represent correctly the increase of knowledge produced by actual experiments. As explained above, in order to overcome this limitation, we will extend our notion of model with a set of observables: variables to which the agent has direct epistemic access. The semantics we propose will enforce knowledge of the values of the observables by guaranteeing that the observables take a constant value in a model. Furthermore, we will adapt the notion of intervention to ensure that it preserves this property. But first we need to enrich the notion of a signature so that it singles out a set of observables.[23]

DEFINITION 5.1 (Signature and observables).
A *signature* with observables is a quadruple $\langle \mathcal{U}, \mathcal{V}, \mathcal{O}, \mathcal{R} \rangle$ where $\langle \mathcal{U}, \mathcal{V}, \mathcal{R} \rangle$ is as in Definition 2.1 and $\mathcal{O} \subseteq \mathcal{U} \cup \mathcal{V}$. The set $\mathcal{O}$ is called the set of *observables*.

Next we introduce the notion of an *epistemic causal model with observables*. This is an epistemic causal model with one extra property: the agent knows at least the value of all observables. This is warranted by the final condition of the following definition.

DEFINITION 5.2 (Epistemic causal model with observables).
Consider a signature $\mathscr{S} = \langle \mathcal{U}, \mathcal{V}, \mathcal{O}, \mathcal{R} \rangle$. An *epistemic causal model with observables* is a triple $\langle \mathscr{S}, \mathcal{F}, \mathcal{T} \rangle$ such that $\langle \langle \mathcal{U}, \mathcal{V}, \mathcal{R} \rangle, \mathcal{F}, \mathcal{T} \rangle$ is an epistemic causal model and, for each $\mathcal{A}, \mathcal{A}' \in \mathcal{T}$, $\mathcal{A}(\mathcal{O}) = \mathcal{A}'(\mathcal{O})$.

Now we modify the definition of intervention. As mentioned, the important difference in the new notion of intervention is that now the agent has epistemic access to the effects of the intervention on the observable variables in the actual world. This is stated by the last condition in Definition 5.3. Based on this update of our basic definitions we define a new semantics 'with observables' for the language $\mathcal{L}_{PAKC}$ in Definition 5.4.

DEFINITION 5.3 (Interventions on models with observables).
Let $\mathcal{E}$ be an epistemic causal model with observables $\langle \mathscr{S}, \mathcal{F}, \mathcal{T} \rangle$ for $\mathscr{S} = \langle \mathcal{U}, \mathcal{V}, \mathcal{O}, \mathcal{R} \rangle$; take $\mathcal{A} \in \mathcal{T}$. The epistemic causal model $\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}} = \langle \mathscr{S}, \mathcal{F}_{\vec{X}=\vec{x}}, \mathcal{T}^{\mathcal{F},\mathcal{A}}_{\vec{X}=\vec{x}} \rangle$, resulting from an intervention setting the values of variables in $\vec{X}$ to $\vec{x}$, is such that

- $\mathcal{F}_{\vec{X}=\vec{x}}$ is as in Definition 3.2, and
- $\mathcal{B} \in \mathcal{T}^{\mathcal{F},\mathcal{A}}_{\vec{X}=\vec{x}}$ iff $\mathcal{B} \in \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}}$ and, for all $O \in \mathcal{O}$, $\mathcal{B}(O) = \mathcal{A}_{\vec{X}=\vec{x}}(O)$.

[23]As explained in Footnote 2, we look for a formal system that is as general (and thus as flexible) as possible, without introducing restrictions (on exogenous variables) that are not necessary. Thus, Definition 5.1 treats all variables as potentially observable, even though that might not be how exogenous variables are normally conceptualized.

DEFINITION 5.4 (Semantics with observables).
The formulas of $\mathcal{L}_{PAKC}$ are evaluated over pairs $(\mathcal{E},\mathcal{A})$, where $\mathcal{E} = \langle \mathscr{S}, \mathcal{F}, \mathcal{T}\rangle$ is an epistemic causal model with observables and $\mathcal{A} \in \mathcal{T}$. The satisfaction relation $(\mathcal{E},\mathcal{A}) \models \varphi$ is defined by the inductive clauses given in Definition 3.4, with the following exception:

$$(\mathcal{E},\mathcal{A}) \models [\vec{X}{=}\vec{x}]\gamma \qquad \text{iff} \qquad (\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}}, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \gamma.$$

Wherever there is need to distinguish between this semantics with observables and the one in Section 3, we will write $\models^O$ for the satisfaction relation in the semantics with observables and $\models^W$ for the earlier semantics without observables. We point out, however, that the semantics $\models^W$ can be seen a special case of $\models^O$: the case when the set $\mathcal{O}$ of observables is empty.

Let us illustrate how this extended system works by going back to the inference discussed in Section 4 in the context of Example 1.1: we will show that in the extended system pushing the button will tell the agent that the battery of the flashlight is empty. In other words, we show that $[P{=}1]K(B{=}0)$ holds in the epistemic causal model $\mathrm{E} = \langle \mathscr{S}, \mathcal{F}, \mathcal{T}\rangle$ extended with the set of observables $\{P, L\}$ (i.e. we assume that the agent can observe the button and the light, but not whether the batteries of the flashlight are charged).

$\langle \mathscr{S}, \mathcal{F}, \mathcal{T}\rangle, \mathcal{A}_1 \models^O [P{=}1]K(B{=}0)$ is the case if for each assignment $\mathcal{B}$ in the resulting knowledge state after pushing the button ($\mathcal{T}^{\mathcal{F},\mathcal{A}_1}_{P=1}$) the battery is empty ($B = 0$). So, let's calculate the resulting knowledge state. We know that $\mathcal{T} = \{\mathcal{A}_1, \mathcal{A}_2\}$, where $\mathcal{A}_1$ is the possibility that the battery is empty, the button not pressed and the light is off, and $\mathcal{A}_2$ is the same possibility, except with a full battery. After the intervention we get $\mathcal{T}^{\mathcal{F}}_{P=1} = \{\mathcal{A}^{\mathcal{F}}_{1,P=1}, \mathcal{A}^{\mathcal{F}}_{2,P=1}\}$ where in $\mathcal{A}^{\mathcal{F}}_{1,P=1}$ the light is still off (because the battery is empty), but in $\mathcal{A}^{\mathcal{F}}_{2,P=1}$ the light is on. Now, we have to check which of these possibilities matches what the agent actually observes. In this case the agent can see the button and the light, and in the actual world $\mathcal{A}_1$ pressing the button will not turn on the light. Then, the possibility $\mathcal{A}^{\mathcal{F}}_{2,P=0}$ will be removed from her knowledge state; only the possibility $\mathcal{A}^{F}_{1,P=0}$ remains. In this world the battery is empty. Thus after the experiment described by $[P = 0]$ the agent knows that the battery is empty.

This example also immediately illustrates that this new semantics does not validate the rule **NL** (*no learning*). We have just seen that $\langle \mathscr{S}, \mathcal{F}, \mathcal{T}\rangle, \mathcal{A}_1 \models^O [P{=}1]K(B{=}0)$. However, $\langle \mathscr{S}, \mathcal{F}, \mathcal{T}\rangle \not\models^O K[P{=}1](B{=}0)$ (i.e. hypothetically considering the intervention of pushing the button will not make the agent believe that the battery is empty). For this sentence to be true in the given model we need that, in all the possibilities in the agent's knowledge state $\mathcal{T}$ after pressing the button, the battery is empty. However, this is not true for the possibility $\mathcal{A}_2 \in \mathcal{T}$. In this assignment the battery is charged, the button not pushed and the lamp is off. After pressing the button the battery will still be charged, i.e. $B = 0$ will not hold. Thus, the left-to-right direction of the axiom **CM** from the deduction system L_{PAKC} is not valid in the semantics with observables.

However, the opposite direction, **PR**, is still valid.

$$\textbf{PR} \quad K[\vec{X} = \vec{x}]\phi \to [\vec{X} = \vec{x}]K\phi.$$

PR is an instance of the principle known as *perfect recall* in the context of dynamic logics. In its general formulation, this principle asserts that if the outcome of an action (in our case, an intervention) is known before the action is performed, this piece of information is not forgotten after the action takes place. The reason why this principle still holds is that throughout this paper we have always assumed that all interventions are public knowledge. In other words, the value that intervened variables take after an intervention is always known by an agent.[24] Therefore, any changes

[24] We will come back to this assumption in Section 6.

TABLE 4. Additional rules for L^{O}_{PAKC}.

OC	$[\vec{X}=\vec{x}]\bigvee_{\vec{o}\in\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{O}})} K\boldsymbol{\mathcal{O}}=\vec{o}$, for each $O\in\boldsymbol{\mathcal{O}}$
PD	$[\vec{X}=\vec{x}]K\psi \leftrightarrow \bigvee_{\vec{o}\in\boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{O}})}\Big([\vec{X}=\vec{x}]\mathcal{O}=\vec{o}\wedge\big[[\vec{X}=\vec{x}]\mathcal{O}=\vec{o}!\,\big]K[\vec{X}=\vec{x}]\psi\Big)$

a hypothetical intervention brings to the epistemic state of the agent (left side of the implication) will also be there in case the intervention takes actually place (right side of the implication). As a consequence, whatever the agent can infer from a hypothetical intervention she will also infer from the actual intervention. Hence, **PR** is valid. However, in the case of an actual intervention the agent can infer more. She might observe consequences of the action on variables that she can observe, which will add to the knowledge that she has. This additional information an agent can gain in case of an actual intervention renders the right-to-left side of the implication (*no learning*) invalid.

PROPOSITION 5.5 (Perfect recall).
Let $\mathcal{E}=\langle\mathscr{S},\mathcal{F},\mathcal{T}\rangle$ be an epistemic causal model with observables, $\mathcal{A}\in\mathcal{T}$, and suppose $(\mathcal{E},\mathcal{A})\models K[\vec{X}{=}\vec{x}]\psi$. Then $(\mathcal{E},\mathcal{A})\models[\vec{X}{=}\vec{x}]K\psi$.

PROOF. If $(\mathcal{E},\mathcal{A})\models K[\vec{X}{=}\vec{x}]\psi$, then for all $\mathcal{B}\in\mathcal{T}$, $(\mathcal{E},\mathcal{B})\models[\vec{X}{=}\,\vec{x}]\psi$. In particular, this holds for all $\mathcal{B}\in\mathcal{T}$ such that $\mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}\in\mathcal{T}^{\mathcal{F},\mathcal{A}}_{\vec{X}=\,\vec{x}}$. So, for all such $\mathcal{B}$, $(\mathcal{E}^{\mathcal{B}}_{\vec{X}=\vec{x}},\mathcal{B}^{\mathcal{F}}_{\vec{X}=\,\vec{x}})\models\psi$. But now observe that, for all such $\mathcal{B}$, $\mathcal{T}^{\mathcal{B}}_{\vec{X}=\vec{x}}=\{\mathcal{C}^{\mathcal{F}}_{\vec{X}=\vec{x}}\mid\mathcal{C}\in\mathcal{T}$ and $\mathcal{C}^{\mathcal{F}}_{\vec{X}=\vec{x}}(O)=\mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}(O)$ for all $O\in\mathcal{O}_T\}=\{\mathcal{C}^{\mathcal{F}}_{\vec{X}=\vec{x}}\mid\mathcal{C}\in\mathcal{T}$ and $\mathcal{C}^{\mathcal{F}}_{\vec{X}=\vec{x}}(O)=\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}(O)$ for all $O\in\mathcal{O}_T\}=\mathcal{T}^{\mathcal{F},\mathcal{A}}_{\vec{X}=\vec{x}}$. So $(\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}},\mathcal{B}_{\vec{X}=\vec{x}})\models\psi$ for all $\mathcal{B}_{\vec{X}=\vec{x}}\in\mathcal{T}^{\mathcal{F},\mathcal{A}}_{\vec{X}=\vec{x}}$; thus $(\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}},\mathcal{A}_{\vec{X}=\vec{x}})\models K\psi$. So $(\mathcal{E},\mathcal{A})\models[\vec{X}{=}\vec{x}]K\psi$. $\square$

Axiom system Our considerations on the failure of the 'no learning' principle highlight the fact that the proof theory of in the semantics with observables will differ from that in the case without. In particular, we cannot rely on the axiom **CM**, which played an important role in the elimination of public announcements in L_{PAKC}.

We then propose the following axiom system L^{O}_{PAKC} (once more, parametrized by a signature $\mathscr{S}$). L^{O}_{PAKC} is defined from L_{PAKC} by removing the axiom **CM** and adding the principles in Table 4. As for L_{PAKC}, each axiom of L^{O}_{PAKC} is taken also in the special case that the set of intervened variables $\vec{X}$ is empty.

The axiom **OC** expresses the fact that the observables always take a constant value in the set of valuations, i.e. their value is always known to the experimenter—before and after the experiment. Axiom **PD** (*prediction axiom*) describes how the knowledge of the agent after an intervention is characterized by the knowledge she may have before any intervention occurs, just as axiom **CM** does in the case without observables. However, while in the case without observables the knowledge the agent needs to have prior to the intervention is full knowledge of the outcome (*no learning*), now she may also make use of her knowledge of the effect that the intervention would have on the observables. Indeed, **PD** essentially states that after an intervention the agent knows ψ if and only if, *after learning how the values of the observables are affected by the intervention*, she knows that the intervention makes ψ true. By characterizing knowledge after an intervention in terms of knowledge before the intervention, axiom **PD** plays a crucial role in the procedure of elimination of the announcement operators, by allowing us to extract occurrences of K from the scope of an intervention (see Proposition A.17 in the Appendix). A similar role was played by axiom **CM** in the completeness proof for L_{PAKC}.

