Annals of Pure and Applied Logic 173 (2022) 103159

Contents lists available at ScienceDirect

Annals of Pure and Applied Logic

www.elsevier.com/locate/apal

Embedding causal team languages into predicate logic ☆

Fausto Barbero [a,*], Pietro Galliani [b]

[a] University of Sheffield, 211 Portobello, S1 4DP, Sheffield, United Kingdom
[b] Free University of Bozen-Bolzano, Universitätsplatz 1 – Piazza Università 1, 39100, Bozen-Bolzano, Italy

ARTICLE INFO

Article history:
Available online 22 June 2022

MSC:
03C80
03B60
03B70
03B80
03B25
62-09

Keywords:
Interventionist counterfactuals
Causal teams
Dependence logic
Embedding
Bernays-Schönfinkel-Ramsey fragment
Cyclic causation

ABSTRACT

Causal team semantics ([2]) supports causal-observational languages, which enrich the languages for deterministic causation ([11,18]) with dependencies and other team-specific operators. Handling the causal aspects of these languages requires a richer semantics than propositional team semantics; nonetheless, in this paper we show that the causal-observational languages considered in [2] can be embedded into first-order dependence logic by means of a translation and a careful choice of models. We show that, in some significant cases, the translation can be refined to an embedding into the Bernays-Schönfinkel-Ramsey fragment of dependence logic or, in the restricted case of recursive causal models, into the existential fragment. As an application, we use the embeddings to show the decidability of a satisfiability problem for the causal-observational languages. Along the way, we question the correctness of the semantics for interventionist counterfactuals proposed by Halpern ([18]) and propose an alternative one which behaves as usual in the uncontroversial recursive case.

1. Introduction

The main objective of empirical studies is, arguably, to confirm or falsify our hypotheses on causal relationships among variables. The traditional statistical methods do not directly address causation; they only analyze *associations* among variables as they appear in the recorded data. The methods of causal inference ([28,29]) explain how the empirical information on associations can be combined with causal assumptions in order to prove or reject further causal relationships. In this context, one reasons about

☆ The research of the first author was supported by the Academy of Finland grants n.286991 and n.316460, and by the Deutsche Forschungsgemeinschaft grant VI 1045/1-1. The research of the second author was supported by the Free University of Bozen-Bolzano.

* Corresponding author.
E-mail addresses: f.barbero@sheffield.ac.uk (F. Barbero), pietro.galliani@unibz.it (P. Galliani).

https://doi.org/10.1016/j.apal.2022.103159

the interaction of dependencies of the associational type (which we may call *contingent* dependencies) and causal dependencies. It is therefore natural to search for semantic frameworks and logical languages that may account for both kinds of dependencies; let us see what has been done in this direction.

The standard logical approach to causal reasoning ([11,18,8,35,19,17]) makes use of *causal models* – more specifically, those known as (deterministic) *structural equation models*, which consist of a valuation together with a set of functions describing causal laws. The analysis of causal dependence is reduced to the study of so-called *interventionist counterfactuals* of the form $\mathbf{X} = \mathbf{x} \boxright \psi$, whose intended interpretation is:

if we were to fix the (tuple of) variables $\mathbf{X}$ to the (tuple of) values $\mathbf{x}$, then ψ would hold.

This apparatus is usually modified so that it can handle probabilistic reasoning, and it is most often in such form that it is used towards applications (as e.g. in [28,29]). However, neither the usual deterministic approach nor the probabilistic one seem to account for those kinds of contingent dependencies that are qualitative (i.e. non-probabilistic) in nature. [1,2] then proposed to modify the semantics of causal models using ideas from *team semantics* ([23,31,14,12]). By replacing the single valuation with a set of valuations (*team*) and by modifying appropriately the satisfaction clauses, the resulting *causal team semantics* allows the correct modelling of non-probabilistic contingent dependencies together with counterfactuals and causal dependencies.[1 2] A typical example of the contingent dependencies that can be studied with the help of teams is *functional determinacy*, usually symbolized by *dependence atoms* $=(\mathbf{X}; Y)$, whose intended meaning is:

whenever two valuations agree on the value of $\mathbf{X}$, they also agree on the value of Y.

Among the causal-observational languages proposed in [1,2] as test cases for causal team semantics is a language $\mathcal{COD}$ that can be seen as a generalization of *propositional dependence logic*. Aside from superficial differences, the language $\mathcal{COD}$ enriches propositional dependence logic with dynamic operators corresponding to observation (*selective implication*) and intervention (interventionist counterfactuals). The former is definable in propositional dependence logic, while the latter is not.[3] This is just one instance of the motto that causation is not reducible to association.