THEOREM 5.6 (Axiom system for $\mathcal{L}_{PAKC}$ over models with observables).
Let $\mathscr{S} = \langle \mathcal{U}, \mathcal{V}, \mathcal{O}, \mathcal{R} \rangle$ be a finite signature with observables. The axiom system L^{O}_{PAKC} is sound and strongly complete for the language $\mathcal{L}_{PAKC}$ based on $\mathscr{S}$ with respect to epistemic (recursive) causal models with observables.

PROOF. See Appendix A.4. □

One may wonder what kind of relationship subsists between the two interpretations of $\mathcal{L}_{PAKC}$ (i.e. with and without observables). It can be proved that, as long as only finite signatures are considered—or more generally, only finite sets of observables—$\mathcal{L}_{PAKC}$, as interpreted with observables, can be embedded into its version without observables. So, in principle, reasoning about real experiments could also be conducted within the semantics without observables; but this comes at the cost of an increase in the size and complexity of formulas. The involved details of this argument will be presented in a forthcoming manuscript.

6 Conclusions

This paper has given some steps towards the integration of causal and epistemic reasoning, providing an adequate semantics, a language combining interventionist counterfactuals with (dynamic) epistemic operators and a sound and complete system of inference. Our deductive system models the thought of an agent reasoning about the consequences of hypothetical and real experiments. It describes what an agent may deduce from her/his *a priori* pool of knowledge when considering a hypothetical intervention, but also what she could learn from performing this experiment in the actual world. For this later part the concept of observables was crucial. Observables allow us to model measuring the value of variables when performing experiment. Through the addition of observables to our logic we are able to model actual learning from experiments. In this respect the system proposed here goes substantially beyond [9].

Still, the framework developed here is just a first step towards providing a logic of reasoning about experiments. There are aspects of reasoning about experiments that our proposal cannot yet account for. Some of these limitations are due to assumptions made by the models of epistemic and causal reasoning that we combined. For instance, the particular model of causal information we have worked with, structural causal models, assume that the value of a variable is fully determined by the value of its causal parents. But most causal relationships science deals with are of a probabilistic/statistical nature: if the values of the causal parents are such and such, then there is a probability p that the value of X is x. This could be modelled better using causal Bayes nets as a starting point. In such case, one might want to use a probabilistic setting for modelling *uncertainty* [see, e.g. 24] and a matching extension for dealing with change in uncertainty [e.g. the already mentioned 3, 31, 32].

There are also some assumptions about the interaction between epistemic and causal reasoning that could be reconsidered in future work. One of them, for instance, concerns the way we implemented interventions. The logic proposed here treats an intervention as public knowledge: an epistemic state is automatically updated with any intervention that takes place – all experiments are observed. Whether the agent can also see the effects of the intervention depends on which variables she can observe. But she will certainly be aware of the change in the value of variable(s) that are in the scope of the intervention. We have chosen this implementation of intervention because we wanted to push here the perspective of intervention as just another action in dynamic logic. For this reason we wanted to make it as similar as possible to the operation of public announcement. Furthermore, in the single agent setting it seems natural to assume that this agent is reasoning about experiments she is performing. However, in a multi-agent setting, this assumption can be given up.

There is one assumption we have made here that deserves particular attention. As explained in the introduction, the goal of the formal work reported in this paper is to come up with a qualitative logic of reasoning about experiments. One important function of experiments is to provide us information about causal relationships. However, this cannot be modelled by the system proposed here. This system takes the causal dependencies to be given to the agent. The only information an agent can gain from doing experiments is information about the actual value of variables. One way to solve this problem is to introduce uncertainty about the right causal network, for instance [as in 8] by letting in the epistemic state not only the assignment, but also the causal model vary. Making this change is not hard, but it raises the interesting and difficult question what kind of regularities we want to implement concerning learning causal relationships. This is something we want to study in future work, together with the question how such a framework would relate to the algorithms for learning causal networks that are used in various scientific applications, such as [38, 40].

A third assumption concerns the structure of the causal laws. As is standard in the literature, we assumed here that causal structures are not circular. Models without this restriction are mostly used to represent complex feedback systems rather than everyday scenarios such as we considered in this paper; and their general formal treatment would lead us away from the safe methods of modal logic, since on acyclic structures interventions may not guarantee the existence and uniqueness of an actual world. There is however a ‘safe’ class of causal models, not necessarily acyclic, on which interventions uniquely identify one actual world: the so-called *unique-solution* systems [17]. It might be interesting to extend our work in this direction.

As mentioned before, our framework has many points in common with causal team semantics. In [9] we provide a translation between the two approaches. Our semantics has the advantage of encoding explicitly the actual state of the world (in particular, the actual value of variables). For our purposes having access to the actual value of variables was crucial, because it enabled us to account for learning from experiments. Actual values play also an important role in many definitions of *token causation*, i.e. causation between events [23, 27, 47]. Our framework might help with formalizing these definitions and comparing them to each other. In future work we plan to consider richer languages with hybrid features that will make even more use of the available information about the actual world. For example, one could distinguish in the language the actual action (intervention) from observing the action; and operators for *fixing the actual value* of a variable can help to spell out definitions of token causation.

Another family of logics that our proposal is related to are the logics of dependence and independence. As demonstrated in [2, 45], the notion of functional dependence can be reduced in terms of knowledge operators. This raised the question whether other forms of correlational dependence may be described in terms of epistemic operators. The literature on team semantics considered a large variety of such (in)dependence notions, among which the most studied are the *independence atoms* [21] and *inclusion atoms* [19]. More generally, in the kinds of frameworks developed in the present paper, we can investigate how a notion of causal dependence can be decomposed into epistemic and purely causal aspects. We believe this might contribute to the everlasting problem of separating clearly the epistemology and the ontology of causation.

Finally, in future work we plan to extend the setting to a multi-agent system. This involves considering not only different agents with potentially different knowledge and different variables they can observe, but also epistemic attitudes for groups (e.g. distributed and common knowledge) and the effect of inter-agent communication. One advantage this will bring is the potential to contribute to the discussion about causal agency and the role of causation in the study of responsibility within AI [see, for instance, 5].

A Appendix

A.1 Proof of Theorem 2.10

Soundness. For **P** and **MP** this is straightforward; for $\mathbf{A_1}$–$\mathbf{A_5}$ and $\mathbf{A}_\neg$ this has been proved in [22]. Axioms $\mathbf{A_6}$ and $\mathbf{A}_\neg$ are sound because we work on *recursive* causal models. For the first, a recursive $\mathcal{F}$ produces a relation $\hookrightarrow_{\mathcal{F}}$ without cycles (Footnote 11), syntactically characterized by $\rightsquigarrow$. For the second, in a recursive causal model, the value of each variable is uniquely determined.

For $\mathbf{A_7}$ note how, for any assignment $\vec{X}{=}\,\vec{x}$, the valuations $\mathcal{A}$ and $\mathcal{A}^{\mathcal{F}}_{\vec{X}=\,\vec{x}}$ coincide in the value of *exogenous* variables not occurring in $\vec{X}$ (Definition 2.6). For $\mathbf{A_{[]}}$, an intervention with the empty assignment does not affect the given causal model. Finally, $\mathbf{A_{[][]}}$states that, when two interventions are performed in a row, the second overrides the first in the variables they both act upon.

Completeness. For completeness, it will be shown that axioms $\mathbf{A_{[]}}$, $\mathbf{A}_\neg$, $\mathbf{A}_\wedge$ and $\mathbf{A_{[][]}}$ define a translation from $\mathcal{L}_C$ to a language $\mathcal{L}'_C$ for which axioms $\mathbf{A_1}$–$\mathbf{A_7}$, **P** and rule **MP** are complete. Here are the steps.

Completeness, step 1: from $\mathcal{L}_C$ to $\mathcal{L}'_C$. Formulas γ of the language $\mathcal{L}'_C$ [22] consists of Boolean combinations of expressions of the form $[\vec{X}{=}\,\vec{x}]Y{=}y$. They are semantically interpreted over causal models as before.

Definition A.1 (Translation tr_1).
Define $\mathrm{tr}_1 : \mathcal{L}_C \to \mathcal{L}'_C$ as

$$
\begin{array}{rr}
\mathrm{tr}_1(Y{=}y) := [\;]Y{=}y & \mathrm{tr}_1([\vec{X}{=}\,\vec{x}]Y{=}y) := [\vec{X}{=}\,\vec{x}]Y{=}y \\
\mathrm{tr}_1(\neg\varphi) := \neg\,\mathrm{tr}_1(\varphi) & \mathrm{tr}_1([\vec{X}{=}\,\vec{x}]\neg\varphi) := \mathrm{tr}_1(\neg[\vec{X}{=}\,\vec{x}]\varphi) \\
\mathrm{tr}_1(\varphi_1 \wedge \varphi_2) := \mathrm{tr}_1(\varphi_1) \wedge \mathrm{tr}_1(\varphi_2) & \mathrm{tr}_1([\vec{X}{=}\,\vec{x}](\varphi_1 \wedge \varphi_2)) := \mathrm{tr}_1([\vec{X}{=}\,\vec{x}]\varphi_1 \wedge [\vec{X}{=}\,\vec{x}]\varphi_2) \\
 & \mathrm{tr}_1([\vec{X}{=}\,\vec{x}][\vec{Y}{=}\,\vec{y}]\varphi) := \mathrm{tr}_1([\vec{X}'{=}\,\vec{x}', \vec{Y}{=}\,\vec{y}]\varphi)
\end{array}
$$

with $\vec{X}'{=}\,\vec{x}'$ the subassignment of $\vec{X}{=}\,\vec{x}$ for $\vec{X}' := \vec{X} \setminus \vec{Y}$.

The following proposition indicates the crucial properties of tr_1.

Proposition A.2
For every $\varphi \in \mathcal{L}_C$,

(i) $\mathrm{tr}_1(\varphi) \in \mathcal{L}'_C$,
(ii) $\vdash_{\mathsf{L}_C} \varphi \leftrightarrow \mathrm{tr}_1(\varphi)$,
(iii) $\models \varphi \leftrightarrow \mathrm{tr}_1(\varphi)$.

PROOF. Sketch The proofs of this proposition use induction on a notion of *complexity* $\mathrm{c} : \mathcal{L}_C \to \mathbb{N} \setminus \{0\}$, defined as

$$\mathrm{c}(Y{=}y) := 1 \qquad \mathrm{c}(\varphi_1 \wedge \varphi_2) := 1 + \max\{\mathrm{c}(\varphi_1), \mathrm{c}(\varphi_2)\}$$

$$\mathrm{c}(\neg\varphi) := 1 + \mathrm{c}(\varphi) \qquad \mathrm{c}([\vec{X}{=}\vec{x}]\varphi) := 2\,\mathrm{c}(\varphi)$$

Note: for all assignments $\vec{X}{=}\vec{x}$ and $\vec{Y}{=}\vec{y}$, and all formulas $\varphi, \varphi_1, \varphi_2 \in \mathcal{L}_C$,

$$\begin{array}{r}
\hline
\mathrm{c}(\varphi) \geqslant \mathrm{c}(\psi) \text{ for every } \psi \in \mathrm{sub}(\varphi) \\
\hline
\mathrm{c}([\vec{X}{=}\vec{x}]\neg\varphi) > \mathrm{c}(\neg[\vec{X}{=}\vec{x}]\varphi) \\
\mathrm{c}([\vec{X}{=}\vec{x}](\varphi_1 \wedge \varphi_2)) > \mathrm{c}([\vec{X}{=}\vec{x}]\varphi_1 \wedge [\vec{X}{=}\vec{x}]\varphi_2) \\
\mathrm{c}([\vec{X}{=}\vec{x}][\vec{Y}{=}\vec{y}]\varphi) > \mathrm{c}([\vec{X}'{=}\vec{x}', \vec{Y}{=}\vec{y}]\varphi) \\
\hline
\end{array}$$

Thus, in every case in Definition A.2, the complexity of the formulas under tr_1 on the right-hand side is strictly smaller than the complexity of the formula under tr_1 on the left-hand side. Hence, tr_1 is a proper recursive translation. With this, the first and the second item can be proved by induction on $\mathrm{c}(\varphi)$ (the first using tr_1's definition and the inductive hypothesis (IH); the second using, additionally, propositional reasoning and axioms $\mathbf{A}_{[]}$, $\mathbf{A}_{\neg}$, $\mathbf{A}_{\wedge}$ and $\mathbf{A}_{[][]}$). The third item follows from the second and the system's soundness.

Completeness, step 2: a canonical model for $\mathcal{L}'_C$. The next step is to show that , the fragment of without axioms $\mathbf{A}_{[]}$, $\mathbf{A}_{\neg}$, $\mathbf{A}_{\wedge}$ and $\mathbf{A}_{[][]}$, is complete for $\mathcal{L}'_C$ over recursive causal models. This is done by showing, via the construction of a canonical model, that any L'_C-consistent set of $\mathcal{L}'_C$-formulas is satisfiable in a recursive causal model. The construction is almost exactly as that in [22], so here we will only describe the main tasks. □

Fix the signature $\mathscr{S} = \langle \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{R}} \rangle$. Let $\mathbb{C}_C$ be the set of all maximally L'_C-consistent sets of $\mathcal{L}'_C$-formulas. The first task is to show that each $\Gamma \in \mathbb{C}_C$ gives raise to a recursive causal model. A valuation $\mathcal{A}^\Gamma$ is defined by picking as $\mathcal{A}(Z)$ (for $Z \in \boldsymbol{\mathcal{W}}$) the unique z such that $[\ ]Z{=}z \in \Gamma$. Similarly, the structural function for an endogenous variable V is read off from the formulas of the form $[\vec{U}{=}\vec{u}, \vec{Y}{=}\vec{y}]V{=}v$ (with $\vec{U} = \boldsymbol{\mathcal{U}}$ and $\vec{Y} = \boldsymbol{\mathcal{V}} \setminus \{V\}$) that occur in Γ. It can be shown that the resulting structure, $\langle \mathscr{S}, \mathcal{F}^\Gamma, \mathcal{A}^\Gamma \rangle$ is a proper recursive causal model (i.e. $\mathcal{F}^\Gamma$ is recursive and $\mathcal{A}^\Gamma$ complies with $\mathcal{F}^\Gamma$).

The second task is to prove a truth lemma: every $\gamma \in \mathcal{L}'{}_C$ is such that $\langle \mathscr{S}, \mathcal{F}^\Gamma, \mathcal{A}^\Gamma \rangle \models \gamma$ if and only if $\gamma \in \Gamma$. The crucial base case is as in [22]; the Boolean cases rely on the IH. With it, one can prove the following.

THEOREM A.3 (Completeness of L'_C for $\mathcal{L}'_C$).
The system L'_C is strongly complete for $\mathcal{L}'_C$ based on $\mathscr{S}$ w.r.t. recursive causal models for $\mathscr{S}$. In other words, for every set of $\mathcal{L}'_C$-formulas $\Gamma^- \cup \{\gamma\}$, if $\Gamma^- \models \gamma$ then $\Gamma^- \vdash_{\mathsf{L}'_C} \gamma$.

Finally, the argument for strong completeness of L_C for formulas in $\mathcal{L}_C$ is as follows. First, take $\Psi \cup \{\varphi\} \subseteq \mathcal{L}_C$ and suppose $\Psi \models \varphi$. Since $\models \varphi' \leftrightarrow \mathrm{tr}_1(\varphi')$ for every $\varphi' \in \mathcal{L}_C$ (Proposition A.2.3), by defining $\mathrm{tr}_1(\Psi) := \{\mathrm{tr}_1(\psi) \mid \psi \in \Psi\}$ it follows that $\mathrm{tr}_1(\Psi) \models \mathrm{tr}_1(\varphi)$. Now, $\mathrm{tr}_1(\varphi') \in \mathcal{L}'_C$ for every $\varphi' \in \mathcal{L}_C$ (Proposition A.2.1); hence, $\mathrm{tr}_1(\Psi) \cup \{\mathrm{tr}_1(\varphi)\} \subseteq \mathcal{L}'_C$. Therefore, the just obtained $\mathrm{tr}_1(\Psi) \models \mathrm{tr}_1(\varphi)$ and Theorem A.3 imply $\mathrm{tr}_1(\Psi) \vdash_{\mathsf{L}'_C} \mathrm{tr}_1(\varphi)$. Since L'_C is a subsystem of L_C, it follows that $\mathrm{tr}_1(\Psi) \vdash_{\mathsf{L}_C} \mathrm{tr}_1(\varphi)$. But $\vdash_{\mathsf{L}_C} \varphi' \leftrightarrow \mathrm{tr}_1(\varphi')$ for every $\varphi' \in \mathcal{L}_C$ (Proposition A.2.2); hence, $\Psi \vdash_{\mathsf{L}_C} \varphi$, as required.

A.2 Proof of Theorem 3.5

Soundness. Axioms and rules in Table 1 do not involve the K operator, so their truth-value in a pointed epistemic causal model $(\langle \mathscr{S}, \mathcal{F}, \mathcal{T}\rangle, \mathcal{A})$ only depends on the causal model $\langle \mathscr{S}, \mathcal{F}, \mathcal{A}\rangle$, for which they are sound (Theorem 2.10).

For Table 2, axioms and rules in the *epistemic* part are sound on relational structures with a single equivalence relation [12, 16], which is equivalent to having a simple set of epistemic alternatives, as epistemic causal models have. For **CM** recall: (i) $K[\vec{X}=\vec{x}]\xi$ holds at $(\langle \mathscr{S}, \mathcal{F}, \mathcal{T}\rangle, \mathcal{A})$ iff ξ holds at $(\langle \mathscr{S}, \mathcal{F}_{\vec{X}=\vec{x}}, \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}}\rangle, \mathcal{A}'^{\mathcal{F}}_{\vec{X}=\vec{x}})$ for every $\mathcal{A}' \in \mathcal{T}$, and (ii) $[\vec{X}=\vec{x}]K\xi$ holds at $(\langle \mathscr{S}, \mathcal{F}, \mathcal{T}\rangle, \mathcal{A})$ iff ξ holds at $(\langle \mathscr{S}, \mathcal{F}_{\vec{X}=\vec{x}}, \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}}\rangle, (\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}})')$ for every $(\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}})' \in \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}}$. Then note how, by Definition 3.2, the set of relevant valuations for the first, $\{\mathcal{A}'^{\mathcal{F}}_{\vec{X}=\vec{x}} \mid \mathcal{A}' \in \mathcal{T}\}$, is exactly as that for the second, $\mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}}$. For recall: all valuations in $\mathcal{T}$ comply with the same structural functions.

Completeness. Again, by translation, now also using the following rule.

LEMMA A.4
Let ξ_1, ξ_2 be $\mathcal{L}_{KC}$formulas.

$$\mathbf{RE_K}\text{ :if } \vdash_{\mathsf{L}_{KC}} \xi_1 \leftrightarrow \xi_2 \text{ then } \vdash_{\mathsf{L}_{KC}} K\xi_1 \leftrightarrow K\xi_2.$$

Completeness, step 1: from $\mathcal{L}_{KC}$ to $\mathcal{L}'_{KC}$. Formulas δ in the language $\mathcal{L}'_{KC}$ are built by using Boolean and K operators on 'atoms' of the form $[\vec{X}=\vec{x}]Y{=}y$. They are semantically interpreted just as formulas in $\mathcal{L}_{KC}$.

Here is the translation.

DEFINITION A.5 (Translation tr_2).
Define $\mathrm{tr}_2 : \mathcal{L}_{KC} \to \mathcal{L}'_{KC}$ as

$$\begin{array}{rr}
\mathrm{tr}_2(Y{=}y) := [\,]Y{=}y & \mathrm{tr}_2([\vec{X}=\vec{x}]Y{=}y) := [\vec{X}=\vec{x}]Y{=}y \\
\mathrm{tr}_2(\neg\xi) := \neg\,\mathrm{tr}_2(\xi) & \mathrm{tr}_2([\vec{X}=\vec{x}]\neg\xi) := \mathrm{tr}_2(\neg[\vec{X}=\vec{x}]\xi) \\
\mathrm{tr}_2(\xi_1 \wedge \xi_2) := \mathrm{tr}_2(\xi_1) \wedge \mathrm{tr}_2(\xi_2) & \mathrm{tr}_2([\vec{X}=\vec{x}](\xi_1 \wedge \xi_2)) := \mathrm{tr}_2([\vec{X}=\vec{x}]\xi_1 \wedge [\vec{X}=\vec{x}]\xi_2) \\
\mathrm{tr}_2(K\xi) := K\,\mathrm{tr}_2(\xi) & \mathrm{tr}_2([\vec{X}=\vec{x}]K\xi) := \mathrm{tr}_2(K[\vec{X}=\vec{x}]\xi) \\
 & \mathrm{tr}_2([\vec{X}=\vec{x}][\vec{Y}=\vec{y}]\xi) := \mathrm{tr}_2([\vec{X}'=\vec{x}', \vec{Y}=\vec{y}]\xi)
\end{array}$$

with $\vec{X}'=\vec{x}'$ the subassignment of $\vec{X}=\vec{x}$ for $\vec{X}' := \vec{X} \setminus \vec{Y}$.

The following proposition contains tr_2's crucial properties.

PROPOSITION A.6
For every $\xi \in \mathcal{L}_{KC}$,

(i) $\mathrm{tr}_2(\xi) \in \mathcal{L}'_{KC}$,
(ii) $\vdash_{\mathsf{L}_{KC}} \xi \leftrightarrow \mathrm{tr}_2(\xi)$,
(iii) $\models \xi \leftrightarrow \mathrm{tr}_2(\xi)$.

PROOF. The proofs rely again on the formulas' complexity $\mathrm{c} : \mathcal{L}_{KC} \to \mathbb{N} \setminus \{0\}$, this time defined as

$$\mathrm{c}(Y{=}y) := 1 \qquad \mathrm{c}(K\xi) := 1 + c(\xi)$$
$$\mathrm{c}(\neg\xi) := 1 + \mathrm{c}(\xi) \qquad \mathrm{c}([\vec{X}{=}\vec{x}]\xi) := 2\,\mathrm{c}(\xi)$$
$$\mathrm{c}(\xi_1 \wedge \xi_2) := 1 + \max\{\mathrm{c}(\xi_1), \mathrm{c}(\xi_2)\}$$

Note: for every assignments $\vec{X}{=}\vec{x}$ and $\vec{Y}{=}\vec{y}$, and every $\xi, \xi_1, \xi_2 \in \mathcal{L}_{KC}$,

$$\begin{array}{rcl}
\hline
 & & \mathrm{c}(\xi) \geqslant \mathrm{c}(\psi) \text{ for every } \psi \in \mathrm{sub}(\xi) \\
\hline
\mathrm{c}([\vec{X}{=}\vec{x}]\neg\xi) & > & \mathrm{c}(\neg[\vec{X}{=}\vec{x}]\xi) \\
\mathrm{c}([\vec{X}{=}\vec{x}](\xi_1 \wedge \xi_2)) & > & \mathrm{c}([\vec{X}{=}\vec{x}]\xi_1 \wedge [\vec{X}{=}\vec{x}]\xi_2) \\
\mathrm{c}([\vec{X}{=}\vec{x}]K\xi) & > & \mathrm{c}(K[\vec{X}{=}\vec{x}]\xi) \\
\mathrm{c}([\vec{X}{=}\vec{x}][\vec{Y}{=}\vec{y}]\xi) & > & \mathrm{c}([\vec{X}'{=}\vec{x}', \vec{Y}{=}\vec{y}]\xi) \\
\hline
\end{array}$$

Thus, tr_2 is a proper recursive translation. The proof of first two properties use induction on $\mathrm{c}(\xi)$. The base case is for formulas ξ with $\mathrm{c}(\xi) = 1$ $(Y{=}y)$; the inductive case is for formulas ξ with $\mathrm{c}(\xi) > 1$ $(\neg\xi, \xi_1 \wedge \xi_2, K\xi,\ [\vec{X}{=}\vec{x}]Y{=}y, [\vec{X}{=}\vec{x}]\neg\xi, [\vec{X}{=}\vec{x}](\xi_1 \wedge \xi_2),\ [\vec{X}{=}\vec{x}]K\xi$ and $[\vec{X}{=}\vec{x}][\vec{Y}{=}\vec{y}]\xi)$.

(i) The IH states that $\mathrm{tr}_2(\psi) \in \mathcal{L}'_{KC}$ for all formulas $\psi \in \mathcal{L}_{KC}$ with $\mathrm{c}(\psi) < \mathrm{c}(\xi)$. All cases are straightforward from tr_2's definition and the respective IH.

(ii) The IH states that $\vdash_{\mathsf{L}_{KC}} \psi \leftrightarrow \mathrm{tr}_2(\psi)$ holds for all formulas $\psi \in \mathcal{L}_{KC}$ with $\mathrm{c}(\psi) < \mathrm{c}(\xi)$. All cases follow from tr_2's definition, the IH, propositional reasoning and the axiom system ($\mathbf{A_{[]}}$ for formulas of the form $Y{=}y$, $\mathbf{RE_K}$ on Lemma A.4 for formulas of the form $K\xi$, $\mathbf{A_\neg}$ for formulas of the form $[\vec{X}{=}\vec{x}]\neg\xi$, $\mathbf{A_\wedge}$ for formulas of the form $[\vec{X}{=}\vec{x}](\xi_1 \wedge \xi_2)$, **CM** for formulas of the form $[\vec{X}{=}\vec{x}]K\xi$, $\mathbf{A_{[][]}}$ for formulas of the form $[\vec{X}{=}\vec{x}][\vec{Y}{=}\vec{y}]\xi$).

(iii) By the previous item, $\vdash_{\mathsf{L}_{KC}} \xi \leftrightarrow \mathrm{tr}_2(\xi)$. But is sound within recursive causal models; therefore, $\models \xi \leftrightarrow \mathrm{tr}_2(\xi)$.

Completeness, step 2: a canonical model for $\mathcal{L}'_{KC}$. Next it is shown that L'_{KC}, the fragment of L_{KC} without axioms $\mathbf{A_{[]}}$, $\mathbf{A_\neg}$, $\mathbf{A_\wedge}$ and $\mathbf{A_{[][]}}$ is complete for $\mathcal{L}'_{KC}$ over epistemic (recursive) causal models. This is done by showing, via the construction of a canonical model, that any L'_{KC}-consistent set of $\mathcal{L}'_{KC}$-formulas is satisfiable in an epistemic (recursive) causal model. The construction uses the ideas of [22] for building a recursive causal model, and then ideas from [16] to extend it to a recursive epistemic causal model, taking additional care of guaranteeing that all valuations comply with the same set of structural functions.

Let $\mathbb{C}_{KC}$ be the set of all maximally L'_{KC}-consistent sets of $\mathcal{L}'_{KC}$-formulas; take $\Delta \in \mathbb{C}_{KC}$. Since $\mathcal{L}'_{KC}$ extends $\mathcal{L}'_C$ and extends L'_C, the set $\Delta|_{\mathcal{L}'_C}$ (the restriction of Δ to formulas in $\mathcal{L}'_C$) is a maximally L'_C-consistent set of $\mathcal{L}'_C$-formulas. Then (completeness proof of Theorem 2.10), each such Δ defines a set of structural functions for $\boldsymbol{\mathcal{V}}$, namely $\mathcal{F}^{\Delta|_{\mathcal{L}'_C}}$. This is the basis for defining an epistemic causal model $\mathcal{E}^\Delta = \langle \mathscr{S}, \mathcal{F}^{\Delta|_{\mathcal{L}'_C}}, \mathcal{T}^\Delta \rangle$, where the valuations in $\mathcal{T}^\Delta$ are built as follows. First, define the set $\mathbb{C}^\Delta_{KC} \subseteq \mathbb{C}_{KC}$ containing the elements of $\mathbb{C}_{KC}$ whose structural functions coincide with those of

Δ (obviously, $\Delta \in \mathbb{C}^{\Delta}_{KC}$). Then define $\mathcal{T}^{\Delta}$ as those sets in $\mathbb{C}^{\Delta}_{KC}$ that fit the epistemic information in Δ.[25]

It can be shown, by structural induction on $\delta \in \mathcal{L}'_{KC}$, that $\mathcal{E}^{\Delta}$ satisfies a truth lemma. The base case (formulas of the form $[\vec{X}=\vec{x}]Z=z$) is the base case of the truth lemma for Theorem 2.10; the Boolean cases use their IH. For $K\xi$, the proof goes as in standard epistemic logic, using the fact that, by **KL**, the formulas in Δ defining $\mathcal{F}^{\Delta|_{\mathcal{L}'_C}}$ are also in every other element of $\mathbb{C}^{\Delta}_{KC}$. Moreover, $\mathcal{E}^{\Delta}$ is an epistemic recursive causal model (every valuation in $\mathcal{T}^{\Delta}$ complies with $\mathcal{F}^{\Delta|_{\mathcal{L}'_C}}$, which is recursive). Thus, we get the following. □

THEOREM A.7 (Completeness of L'_{KC} for $\mathcal{L}'_{KC}$).
Given a signature $\mathscr{S}$, the system L'_{KC} is strongly complete for $\mathcal{L}'_{KC}$ w.r.t. epistemic recursive causal models.

Finally, the argument for strong completeness of L_{KC} for formulas in $\mathcal{L}_{KC}$ is analogous to that in the proof of Theorem 2.10 (Page 52).

A.3 Proof of Theorem 3.8

Soundness. For axioms and rules in Tables 1 and 2 see Theorem 3.5. Soundness of $\mathbf{N}_=$ and $\mathbf{N}_!$ comes from the fact that both an intervention and an announcement on an epistemic (recursive) causal model returns an epistemic (recursive) causal model (Definitions 3.2 and 3.6, respectively). For $\mathbf{{}_!RE}$ note that, if $\mathcal{E}$ is an epistemic causal model and $\models \alpha_1 \leftrightarrow \alpha_2$, then $\mathcal{E}^{\alpha_1}$ and $\mathcal{E}^{\alpha_2}$ are identical. Then, while $\mathbf{!}_=$, $\mathbf{!}_\neg$, $\mathbf{!}_\wedge$, $\mathbf{!_K}$ and $\mathbf{K_!}$ are standard axioms for a modality $[\alpha!]$ describing the effects of a model operation defined as a partial and deterministic function [see, e.g. 46], $\mathbf{!_!}$is well-known from public announcement logic [44, Section 7.4].

Finally, for , take any epistemic causal model $\langle \mathcal{F}, \mathcal{T} \rangle$ (for simplicity, omit the signature) and any $\mathcal{A} \in \mathcal{T}$. Expanding the axiom's left-hand side yields

$$\begin{aligned}
(\langle \mathcal{F}, \mathcal{T} \rangle, \mathcal{A}) \models [\vec{X}=\vec{x}][\alpha!]\chi \ \text{ iff } & \ (\langle \mathcal{F}_{\vec{X}=\vec{x}}, \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}} \rangle, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models [\alpha!]\chi \\
\text{iff } & \ (\langle \mathcal{F}_{\vec{X}=\vec{x}}, \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}} \rangle, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \alpha \\
& \quad \text{implies } (\langle \mathcal{F}_{\vec{X}=\vec{x}}, (\mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}})^{\alpha} \rangle, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \chi
\end{aligned}$$

and expanding its right-hand yields

$$\begin{aligned}
(\langle \mathcal{F}, \mathcal{T} \rangle, \mathcal{A}) \models [[\vec{X}=\vec{x}]\alpha!][\vec{X}=\vec{x}]\chi \ \text{ iff } & \ (\langle \mathcal{F}, \mathcal{T} \rangle, \mathcal{A}) \models [\vec{X}=\vec{x}]\alpha \\
& \quad \text{implies } (\langle \mathcal{F}, \mathcal{T}^{[\vec{X}=\vec{x}]\alpha} \rangle, \mathcal{A}) \models [\vec{X}=\vec{x}]\chi \\
\text{iff } & \ (\langle \mathcal{F}_{\vec{X}=\vec{x}}, \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}} \rangle, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \alpha \\
& \quad \text{implies } (\langle \mathcal{F}_{\vec{X}=\vec{x}}, (\mathcal{T}^{[\vec{X}=\vec{x}]\alpha})^{\mathcal{F}}_{\vec{X}=\vec{x}} \rangle, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \chi
\end{aligned}$$

[25] This is done in two stages. First, define an epistemic indistinguishability relation $R^{\Delta} \subseteq \mathbb{C}^{\Delta}_{KC} \times \mathbb{C}^{\Delta}_{KC}$ as typically done in modal canonical models [see, e.g. 12, 16]. Due to axioms **T**, **4** and **5**, the relation R^{Δ} is an equivalence relation (in particular, $(\Delta, \Delta) \in R^{\Delta}$). Then, $\mathcal{T}^{\Delta}$ contains the valuation function induced by each $\Psi \in \mathbb{C}^{\Delta}_{C}$ satisfying $R^{\Delta}\Delta\Psi$ (since $(\Delta, \Delta) \in R^{\Delta}$, we have $\mathcal{A}^{\Delta|_{\mathcal{L}'_C}} \in \mathcal{T}^{\Delta}$).

The resulting implications have the same antecedent. The consequents are also the same, since their set of valuation functions coincide:

$$\begin{aligned}\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}} \in (\mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}})^{\alpha} \quad &\text{iff} \quad \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}} \in \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}} \text{ and } (\langle \mathcal{F}_{\vec{X}=\vec{x}}, \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}}\rangle, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \alpha \\ &\text{iff} \quad \mathcal{A} \in \mathcal{T} \text{ and } (\langle \mathcal{F}_{\vec{X}=\vec{x}}, \mathcal{T}^{\mathcal{F}}_{\vec{X}=\vec{x}}\rangle, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \alpha \\ &\text{iff} \quad \mathcal{A} \in \mathcal{T} \text{ and } (\langle \mathcal{F}, \mathcal{T}\rangle, \mathcal{A}) \models [\vec{X}=\vec{x}]\alpha \\ &\text{iff} \quad \mathcal{A} \in \mathcal{T}^{[\vec{X}=\vec{x}]\alpha} \\ &\text{iff} \quad \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}} \in (\mathcal{T}^{[\vec{X}=\vec{x}]\alpha})^{\mathcal{F}}_{\vec{X}=\vec{x}}.\end{aligned}$$

This completes this part of the proof.

Completeness. Again, by translation, now also using the following rules.

LEMMA A.8

(i) **RE$_!$**: if $\vdash_{\mathsf{L}_{PAKC}} \chi_1 \leftrightarrow \chi_2$ then $\vdash_{\mathsf{L}_{PAKC}} [\alpha!\,]\chi_1 \leftrightarrow [\alpha!\,]\chi_2$,
(ii) **RE$_=$**: If $\vdash_{\mathsf{L}_{PAKC}} \chi_1 \leftrightarrow \chi_2$ then $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}]\chi_1 \leftrightarrow [\vec{X}=\vec{x}]\chi_2$.
(iii) **=!=**: $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha!\,][\vec{Y}=\vec{y}]\chi \leftrightarrow [[\vec{X}=\vec{x}]\alpha!\,][\vec{X}'=\vec{x}', \vec{Y}=\vec{y}]\chi$, with the expression $\vec{X}'=\vec{x}'$ the subassignment of $\vec{X}=\vec{x}$ for $\vec{X}' := \vec{X} \setminus \vec{Y}$.

PROOF. Proofs for the first two rules are standard. For the third, while

$$\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha!\,][\vec{Y}=\vec{y}]\chi \leftrightarrow [[\vec{X}=\vec{x}]\alpha!\,][\vec{X}=\vec{x}][\vec{Y}=\vec{y}]\chi$$

is an instance of $=_!$, from $\mathbf{A}_{[][]}$ we get

$$\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\vec{Y}=\vec{y}]\chi \leftrightarrow [\vec{X}'=\vec{x}', \vec{Y}=\vec{y}]\chi$$

with $\vec{X}'=\vec{x}'$ the subassignment of $\vec{X}=\vec{x}$ for $\vec{X}' := \vec{X} \setminus \vec{Y}$. Hence, using the second equivalence and **RE$_!$** on the right-hand side of the first,

$$\vdash_{\mathsf{L}_{PAKC}} [[\vec{X}=\vec{x}]\alpha!\,][\vec{X}=\vec{x}][\vec{Y}=\vec{y}]\chi \leftrightarrow [[\vec{X}=\vec{x}]\alpha!\,][\vec{X}'=\vec{x}', \vec{Y}=\vec{y}]\chi.$$

Thus, by propositional reasoning from the first and the third,

$$\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha!\,][\vec{Y}=\vec{y}]\chi \leftrightarrow [[\vec{X}=\vec{x}]\alpha!\,][\vec{X}'=\vec{x}', \vec{Y}=\vec{y}]\chi.$$

In this case, the translation uses two steps: (i) from $\mathcal{L}_{PAKC}$ to an intermediate language $\mathcal{L}'_{PAKC}$ whose formulas θ are built by using of Boolean operators, K and $[\alpha!\,]$ over 'atoms' of the form $[\vec{X}=\vec{x}]Y=y$ (evaluated over epistemic causal models in the natural way), and (ii) from to the final target $\mathcal{L}'_{KC}$.

Completeness, step 1: from $\mathcal{L}_{PAKC}$ to $\mathcal{L}'_{PAKC}$. The translation from $\mathcal{L}_{PAKC}$ to $\mathcal{L}'_{PAKC}$ uses axioms $\mathbf{A}_{[]}$, $\mathbf{A}_{\neg}$, $\mathbf{A}_{\wedge}$ and **CM** for pushing intervention operators through $\neg$, $\wedge$ and K, and axioms $\mathbf{!}_{=}$, $\mathbf{!}_{\neg}$, $\mathbf{!}_{\wedge}$, $\mathbf{!}_{\mathbf{K}}$, $\mathbf{!}_{!}$ and **=!=** for eliminating public announcement operators *inside the scope of interventions*. □

DEFINITION A.9 (Translation tr_3).
Define $\mathrm{tr}_3 : \mathcal{L}_{PAKC} \to \mathcal{L}'_{PAKC}$ as

$$\begin{aligned}
\mathrm{tr}_3(Y{=}y) &:= [\,]Y{=}y \\
\mathrm{tr}_3(\neg\chi) &:= \neg\,\mathrm{tr}_3(\chi) \\
\mathrm{tr}_3(\chi_1 \wedge \chi_2) &:= \mathrm{tr}_3(\chi_1) \wedge \mathrm{tr}_3(\chi_2) \\
\mathrm{tr}_3(K\chi) &:= K\,\mathrm{tr}_3(\chi) \\
\mathrm{tr}_3([\alpha!\,]\chi) &:= [\mathrm{tr}_3(\alpha)!\,]\,\mathrm{tr}_3(\chi)
\end{aligned}$$

$$\begin{aligned}
\mathrm{tr}_3([\vec{X}{=}\vec{x}]Y{=}y) &:= [\vec{X}{=}\vec{x}]Y{=}y \\
\mathrm{tr}_3([\vec{X}{=}\vec{x}]\neg\chi) &:= \mathrm{tr}_3(\neg[\vec{X}{=}\vec{x}]\chi) \\
\mathrm{tr}_3([\vec{X}{=}\vec{x}](\chi_1 \wedge \chi_2)) &:= \mathrm{tr}_3([\vec{X}{=}\vec{x}]\chi_1 \wedge [\vec{X}{=}\vec{x}]\chi_2) \\
\mathrm{tr}_3([\vec{X}{=}\vec{x}]K\chi) &:= \mathrm{tr}_3(K[\vec{X}{=}\vec{x}]\chi) \\
\mathrm{tr}_3([\vec{X}{=}\vec{x}][\vec{Y}{=}\vec{y}]\chi) &:= \mathrm{tr}_3([\vec{X}'{=}\vec{x}', \vec{Y}{=}\vec{y}]\chi)
\end{aligned}$$

$$\begin{aligned}
\mathrm{tr}_3([\vec{X}{=}\vec{x}][\alpha!\,]Y{=}y) &:= \mathrm{tr}_3([\vec{X}{=}\vec{x}](\alpha \to Y{=}y)) \\
\mathrm{tr}_3([\vec{X}{=}\vec{x}][\alpha!\,]\neg\chi) &:= \mathrm{tr}_3([\vec{X}{=}\vec{x}](\alpha \to \neg[\alpha!\,]\chi)) \\
\mathrm{tr}_3([\vec{X}{=}\vec{x}][\alpha!\,](\chi_1 \wedge \chi_2)) &:= \mathrm{tr}_3([\vec{X}{=}\vec{x}]([\alpha!\,]\chi_1 \wedge [\alpha!\,]\chi_2)) \\
\mathrm{tr}_3([\vec{X}{=}\vec{x}][\alpha!\,]K\chi) &:= \mathrm{tr}_3([\vec{X}{=}\vec{x}](\alpha \to K(\alpha \to [\alpha!\,]\chi))) \\
\mathrm{tr}_3([\vec{X}{=}\vec{x}][\alpha_1!\,][\alpha_2!\,]\chi) &:= \mathrm{tr}_3([\vec{X}{=}\vec{x}][\alpha_1 \wedge [\alpha_1!\,]\alpha_2!\,]\chi) \\
\mathrm{tr}_3([\vec{X}{=}\vec{x}][\alpha!\,][\vec{Y}{=}\vec{y}]\chi) &:= \mathrm{tr}_3([[\vec{X}{=}\vec{x}]\alpha!\,][\vec{X}'{=}\vec{x}', \vec{Y}{=}\vec{y}]\chi)
\end{aligned}$$

with $\vec{X}'{=}\vec{x}'$ the subassignment of $\vec{X}{=}\vec{x}$ for $\vec{X}' := \vec{X} \setminus \vec{Y}$.

PROPOSITION A.10
For every $\chi \in \mathcal{L}_{PAKC}$,

(i) $\mathrm{tr}_3(\chi) \in \mathcal{L}'_{PAKC}$,
(ii) $\vdash_{\mathsf{L}_{PAKC}} \chi \leftrightarrow \mathrm{tr}_3(\chi)$,
(iii) $\models \chi \leftrightarrow \mathrm{tr}_3(\chi)$.

PROOF. By the formulas' complexity $\mathrm{c} : \mathcal{L}_{PAKC} \to \mathbb{N} \setminus \{0\}$, this time defined as

$$\begin{aligned}
\mathrm{c}(Y{=}y) &:= 1 & \mathrm{c}(K\chi) &:= 1 + c(\chi) \\
\mathrm{c}(\neg\chi) &:= 1 + \mathrm{c}(\chi) & c([\alpha!\,]\chi) &:= (7 + \mathrm{c}(\alpha))\,\mathrm{c}(\chi) \\
\mathrm{c}(\chi_1 \wedge \chi_2) &:= 1 + \max\{\mathrm{c}(\chi_1), \mathrm{c}(\chi_2)\} & \mathrm{c}([\vec{X}{=}\vec{x}]\chi) &:= 2\,\mathrm{c}(\chi)
\end{aligned}$$

Thus[26] , for all assignments $\vec{X}=\vec{x}$, $\vec{Y}=\vec{y}$ and all formulas $\chi, \chi_1, \chi_2 \in \mathcal{L}_{PAKC}$,

$$
\begin{array}{rcl}
\hline
\multicolumn{3}{c}{\mathrm{c}(\chi) \geqslant \mathrm{c}(\psi) \text{ for every } \psi \in \mathrm{sub}(\chi),} \\
\hline
\mathrm{c}([\vec{X}=\vec{x}]\neg\chi) & > & \mathrm{c}(\neg[\vec{X}=\vec{x}]\chi), \\
\mathrm{c}([\vec{X}=\vec{x}](\chi_1 \wedge \chi_2)) & > & \mathrm{c}([\vec{X}=\vec{x}]\chi_1 \wedge [\vec{X}=\vec{x}]\chi_2) \\
\mathrm{c}([\vec{X}=\vec{x}]K\chi) & > & \mathrm{c}(K[\vec{X}=\vec{x}]\chi) \\
\hline
\mathrm{c}([\vec{X}=\vec{x}][\alpha!\,]Y{=}y) & > & \mathrm{c}([\vec{X}=\vec{x}](\alpha \to Y{=}y)) \\
\mathrm{c}([\vec{X}=\vec{x}][\alpha!\,]\neg\chi) & > & \mathrm{c}([\vec{X}=\vec{x}](\alpha \to \neg[\alpha!\,]\chi)) \\
\mathrm{c}([\vec{X}=\vec{x}][\alpha!\,](\chi_1 \wedge \chi_2)) & > & \mathrm{tr}_3([\vec{X}=\vec{x}]([\alpha!\,]\chi_1 \wedge [\alpha!\,]\chi_2)) \\
\mathrm{c}([\vec{X}=\vec{x}][\alpha!\,]K\chi) & > & \mathrm{c}([\vec{X}=\vec{x}](\alpha \to K(\alpha \to [\alpha!\,]\chi))) \\
\mathrm{c}([\vec{X}=\vec{x}][\alpha_1!\,][\alpha_2!\,]\chi) & > & \mathrm{c}([\vec{X}=\vec{x}][\alpha_1 \wedge [\alpha_1!\,]\alpha_2!\,]\chi) \\
\mathrm{c}([\vec{X}=\vec{x}][\alpha!\,][\vec{Y}=\vec{y}]\chi) & > & \mathrm{c}([[\vec{X}=\vec{x}]\alpha!\,][\vec{X'}=\vec{x'}, \vec{Y}=\vec{y}]\chi) \text{ with } \vec{X'}=\vec{x'} \text{ the} \\
 & & \text{subassignment of } \vec{X}=\vec{x} \text{ for } \vec{X'} := \vec{X} \setminus \vec{Y} \\
\mathrm{c}([\vec{X}=\vec{x}][\vec{Y}=\vec{y}]\chi) & > & \mathrm{c}([\vec{X'}=\vec{x'}, \vec{Y}=\vec{y}]\chi) \\
\hline
\end{array}
$$

The first is proved by structural induction (with $\mathrm{sub}([\alpha!\,]\chi) := \{[\alpha!\,]\chi\} \cup \mathrm{sub}(\alpha) \cup \mathrm{sub}(\chi)$). Items in the second block are proved as their respective counterparts in Proposition A.6. For items in the third block, all but the last two are consequences of their 'interventionless' counterparts, which are analogous to the cases in public announcement logic [44, Lemma 7.22]. For the next to last, note that $\mathrm{c}([\vec{X}=\vec{x}][\alpha!\,][\vec{Y}=\vec{y}]\chi) = (28 + 4\,\mathrm{c}(\alpha))\,\mathrm{c}(\chi)$, yet $\mathrm{c}([[\vec{X}=\vec{x}]\alpha!\,][\vec{X'}=\vec{x'}, \vec{Y}=\vec{y}]\chi) = (14 + 4\,\mathrm{c}(\alpha))\,\mathrm{c}(\chi)$. The last is straightforward.

For the properties, the first two are proved by induction on $\mathrm{c}(\chi)$: base cases are for formulas χ with $\mathbf{c}(\boldsymbol{\chi}) = \mathbf{1}$ (i.e. $Y{=}y$), and inductive cases are for formulas χ with $\mathbf{c}(\boldsymbol{\chi}) > \mathbf{1}$ (i.e. $\neg\chi$, $\chi_1 \wedge \chi_2$, $K\chi$, $[\alpha!\,]\chi$, $[\vec{X}=\vec{x}]Y{=}y$, $[\vec{X}=\vec{x}]\neg\chi$, $[\vec{X}=\vec{x}](\chi_1 \wedge \chi_2)$, $[\vec{X}=\vec{x}]K\chi$, $[\vec{X}=\vec{x}][\alpha!\,]Y{=}y$, $[\vec{X}=\vec{x}][\alpha!\,]\neg\chi$, $[\vec{X}=\vec{x}][\alpha!\,](\chi_1 \wedge \chi_2)$, $[\vec{X}=\vec{x}][\alpha!\,]K\chi$, $[\vec{X}=\vec{x}][\alpha_1!\,][\alpha_2!\,]\chi$, $[\vec{X}=\vec{x}][\alpha!\,][\vec{Y}=\vec{y}]\chi$ and $[\vec{X}=\vec{x}][\vec{Y}=\vec{y}]\chi$).

(i) The proof is analogous to that in Proposition A.6. The base case is straightforward, and the inductive cases use the above-listed properties of c to generate appropriate IHs, then using the definition of tr_3.

(ii) The proof is analogous to that in Proposition A.6. The base case follows from axiom $\mathbf{A}_{[]}$. The inductive cases rely on the properties of c to generate appropriate IHs, and then use the definition of tr_3 with propositional reasoning (cases $\neg\chi$, $\chi_1 \wedge \chi_2$ and $[\vec{X}=\vec{x}]Y{=}y$) plus $\mathbf{RE_K}$ (case $K\chi$), $\mathbf{{}_!RE}$ and $\mathbf{RE_!}$ from Lemma A.8 (case $[\alpha!\,]\chi$),[27] axiom $\mathbf{A}_\neg$ (case $[\vec{X}=\vec{x}]\neg\chi$), axiom $\mathbf{A}_\wedge$ (case $[\vec{X}=\vec{x}](\chi_1 \wedge \chi_2)$ and axiom **CM** (case $[\vec{X}=\vec{x}]K\chi$). Cases $[\vec{X}=\vec{x}][\alpha!\,]Y{=}y$, $[\vec{X}=\vec{x}][\alpha!\,]\neg\chi$, $[\vec{X}=\vec{x}][\alpha!\,](\chi_1 \wedge \chi_2)$, $[\vec{X}=\vec{x}][\alpha!\,]K\chi$ and $[\vec{X}=\vec{x}][\alpha_1!\,][\alpha_2!\,]\chi$ follow the same pattern: use the properties of c to obtain an appropriate IH, then apply $\mathbf{RE}_=$

[26] [44, Definition 7.21] use rather $\mathrm{c}([\alpha!\,]\chi) := (4 + \mathrm{c}(\alpha))\,\mathrm{c}(\chi)$. Here, 7 is the least natural number guaranteeing $\mathrm{c}([\vec{X}=\vec{x}][\alpha!\,]K\chi) > \mathrm{c}([\vec{X}=\vec{x}](\alpha \to K(\alpha \to [\alpha!\,]\chi)))$.

[27] Indeed, since $\alpha, \chi \in \mathrm{sub}([\alpha!\,]\chi)$, from IH it follows that $\vdash_{\mathsf{L}_{PAKC}} \alpha \leftrightarrow \mathrm{tr}_3(\alpha)$ and $\vdash_{\mathsf{L}_{PAKC}} \chi \leftrightarrow \mathrm{tr}_3(\chi)$. From the first and $\mathbf{{}_!RE}$ it follows that $\vdash_{\mathsf{L}_{PAKC}} [\alpha!\,]\chi \leftrightarrow [\mathrm{tr}_3(\alpha)!\,]\chi$; from the second and $\mathbf{RE_!}$ it follows that $\vdash_{\mathsf{L}_{PAKC}} [\mathrm{tr}_3(\alpha)!\,]\chi \leftrightarrow [\mathrm{tr}_3(\alpha)!\,]\,\mathrm{tr}_3(\chi)$. Then, by propositional reasoning (**P** and **MP**), $\vdash_{\mathsf{L}_{PAKC}} [\alpha!\,]\chi \leftrightarrow [\mathrm{tr}_3(\alpha)!\,]\,\mathrm{tr}_3(\chi)$, which by definition of tr_3 is the required $\vdash_{\mathsf{L}_{PAKC}} [\alpha!\,]\chi \leftrightarrow \mathrm{tr}_3([\alpha!\,]\chi)$.

(Lemma A.8) over the corresponding axiom, and finally use the definition of tr_3.[28] The case $[\vec{X}=\vec{x}][\vec{Y}=\vec{y}]\chi$ relies on axiom , just as its counterpart in Proposition A.2. For the lone case left,

- Case $[\vec{X}=\vec{x}][\alpha!][\vec{Y}=\vec{y}]\chi$. Use again IH, now on $c([\vec{X}=\vec{x}][\alpha!][\vec{Y}=\vec{y}]\chi) > c([[\vec{X}=\vec{x}]\alpha!][\vec{X}'=\vec{x}',\vec{Y}=\vec{y}]\chi)$ (with $\vec{X}'=\vec{x}'$ as indicated above) to obtain $\vdash_{\mathsf{L}_{PAKC}} [[\vec{X}=\vec{x}]\alpha!][\vec{X}'=\vec{x}',\vec{Y}=\vec{y}]\chi \leftrightarrow tr_3([[\vec{X}=\vec{x}]\alpha!][\vec{X}'=\vec{x}',\vec{Y}=\vec{y}]\chi)$. But $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha!][\vec{Y}=\vec{y}]\chi \leftrightarrow [[\vec{X}=\vec{x}]\alpha!][\vec{X}'=\vec{x}',\vec{Y}=\vec{y}]\chi$ (axiom ; see Lemma A.8). Then, by propositional reasoning, it follows that $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha!][\vec{Y}=\vec{y}]\chi \leftrightarrow tr_3([[\vec{X}=\vec{x}]\alpha!][\vec{X}'=\vec{x}',\vec{Y}=\vec{y}]\chi)$. This is enough, as the right-hand side of the latter is, by definition, $tr_3([\vec{X}=\vec{x}][\alpha!][\vec{Y}=\vec{y}]\chi)$.

(iii) By the previous item, $\vdash_{\mathsf{L}_{PAKC}} \chi \leftrightarrow tr_3(\chi)$. But is sound for recursive causal models; therefore, $\models \chi \leftrightarrow tr_3(\chi)$. □

Completeness, step 2: from $\mathcal{L}'_{PAKC}$ to $\mathcal{L}'_{KC}$. The translation from $\mathcal{L}'_{PAKC}$ to $\mathcal{L}'_{KC}$ uses $\mathbf{!}_=$, $\mathbf{!}_\neg$, $\mathbf{!}_\wedge$, and $\mathbf{!_K}$ for eliminating public announcements.

DEFINITION A.11 (Translation tr_4).
Define $tr_4 : \mathcal{L}'_{PAKC} \to \mathcal{L}'_{KC}$ as

$$tr_4([\vec{X}=\vec{x}]Y{=}y) := [\vec{X}=\vec{x}]Y{=}y$$
$$tr_4(\neg\theta) := \neg\, tr_4(\theta)$$
$$tr_4(\theta_1 \wedge \theta_2) := tr_4(\theta_1) \wedge tr_4(\theta_2)$$
$$tr_4(K\theta) := K\, tr_4(\theta)$$

$$tr_4([\alpha!][\vec{X}=\vec{x}]Y{=}y) := tr_4(\alpha \to [\vec{X}=\vec{x}]Y{=}y)$$
$$tr_4([\alpha!]\neg\theta) := tr_4(\alpha \to \neg[\alpha!]\theta)$$
$$tr_4([\alpha!](\theta_1 \wedge \theta_2)) := tr_4([\alpha!]\theta_1 \wedge [\alpha!]\theta_2)$$
$$tr_4([\alpha!]K\theta) := tr_4(\alpha \to K(\alpha \to [\alpha!]\theta))$$
$$tr_4([\alpha_1!][\alpha_2!]\theta) := tr_4([\alpha_1 \wedge [\alpha_1!]\alpha_2!]\theta)$$

PROPOSITION A.12
For every $\theta \in \mathcal{L}'_{PAKC}$,

(i) $tr_4(\theta) \in \mathcal{L}'_{KC}$,
(ii) $\vdash_{\mathsf{L}_{PAKC}} \theta \leftrightarrow tr_4(\theta)$,
(iii) $\models \theta \leftrightarrow tr_4(\theta)$.

[28] As an example, consider $[\vec{X}=\vec{x}][\alpha!]Y{=}y$. Since $c([\vec{X}=\vec{x}][\alpha!]Y{=}y) > c([\vec{X}=\vec{x}](\alpha \to Y{=}y))$, from IH it follows that $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}](\alpha \to Y{=}y) \leftrightarrow tr_3([\vec{X}=\vec{x}](\alpha \to Y{=}y))$. But applying $\mathbf{RE}_=$ on $\mathbf{!}_=$ yields $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha!]Y{=}y \leftrightarrow [\vec{X}=\vec{x}](\alpha \to Y{=}y)$, so $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha!]Y{=}y \leftrightarrow tr_3([\vec{X}=\vec{x}](\alpha \to Y{=}y))$. Then, from the definition of tr_3, it follows that $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha!]Y{=}y \leftrightarrow tr_3([\vec{X}=\vec{x}][\alpha!]Y{=}y)$. As another example consider $[\vec{X}=\vec{x}][\alpha_1!][\alpha_2!]\chi$. Since $c([\vec{X}=\vec{x}][\alpha_1!][\alpha_2!]\chi) > c([\vec{X}=\vec{x}][\alpha_1 \wedge [\alpha_1!]\alpha_2!]\chi)$, from IH it follows that $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha_1 \wedge [\alpha_1!]\alpha_2!]\chi \leftrightarrow tr_3([\vec{X}=\vec{x}][\alpha_1 \wedge [\alpha_1!]\alpha_2!]\chi)$. But applying $\mathbf{RE}_=$ on $\mathbf{!}_!$ yields $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha_1!][\alpha_2!]\chi \leftrightarrow [\vec{X}=\vec{x}][\alpha_1 \wedge [\alpha_1!]\alpha_2!]\chi$, so $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha_1!][\alpha_2!]\chi \leftrightarrow tr_3([\vec{X}=\vec{x}][\alpha_1 \wedge [\alpha_1!]\alpha_2!]\chi)$. Then, from the definition of tr_3, it follows that $\vdash_{\mathsf{L}_{PAKC}} [\vec{X}=\vec{x}][\alpha_1!][\alpha_2!]\chi \leftrightarrow tr_3([\vec{X}=\vec{x}][\alpha_1 \wedge [\alpha_1!]\alpha_2!]\chi)$.

PROOF. By induction $\mathrm{c} : \mathcal{L}'_{PAKC} \to \mathbb{N} \setminus \{0\}$, this time defined as

$$\mathrm{c}([\vec{X}{=}\vec{x}]Y{=}y) := 1 \qquad \mathrm{c}(K\theta) := 1 + \mathrm{c}(\theta)$$
$$\mathrm{c}(\neg\theta) := 1 + \mathrm{c}(\theta) \qquad \mathrm{c}([\alpha!]\theta) := (7 + \mathrm{c}(\alpha))\,\mathrm{c}(\theta)$$
$$\mathrm{c}(\theta_1 \wedge \theta_2) := 1 + \max\{\mathrm{c}(\theta_1), \mathrm{c}(\theta_2)\}$$

Thus, for every assignment $\vec{X}{=}\vec{x}$ and all formulas $\theta, \theta_1, \theta_2 \in \mathcal{L}'_{PAKC}$,

$$\mathrm{c}(\theta) \geqslant \mathrm{c}(\psi) \text{ for every } \psi \in \mathrm{sub}(\theta),$$

$$\begin{array}{ll} \mathrm{c}([\alpha!][\vec{X}{=}\vec{x}]Y{=}y) > \mathrm{c}(\alpha \to [\vec{X}{=}\vec{x}]Y{=}y) & \mathrm{c}([\alpha!]K\theta) > \mathrm{c}(\alpha \to K(\alpha \to [\alpha!]\theta)) \\ \mathrm{c}([\alpha!]\neg\theta) > \mathrm{c}(\alpha \to \neg[\alpha!]\theta) & \mathrm{c}([\alpha_1!][\alpha_2!]\theta) > \mathrm{c}([\alpha_1 \wedge [\alpha_1!]\alpha_2!]\theta) \\ \mathrm{c}([\alpha!](\theta_1 \wedge \theta_2)) > \mathrm{c}([\alpha!]\theta_1 \wedge [\alpha!]\theta_2) & \end{array}$$

The first item is proved by structural induction; the rest are small variations of the well-known cases from public announcement logic [44, Section 7.4].

The first two proofs use induction on $\mathrm{c}(\theta)$. The base cases are for formulas θ with $\mathbf{c(\theta) = 1}$ (i.e. $[\vec{X}{=}\vec{x}]Y{=}y$); the inductive cases are for formulas θ with $\mathbf{c(\theta) > 1}$ (i.e. $\neg\theta$, $\theta_1 \wedge \theta_2$, $K\theta$, $[\alpha!][\vec{X}{=}\vec{x}]Y{=}y$, $[\alpha!]\neg\theta$, $[\alpha!](\theta_1 \wedge \theta_2)$, $[\alpha!]K\theta$ and $[\alpha_1!][\alpha_2!]\theta$).

(i) Analogous to that in Proposition A.6. The base case is straightforward. For the inductive part, use the properties of c to generate appropriate IHs, and then use the definition of tr_4.

(ii) Analogous to that in Proposition A.6. The base case is by propositional reasoning. The inductive cases rely on the properties of c to generate appropriate IHs, and then use the definition of tr_4 with propositional reasoning (cases $\neg\theta, \theta_1 \wedge \theta_2$) plus $\mathbf{RE_K}$ (case $K\theta$), axiom $\mathbf{!_=}$ (case $[\alpha!][\vec{X}{=}\vec{x}]Y{=}y$), axiom (case $[\alpha!]\neg\theta$), axiom $\mathbf{!_\wedge}$ (case $[\alpha!](\theta_1 \wedge \theta_2)$, axiom $\mathbf{!_K}$ (case $[\alpha!]K\theta$) and axiom $\mathbf{!_!}$(case $[\alpha_1!][\alpha_2!]\theta$).[29]

(iii) By the previous item, $\vdash_{\mathsf{L}_{PAKC}} \theta \leftrightarrow \mathrm{tr}_4(\theta)$. But is sound within recursive causal models; therefore, $\models \theta \leftrightarrow \mathrm{tr}_4(\theta)$. □

Then, for strong completeness of for formulas in , define $\mathrm{tr} : \mathcal{L}_{PAKC} \to \mathcal{L}'_{KC}$ as $\mathrm{tr}(\chi') := \mathrm{tr}_4(\mathrm{tr}_3(\chi'))$. Note: tr is properly defined because $\mathrm{tr}_3(\chi') \in \mathcal{L}'_{PAKC}$ (Proposition A.10.*(i)*). Moreover, for every $\chi' \in \mathcal{L}_{PAKC}$,

- $\mathrm{tr}(\chi') \in \mathcal{L}'_{KC}$ (from Proposition A.12.*(i)*).
- $\vdash_{\mathsf{L}_{PAKC}} \chi' \leftrightarrow \mathrm{tr}(\chi')$ (from Proposition A.10.*(ii)*, Proposition A.10.*(i)* and Proposition A.12.*(ii)*).
- $\models \chi' \leftrightarrow \mathrm{tr}(\chi')$ (from Proposition A.10.*(iii)*, Proposition A.10.*(i)* and Proposition A.12.*(iii)*).

[29] As an example, consider $[\alpha!][\vec{X}{=}\vec{x}]Y{=}y$. Since $\mathrm{c}([\alpha!][\vec{X}{=}\vec{x}]Y{=}y) > \mathrm{c}(\alpha \to [\vec{X}{=}\vec{x}]Y{=}y)$, from IH it follows that $\vdash_{\mathsf{L}_{PAKC}} (\alpha \to [\vec{X}{=}\vec{x}]Y{=}y) \leftrightarrow \mathrm{tr}_4(\alpha \to [\vec{X}{=}\vec{x}]Y{=}y)$. But $\mathbf{!_=}$ yields $\vdash_{\mathsf{L}_{PAKC}} [\alpha!][\vec{X}{=}\vec{x}]Y{=}y \leftrightarrow (\alpha \to [\vec{X}{=}\vec{x}]Y{=}y)$, so $\vdash_{\mathsf{L}_{PAKC}} [\alpha!][\vec{X}{=}\vec{x}]Y{=}y \leftrightarrow \mathrm{tr}_4(\alpha \to [\vec{X}{=}\vec{x}]Y{=}y)$. Then, from the definition of tr_4, it follows that $\vdash_{\mathsf{L}_{PAKC}} [\alpha!][\vec{X}{=}\vec{x}]Y{=}y \leftrightarrow \mathrm{tr}_4([\alpha!][\vec{X}{=}\vec{x}]Y{=}y)$. As another example, consider $[\alpha_1!][\alpha_2!]\theta$. Since $\mathrm{c}([\alpha_1!][\alpha_2!]\theta) > \mathrm{c}([\alpha_1 \wedge [\alpha_1!]\alpha_2!]\theta)$, from IH it follows that $\vdash_{\mathsf{L}_{PAKC}} [\alpha_1 \wedge [\alpha_1!]\alpha_2!]\theta \leftrightarrow \mathrm{tr}_4([\alpha_1 \wedge [\alpha_1!]\alpha_2!]\theta)$. But $\mathbf{!_!}$ yields $\vdash_{\mathsf{L}_{PAKC}} [\alpha_1!][\alpha_2!]\theta \leftrightarrow [\alpha_1 \wedge [\alpha_1!]\alpha_2!]\theta$, so $\vdash_{\mathsf{L}_{PAKC}} [\alpha_1!][\alpha_2!]\theta \leftrightarrow \mathrm{tr}_4([\alpha_1 \wedge [\alpha_1!]\alpha_2!]\theta)$. Then, from the definition of tr_4, it follows that $\vdash_{\mathsf{L}_{PAKC}} [\alpha_1!][\alpha_2!]\theta \leftrightarrow \mathrm{tr}_4([\alpha_1!][\alpha_2!]\theta)$.

With these three properties, the argument for strong completeness of for formulas in is as that for Theorem 2.10 (Page 52).

A.4 Proof of Theorem 5.6

Soundness. The soundness of most axioms and rules over epistemic causal models with observables is straightforward. We will only explicitly consider the axioms from Table 1 and **PD**. The first are taken care of by the following lemma, which can be proved by a straightforward structural induction.

LEMMA A.13
Let ϕ be an $\mathcal{L}_{PAKC}$ formula without occurrences of K or announcement operators. Let $\mathcal{E} = \langle \mathscr{S}, \mathcal{F}, \mathcal{T}\rangle$ be epistemic causal model with observables; take $\mathcal{A} \in \mathcal{T}$. Then:

$$(\mathcal{E}, \mathcal{A}) \models^{O} \phi \quad \text{if and only if} \quad \langle \mathcal{A}, \mathcal{F}\rangle \models \phi.$$

For axiom **PD**, we have the following.

LEMMA A.14
Let $\mathcal{E} = \langle \mathscr{S}, \mathcal{F}, \mathcal{T}\rangle$ be an epistemic causal model with observables and $\mathcal{A} \in \mathcal{T}$. Let $\vec{o} = \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}(\boldsymbol{\mathcal{O}})$; write δ for $[\vec{X} = \vec{x}]\boldsymbol{\mathcal{O}} = \vec{o}$. Then,

$$\mathcal{B} \in \mathcal{T}^{\delta} \text{ implies } (\mathcal{E}^{\delta})^{\mathcal{B}}_{\vec{X}=\vec{x}} = (\mathcal{E}^{\delta})^{\mathcal{A}}_{\vec{X}=\vec{x}} = \mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}} = \mathcal{E}^{\mathcal{B}}_{\vec{X}=\vec{x}}.$$

PROOF. From $\mathcal{B} \in \mathcal{T}^{\delta}$ it follows that $(\mathcal{E}^{\delta}, \mathcal{B}) \models [\vec{X} = \vec{x}]\boldsymbol{\mathcal{O}} = \vec{o}$; then we get $((\mathcal{E}^{\delta})^{\mathcal{B}}_{\vec{X}=\vec{x}}, \mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \boldsymbol{\mathcal{O}} = \vec{o}$, which implies $\mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}(\boldsymbol{\mathcal{O}}) = \vec{o} = \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}(\boldsymbol{\mathcal{O}})$. This yields the first equality; the third is proved analogously, using $\mathcal{E}$ instead of $\mathcal{E}^{\delta}$.

For the second equality, observe that any element of $\mathcal{T}^{\mathcal{F},\mathcal{A}}_{\vec{X}=\vec{x}}$ is of the form $\mathcal{C}^{\mathcal{F}}_{\vec{X}=\vec{x}}$ for some $\mathcal{C} \in \mathcal{T}$. However, by the definition of epistemic model with observables, $\mathcal{C}^{\mathcal{F}}_{\vec{X}=\vec{x}}(\boldsymbol{\mathcal{O}}) = \vec{o}$; thus, $(\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}}, \mathcal{C}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \boldsymbol{\mathcal{O}} = \vec{o}$. Therefore, $(\mathcal{E}, \mathcal{C}) \models [\vec{X} = \vec{x}]\boldsymbol{\mathcal{O}} = \vec{o}$, i.e. $\mathcal{C} \in \mathcal{T}^{\delta}$. Since $\mathcal{C}^{\mathcal{F}}_{\vec{X}=\vec{x}}(\boldsymbol{\mathcal{O}}) = \vec{o}$, then, $\mathcal{C}^{\mathcal{F}}_{\vec{X}=\vec{x}} \in (\mathcal{T}^{\delta})^{\mathcal{F},\mathcal{A}}_{\vec{X}=\vec{x}}$. □

THEOREM A.15
Axiom **PD** is sound on epistemic causal models with observables.

PROOF. If case $\vec{X}$ is the empty set, axiom **PD** reduces to:

$$K\gamma \leftrightarrow \bigvee_{\vec{o} \in \boldsymbol{\mathcal{R}}(O)} (\boldsymbol{\mathcal{O}} = \vec{o} \wedge [\boldsymbol{\mathcal{O}} = \vec{o}!\,]K\gamma).$$

Thus, from left to right, assume $(\mathcal{E}, \mathcal{A}) \models K\gamma$. Let $\vec{o} = \mathcal{A}(\boldsymbol{\mathcal{O}})$. Then $(\mathcal{E}, \mathcal{A}) \models \boldsymbol{\mathcal{O}} = \vec{o}$. By definition of epistemic causal models with observables, $(\mathcal{E}, \mathcal{B}) \models \boldsymbol{\mathcal{O}} = \vec{o}$ for all $\mathcal{B} \in \mathcal{T}$. So $\mathcal{T}^{\boldsymbol{\mathcal{O}}=\vec{o}} = \mathcal{T}$ and therefore $(\mathcal{E}^{\boldsymbol{\mathcal{O}}=\vec{o}}, \mathcal{A}) \models K\gamma$. Thus, $(\mathcal{E}, \mathcal{A}) \models [\boldsymbol{\mathcal{O}} = \vec{o}!\,]K\gamma$. From right to left, assume $(\mathcal{E}, \mathcal{A}) \models \bigvee_{\vec{o} \in \mathcal{O}}(\mathcal{O} = \vec{o} \wedge [\boldsymbol{\mathcal{O}} = \vec{o}!\,]K\gamma)$. Then $(\mathcal{E}, \mathcal{A}) \models \boldsymbol{\mathcal{O}} = \vec{o}$ and $(\mathcal{E}, \mathcal{A}) \models [\boldsymbol{\mathcal{O}} = \vec{o}!\,]K\gamma$ for some $\vec{o} \in \boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{O}})$. The latter implies $(\mathcal{E}^{\boldsymbol{\mathcal{O}}=\vec{o}}, \mathcal{A}) \models K\gamma$; the former implies $\mathcal{E} = \mathcal{E}^{\boldsymbol{\mathcal{O}}=\vec{o}}$, as before. Thus, $(\mathcal{E}, \mathcal{A}) \models K\gamma$.

Now, suppose $\vec{X}$ is non-empty; let $\mathcal{E} = \langle \mathscr{S}, \mathcal{T}, \mathcal{F}\rangle$. From left to right, assume $(\mathcal{E}, \mathcal{A}) \models [\vec{X} = \vec{x}]K\gamma$. We have to show the right-hand side of the axiom for some $\vec{o} \in \boldsymbol{\mathcal{R}}(\boldsymbol{\mathcal{O}})$; we will show it for $\vec{o} := \mathcal{A}_{\vec{X}=\vec{x}}(\boldsymbol{\mathcal{O}})$. The assumption implies

$$(\star) : (\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}}, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models K\gamma.$$

so, for every $\mathcal{B} \in \mathcal{T}^{\mathcal{F},\mathcal{A}}_{\vec{X}=\vec{x}}$, we have $(\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}}, \mathcal{B}) \models \gamma$. Notice that $(\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}}, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \boldsymbol{\mathcal{O}} = \vec{o}$; thus, $(\mathcal{E},\mathcal{A}) \models [\vec{X} = \vec{x}]\boldsymbol{\mathcal{O}} = \vec{o}$. This is the first conjunct we needed to prove. Now write δ for $[\vec{X} = \vec{x}]\boldsymbol{\mathcal{O}} = \vec{o}$. Take $\mathcal{C} \in \mathcal{T}^{\delta}$; we want to verify that $(\mathcal{E}^{\delta},\mathcal{C}) \models [\vec{X} = \vec{x}]\gamma$. By Lemma A.14, we have $(\mathcal{E}^{\delta})^{\mathcal{C}}_{\vec{X}=\vec{x}} = \mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}}$. Thus, $(\mathcal{E}^{\delta},\mathcal{C}) \models [\vec{X} = \vec{x}]\gamma$ iff $((\mathcal{E}^{\delta})^{\mathcal{C}}_{\vec{X}=\vec{x}}, \mathcal{C}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \gamma$ iff $(\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}}, \mathcal{C}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \gamma$, which is true by ($\star$). Since $(\mathcal{E}^{\delta},\mathcal{C}) \models [\vec{X} = \vec{x}]\gamma$ for each $\mathcal{C} \in \mathcal{E}^{\delta}$, we have $(\mathcal{E}^{\delta},\mathcal{A}) \models K[\vec{X} = \vec{x}]\gamma$. We then conclude $(\mathcal{E},\mathcal{A}) \models [\delta!\,]K[\vec{X} = \vec{x}]\gamma$, as needed.

From right to left, assume there is $\vec{o} \in \boldsymbol{\mathcal{O}}$ with $(\mathcal{E},\mathcal{A}) \models [\vec{X} = \vec{x}]\boldsymbol{\mathcal{O}} = \vec{o}$ and $(\mathcal{E},\mathcal{A}) \models [\delta!\,]K[\vec{X} = \vec{x}]\gamma$. For any $\mathcal{C} \in \mathcal{T}^{\mathcal{F},\mathcal{A}}_{\vec{X}=\vec{x}}$ we have $\mathcal{C} = \mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}$ for some $\mathcal{B} \in \mathcal{T}$. Thus, $\mathcal{B} \in \mathcal{T}^{\delta}$ (as $\mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}(\boldsymbol{\mathcal{O}}) = \vec{o}$). So, by Lemma A.14, $(\mathcal{E}^{\delta})^{\mathcal{B}}_{\vec{X}=\vec{x}} = \mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}}$. By assumption, $(\mathcal{E}^{\delta},\mathcal{A}) \models K[\vec{X} = \vec{x}]\gamma$, so $((\mathcal{E}^{\delta})^{\mathcal{B}}_{\vec{X}=\vec{x}}, \mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \gamma$. Therefore $(\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}}, \mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models \gamma$ and then $(\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}}, \mathcal{C}) \models \gamma$. Since $\mathcal{C}$ is an arbitrary assignment in $\mathcal{T}^{\mathcal{F},\mathcal{A}}_{\vec{X}=\vec{x}}$, we have $(\mathcal{E}^{\mathcal{A}}_{\vec{X}=\vec{x}}, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models K\gamma$ and thus $(\mathcal{E},\mathcal{A}) \models [\vec{X} = \vec{x}]K\gamma$. □

Completeness via reduction to the case without observables The first step consists in eliminating the announcement operators. This is done using axioms $\mathbf{!_=}$, $\mathbf{!_\neg}$, $\mathbf{!_\wedge}$, and $\mathbf{!_K}$ and **PD** together with the following generalization of the replacement rules proposed before:

$$\mathbf{RE_{full}}: \quad \text{if } \vdash \chi_1 \leftrightarrow \chi_2 \text{ then } \vdash \phi \leftrightarrow \phi[\chi_2/\chi_1]$$

where $\phi[\chi_2/\chi_1]$ obtained by replacing, in ϕ, some occurrences of χ_1 with χ_2.

PROPOSITION A.16
The rule $\mathbf{RE_{full}}$ is admissible in L^{O}_{PAKC}.

PROOF. (Sketch) The rules $\mathbf{RE_K}$, $\mathbf{RE_=}$, $\mathbf{RE_!}$ and $\mathbf{{}_!RE}$, and are provable as in . The main claim can then be proved by induction on ϕ: the base case is trivial, the Boolean cases rely on using classical logic (+), and the cases for $\phi = K\psi$ (resp. $[\vec{X} = \vec{x}]\psi$, $[\alpha!\,]\psi$) are covered by $\mathbf{RE_K}$ (resp. $\mathbf{RE_=}$, $\mathbf{{}_!RE}$ $+\mathbf{RE_!}$). □

PROPOSITION A.17
(i) Every formula $\phi \in \mathcal{L}_{PAKC}$ is logically equivalent to a formula $\xi_\phi \in \mathcal{L}'_{PAKC}$. Moreover, $\phi \leftrightarrow \xi_\phi$ is derivable in L^{O}_{PAKC}. (ii) Every formula $\xi \in \mathcal{L}'_{PAKC}$ is logically equivalent to a formula $\chi_\xi \in \mathcal{L}'_{KC}$. Moreover, $\xi \leftrightarrow \chi_\xi$ is derivable in L^{O}_{PAKC}.

PROOF. For the first, modify the translation tr_3 (Definition A.9) by taking

$$\mathrm{tr}_3([\vec{X} = \vec{x}]K\gamma) := \bigvee_{\vec{o}\in\mathcal{O}} \Big([\vec{X} = \vec{x}]\mathcal{O} = \vec{o} \wedge \big[[\vec{X} = \vec{x}]\mathcal{O} = \vec{o}!\,\big]K[\vec{X} = \vec{x}]\mathrm{tr}_3(\gamma)\Big).$$

The correctness of this clause is proved by axiom **PD** and rule **RE**. This clause allows removing instances of K from the consequents of counterfactuals. For the second, use the same translation as in the case without observables (Definition A.11); see the proof of Proposition A.12. □

For the completeness of L^{O}_{PAKC}, by Proposition A.17 it suffices to show that $\mathsf{L}^{O}_{KC} := \mathsf{L}^{O}_{PAKC} \setminus \{\mathbf{A_{[]}}, \mathbf{A_\neg}, \mathbf{A_\wedge}, \mathbf{A_{[][]}}, \mathbf{!_=}, \mathbf{!_\neg}, \mathbf{!_\wedge}, \mathbf{!_K}, \mathbf{PD}\}$ is complete for $\mathcal{L}'_{KC}$ (in the semantics with observables). Since $\mathsf{L}^{O}_{KC} = \mathsf{L}_{KC} \cup \{\mathbf{OC}\}$, we have the following.

PROPOSITION A.18
Let $\Gamma \subseteq \mathcal{L}'_{KC}$. Then

$$\Gamma \text{ is } \mathsf{L}^{O}_{KC}\text{-consistent} \quad \text{if and only if} \quad \Gamma \cup \{\mathbf{OC}\} \text{ is } \mathsf{L}_{KC}\text{-consistent}.$$

Thus, for finding a model for a L^{O}_{KC}-consistent set of $\mathcal{L}'_{KC}$ formulas Γ, use the completeness theorem for the case without observables, which provides a pointed model for $\Gamma \cup \{\mathbf{OC}\}$, i.e. an epistemic causal model $\mathcal{E}$ together with an assignment $\mathcal{A}$ such that $(\mathcal{E},\mathcal{A}) \models^{W} \Gamma \cup \{\mathbf{OC}\}$. If we prove that $(\mathcal{E}^{*},\mathcal{A}) \models^{O} \Gamma \cup \{\mathbf{OC}\}$ (where $\mathcal{E}^{*}$ differs from $\mathcal{E}$ only in that its signature has a set of observables), we are done. But this is provided by the following result. We write $\mathbf{OC}_{\mathscr{S}^{*}}$ for the axiom scheme specialized to the signature $\mathscr{S}^{*}$.

PROPOSITION A.19
Let $\mathscr{S} = \langle \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{R}} \rangle$ be a signature and $\mathcal{E} = \langle \mathscr{S}, \mathcal{T}, \mathcal{F} \rangle$ epistemic causal model. Let $\boldsymbol{\mathcal{O}}$ be a subset of $\boldsymbol{\mathcal{U}} \cup \boldsymbol{\mathcal{V}}$ and $\mathscr{S}^{*} = \langle \boldsymbol{\mathcal{U}}, \boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{O}}, \boldsymbol{\mathcal{R}} \rangle$ be the corresponding signature with observables. Suppose that $(\mathcal{E},\mathcal{A}) \models^{W} \mathbf{OC}_{\mathscr{S}^{*}}$ for some $\mathcal{A} \in \mathcal{T}$. Then,

- (i) The tuple $\boldsymbol{\mathcal{O}}$ takes constant value in $\mathcal{T}$; therefore $\mathcal{E}^{*} = \langle \mathscr{S}, \mathcal{T}, \mathcal{F} \rangle$ is an epistemic causal model with observables $\boldsymbol{\mathcal{O}}$.
- (ii) For all $\mathcal{B} \in \mathcal{T}$ and all $\varphi \in \mathcal{L}_{PAKC}$: $(\mathcal{E},\mathcal{B}) \models^{W} \varphi \iff (\mathcal{E}^{*},\mathcal{B}) \models^{O} \varphi$

PROOF. (i) This follows from $(\mathcal{E},\mathcal{A}) \models \bigvee_{\vec{o}\in\mathcal{O}} K(\boldsymbol{\mathcal{O}} = \vec{o})$, the special instance of $\mathbf{OC}_{\mathscr{S}^{*}}$ for $\vec{X} = \emptyset$. (ii) Use induction on φ. For the only non-trivial case, $\varphi = [\vec{X} = \vec{x}]\gamma$, we have $(\mathcal{E},\mathcal{B}) \models^{W} [\vec{X} = \vec{x}]\gamma$ iff $(\mathcal{E}_{\vec{X}=\vec{x}}, \mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models^{W} \gamma$. But $(\mathcal{E},\mathcal{A}) \models^{W} [\vec{X} = \vec{x}] \bigvee_{\vec{o}\in\boldsymbol{\mathcal{O}}} K\boldsymbol{\mathcal{O}} = \vec{o}$, so $(\mathcal{E}_{\vec{X}=\vec{x}}, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models^{W} K\boldsymbol{\mathcal{O}} = \vec{o}$ for some value $\vec{o} \in \boldsymbol{\mathcal{R}}(O)$. Then $(\mathcal{E}_{\vec{X}=\vec{x}})^{*}$, the epistemic causal model with observables that differs from $\mathcal{E}_{\vec{X}=\vec{x}}$ only in that it has signature $\mathscr{S}^{*}$ instead of signature $\mathscr{S}$, is an epistemic causal model with observables $\boldsymbol{\mathcal{O}}$. Then we can use the inductive hypothesis to obtain (*): $((\mathcal{E}_{\vec{X}=\vec{x}})^{*}, \mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models^{O} \gamma$. But now observe that, since $(\mathcal{E}_{\vec{X}=\vec{x}}, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models^{W} K\boldsymbol{\mathcal{O}} = \vec{o}$, also $(\mathcal{E}_{\vec{X}=\vec{x}}, \mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models^{W} K\boldsymbol{\mathcal{O}} = \vec{o}$; thus $(\mathcal{E}_{\vec{X}=\vec{x}})^{*} = \mathcal{E}_{\vec{X}=\vec{x}} = \mathcal{E}^{\mathcal{B}}_{\vec{X}=\vec{x}} = (\mathcal{E}^{*})^{\mathcal{B}}_{\vec{X}=\vec{x}}$. Hence, (*) is equivalent to $((\mathcal{E}^{*})^{\mathcal{B}}_{\vec{X}=\vec{x}}, \mathcal{B}^{\mathcal{F}}_{\vec{X}=\vec{x}}) \models^{O} \gamma$, and then to $(\mathcal{E}^{*},\mathcal{B}) \models^{O} [\vec{X} = \vec{x}]\gamma$. □

THEOREM A.20 (Completeness with observables).
Let $\Gamma \cup \{\phi\}$ be a set of formulas. Then:

$$\Gamma \models^{O} \phi \iff \Gamma \vdash_{\mathsf{L}^{O}_{PAKC}} \phi.$$

PROOF. By Proposition A.17, we can assume that Γ and ϕ are in $\mathcal{L}'_{KC}$. By the usual arguments, the statement is then equivalent to the assertion that every maximally L^{O}_{KC}-consistent set of formulas Γ is true in some pair $(\mathcal{E},\mathcal{A})$, where $\mathcal{E}$ is an epistemic causal model with observables. Now, if Γ is L^{O}_{KC}-consistent, then (Proposition A.18) it is L_{KC}-consistent; thus, by completeness for the case without observables (Theorem 3.8) there is an $(\mathcal{E},\mathcal{A})$ that satisfies Γ. Since Γ is maximally L^{O}_{KC}-consistent, in particular $\mathbf{OC} \in \Gamma$. Thus, by Proposition A.19, there is an $\mathcal{E}^{*}$ with observables such that $(\mathcal{E}^{*},\mathcal{A}) \models \Gamma$, as needed. □

References

[1] C. Areces and B. ten Cate. 14 Hybrid logics. In P. Blackburn, J. van Benthem and F. Wolter., eds, *Handbook of Modal Logic*, pp. 821–868. Elsevier, North Holland, 2006. ISBN 978–0–444-51690-9. doi: 10.1016/S1570-2464(07)80017-6.

[2] A. Baltag. To know is to know the value of a variable. *Advances in Modal Logic*, **11**, 135–155, 2016.

[3] A. Baltag and S. Smets. Probabilistic dynamic belief revision. *Synthese*, **165**, 179–202, 2008. doi: 10.1007/s11229-008-9369-8.

[4] A. Baltag, L. S. Moss and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. In I. Gilboa., ed., *TARK*, pp. 43–56, San Francisco, CA, USA, 1998. Morgan Kaufmann. ISBN 1-55860-563-0. URL http://dl.acm.org/citation.cfm?id=645876.671885.

[5] A. Baltag, I. Canavotto and S. Smets. Causal agency and responsibility: a refinement of STIT logic. In A. Giordani and J. Malinowski., eds, *Logic in High Definition: Trends in Logical Semantics*, pp. 149–176. Springer, 2021. doi: 10.1007/978-3-030-53487-5_8.

[6] B. Finkbeiner and S. Kleinberg. Proceedings 3rd Workshop on formal reasoning about Causation, Responsibility, and Explanations in Science and Technology. In *EPTCS*, B. Finkbeiner and S. Kleinberg., eds, pp. 16–30. Open Publishing Association, 2019.

[7] F. Barbero and G. Sandu. Team semantics for interventionist counterfactuals: observations vs. interventions. *Journal of Philosophical Logic*, **50**, 471–521, 2021. doi: 10.1007/s10992-020-09573-6.

[8] F. Barbero and F. Yang. Counterfactuals and dependencies on causal teams: expressive power and deduction systems. In N. Olivetti, R. Verbrugge, S. Negri and G. Sandu., eds, *13th Conference on Advances in Modal Logic, AiML 2020, Helsinki, Finland, August 24–28, 2020*, pages 73–93. College Publications, 2020. ISBN 978–1–84890-341-8. URL http://www.aiml.net/volumes/volume13/Barbero-Yang.pdf.

[9] F. Barbero, K. Schulz, S. Smets, F. R. Velázquez-Quesada and K. Xie. Thinking about causation: a causal language with epistemic operators. In M. A. Martins and I. Sedlár., eds, *Dynamic Logic. New Trends and Applications–Third International Workshop, DaLí 2020, Prague, Czech Republic, October 9–10, 2020, Revised Selected Papers*, vol. 12569 of Lecture Notes in Computer Science, pp. 17–32. Springer, 2020. ISBN 978-3-030-65839-7. doi: 10.1007/978-3-030-65840-3_2.

[10] F. Barbero, K. Schulz, F. R. Velázquez-Quesada and K. Xie. *Observing interventions: a logic for thinking about experiments*, 2021. URL https://arxiv.org/abs/2111.12978.

[11] B. Bergstein. *What AI still can't do*, 2020. https://www.technologyreview.com/2020/02/19/868178/what-ai-still-cant-do/ [Accessed: June 2020].

[12] P. Blackburn, M. de Rijke and Y. Venema. *Modal logic. Number 53 in Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, Cambridge, UK, 2001. ISBN 0–521–80200-8. doi: 10.1017/CBO9781107050884.

[13] T. Braüner. Hybrid logic. In E. N. Zalta., ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2017 edition, 2017. URL https://plato.stanford.edu/archives/sum2017/entries/logic-hybrid/.

[14] R. Briggs. *Philosophical Studies*, **160**, 139–166, 2012.

[15] R. Fagin and J. Y. Halpern. Reasoning about knowledge and probability. *Journal of the ACM*, **41**, 340–367, 1994. doi: 10.1145/174652.174658.

[16] R. Fagin, J. Y. Halpern, Y. Moses and M. Y. Vardi. *Reasoning about knowledge*. The MIT Press, Cambridge, MA, 1995. ISBN 0–262–06162-7.

[17] D. Galles and J. Pearl. Axioms of causal relevance. *Artificial Intelligence*, **97**, 9–43, 1997. doi: 10.1016/S0004-3702(97)00047-7.

[18] D. Galles and J. Pearl. An axiomatic characterisation of causal counterfactuals. *Foundations of Science*, **3**, 151–182, 1998. doi: 10.1023/A:1009602825894.

[19] P. Galliani. Inclusion and exclusion dependencies in team semantics—on some logics of imperfect information. *Annals of Pure and Applied Logic*, **163**: 68–84, 2012. doi: 10.1016/j.apal.2011.08.005.

[20] J. Gerbrandy and W. Groeneveld. Reasoning about information change. *Journal of Logic, Language and Information*, **6**, 147–169, 1997. doi: 10.1023/A:1008222603071.

[21] E. Grädel and J. A. Väänänen. Dependence and independence. *Studia Logica*, **101**, 399–410, 2013. doi: 10.1007/s11225-013-9479-2.

[22] J. Y. Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, **12**, 317–337, 2000. 10.1613/jair.648.

[23] J. Y. Halpern. *Actual Causality*. The MIT Press, 2016.

[24] J. Y. Halpern. *Reasoning About Uncertainty*, 2nd edn. The MIT Press, Cambridge, MA, USA, 2017. ISBN 978–0–262533-80-5.

[25] J. Y. Halpern, R. van der Meyden and M. Y. Vardi. Complete axiomatizations for reasoning about knowledge and time. *SIAM Journal on Computing*, **33**, 674–703, 2004. doi: 10.1137/S0097539797320906.

[26] J. Hintikka. *Knowledge and Belief*. Cornell University Press, Ithaca, NY, 1962 ISBN 1–904987–08-7.

[27] C. Hitchcock. A tale of two effects. *Philosophical Review*, **110**, 361–396, 2001. doi: 10.1215/00318108-110-3-361.

[28] W. Hodges. Compositional semantics for a language of imperfect information. *Logic Journal of the IGPL*, **5**, 539–563, 1997. doi: 10.1093/jigpal/5.4.539. URL https://doi.org/10.1093/jigpal/5.4.539.

[29] D. Ibeling and T. Icard. On open-universe causal reasoning. In A. Globerson and R. Silva., eds, *Proceedings of the Thirty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI 2019, Tel Aviv, Israel, July 22–25, 2019*, vol. 115 of Proceedings of Machine Learning Research, pp. 1233–1243. AUAI Press, 2019. URL http://proceedings.mlr.press/v115/ibeling20a.html.

[30] D. Ibeling and T. Icard. Probabilistic reasoning across the causal hierarchy. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 10170–10177, 2020. doi: 10.1609/aaai.v34i06.6577. URL https://ojs.aaai.org/index.php/AAAI/article/view/6577.

[31] B. P. Kooi. Probabilistic dynamic epistemic logic. *Journal of Logic, Language and Information*, **12**, 381–408, 2003. doi: 10.1023/A:1025050800836.

[32] K. Martiny and R. Möller. A probabilistic doxastic temporal logic for reasoning about beliefs in multi-agent systems. In *ICAART 2015—Proceedings of the International Conference on Agents and Artificial Intelligence, Volume 2, Lisbon, Portugal, 10–12 January, 2015*, S. Loiseau, J. Filipe, B. Duval and H. J. van den Herik., eds, pp. 277–284. SciTePress, 2015.

[33] R. Nozick. *Philosophical Explanations*. Harvard University Press, Cambridge, Massachusetts, 1981.

[34] R. Parikh and R. Ramanujam. A knowledge based semantics of messages. *Journal of Logic, Language and Information*, **12**, 453–467, 2003. doi: 10.1023/A:1025007018583.

[35] J. Pearl. Causal diagrams for empirical research. *Biometrika*, **82**, 669–688, 1995.

[36] J. Pearl. *Causality. Models, Reasoning, and Inference*, 2nd edn. Cambridge University Press, Cambridge, 2009.

[37] J. Pearl and D. Mackenzie. The book of why. In *The New Science of Cause and Effect*. Penguin Books UK, 2019.

[38] J. Peters, D. Janzing and B. Scholkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*, 1st edn. The MIT Press, 2017.

[39] J. A. Plaza. Logics of public communications. In *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, M. L. Emrich, M. S. Pfeifer, M. Hadzikadic and Z. W. Ras., eds, pp. 201–216. Oak Ridge National Laboratory, ORNL/DSRD-24, Tennessee, USA, 1989.

[40] P. Spirtes, C. Glymour and R. Scheines. *Causation, Prediction, and Search*, 1st edn. Springer New York, 1993.

[41] R. H. Strotz and H. O. Wold. Recursive vs. nonrecursive systems: an attempt at synthesis (part I of a triptych on causal chain systems). *Econometrica: Journal of the Econometric Society*, **28**, 417–427, 1960.

[42] J. Väänänen. *Dependence Logic: A New Approach to Independence Friendly Logic*, vol. 70 of London Mathematical Society Student Texts. Cambridge University Press, 2007.

[43] J. van Benthem. *Logical Dynamics of Information and Interaction*. Cambridge University Press, 2011. ISBN 978–0–521-76579-4.

[44] H. van Ditmarsch, W. van der Hoek and B. Kooi. *Dynamic Epistemic Logic*, vol. 337 of Synthese Library Series. Springer, Dordrecht, The Netherlands, 2008. ISBN 978–1–4020-5838-7. doi: 10.1007/978-1-4020-5839-4.

[45] J. van Eijck, M. Gattinger and Y. Wang. Knowing values and public inspection. In S. Ghosh and S. Prasad., eds, *Logic and Its Applications—7th Indian Conference, ICLA 2017, Kanpur, India, January 5–7, 2017, Proceedings*, vol. 10119 of *Lecture Notes in Computer Science*, pages 77–90. Springer, 2017. ISBN 978-3-662-54068-8. doi: 10.1007/978-3-662-54069-5_7.

[46] Y. Wang and Q. Cao. On axiomatizations of public announcement logic. *Synthese*, **190**, 103–134, 2013. 10.1007/s11229-012-0233-5.

[47] J. Woodward. *Making Things Happen*, vol. 114 of Oxford Studies in the Philosophy of Science. Oxford University Press, 2003.

Received 31 March 2021