In the present paper, the causal-observational languages are compared with *first-order* dependence logic. The latter is a language that incorporates dependence atoms into the syntax of first-order logic. Since causal reasoning is beyond the resources of propositional dependence logic, it may come as a surprise that – as will be shown in the present paper – causal reasoning can be modelled in first-order dependence logic (and, to some extent, already in first-order logic). The key difference between first-order dependence logic and its propositional counterpart is that the former does have some dynamic operators: the quantifiers. The existential and universal quantifiers can both be seen as operators that modify the shape of the team on which a formula is evaluated. In appropriate circumstances, sequences of quantifiers can then be used to "simulate" the effect of the intervention operators.[4] To realize these circumstances, one needs to associate to each causal team T a first-order structure that appropriately encodes the causal laws of T. Once this correspondence is set up, we can define a truth-preserving translation of the causal languages into a fragment

[1] A way of reintegrating probabilistic reasoning in a variant of team semantics is suggested in [1] along the lines of earlier work such as [9]; a fuller development is under preparation. However interesting this line of investigation may be having in mind the applications, in this paper we focus on the simpler deterministic case, which is already of great philosophical interest (see e.g. [17,33,22]).

[2] Causal team semantics has been later generalized so that it can also model uncertainty about the causal laws ([5]). An alternative approach using modal semantics instead of team semantics was given in [3,4]. We will not discuss these alternative approaches in this paper.

[3] An easy formal proof is given in the forthcoming journal version of [5] ([6]).

[4] A reviewer of this paper pointed out the existence of a small literature on quantified *propositional* logics of dependence ([21],[20]). This raises the question (which we leave as an open problem) whether the results presented in this paper might be strengthened in the form of embeddings into quantified propositional logics.

of dependence logic. If in the causal languages we impose some restriction on the use of $\boxright$ within antecedents of selective implications, the formulas in the corresponding target language can be shown to be equivalent to Bernays-Schönfinkel-Ramsey formulas, i.e. formulas with an $\exists^*\forall^*$ prefix and relational vocabulary. The same can be achieved also without syntactical restrictions, provide the semantics is limited to the case of the so-called *unique-solution* causal teams. Furthermore, if one restricts attention to systems of acyclic causal dependencies (*recursive* causal teams), it is possible to give a simpler translation which leads (up to equivalence) into the *existential* fragment of first-order dependence logic. On one side, these translations reveal that first-order dependence logic can be used to model the discourse on causal dependencies. In the opposite direction, they allow us to use known results on dependence logic to shed new light on causal team semantics. To illustrate this point, we will use some results on the satisfiability problem of fragments of first-order (dependence) logic to address some satisfiability issues for the causal-observational languages.

The paper is structured as follows. In section 2 we review (and illustrate with a few examples) the preliminary definitions of causal team semantics and of the causal-observational languages. Contrarily to much of the previous literature on causal teams, we do not restrict attention to recursive models (i.e. acyclic causal laws). In particular, we adopt a very general definition of intervention that was sketched in [1,2] and is here fleshed out for the first time. The definition generalizes an earlier idea of Halpern ([18]) for causal models, and thus we denote the semantics that it induces by $\models^H$. In section 3 we show how to embed the causal-observational languages (interpreted according to $\models^H$) into the Bernays-Schönfinkel-Ramsey fragment of dependence logic; the section opens with the necessary preliminaries on first-order logic, dependence logic and their team semantics. Section 4 presents two alternative definitions of intervention on a causal team, and the alternative semantics they induce. Of these, the former, denoted as $\models^A$, is based on a fully general definition of intervention that we argue to agree better with the usual intuitions about causation than the semantics $\models^H$ does; the latter, denoted as $\models^R$, is the most common definition of intervention from the previous literature on causal teams, which applies only to the recursive case. We present embedding results for both semantics, and we show that $\models^H, \models^A$ and $\models^R$ agree in the recursive case. In section 5 we show that (in the recursive case) the satisfiability problem of the causal languages for a fixed, finite variable domain (but arbitrary ranges) is decidable. To this purpose, we introduce a weaker but more general form of the embedding. In section 6 we draw conclusions and suggest directions for further research.

2. Causal teams and causal languages

2.1. Preliminaries and notation

We will mostly follow the notational conventions from the literature on interventionist counterfactuals; therefore, we use capital letters X, Y, Z... for variables, and small letters x, y, z... for constants (called *values*). Boldface letters $\mathbf{X}$, resp. $\mathbf{x}$, denote (depending on the context) finite sets or sequences of variables, resp. of values. We will specify case by case whether we refer to sets or sequences.

Formal definitions and results in the field of causal inference often need to be formulated in relation to a *signature*, which describes which variables are taken into consideration and over what sets their values are allowed to vary. More precisely, a **signature** σ is a pair (Dom, Ran), where Dom is a nonempty set of variables and Ran is a function that associates to each variable $X \in Dom$ a nonempty set $Ran(X)$ of values (the *range* of X). We will usually assume that $Ran(X)$ has at least two elements, in order to exclude trivial cases. When we use $\mathbf{X}$ to denote a sequence of variables $(X_1, \dots, X_n)$, typically $\mathbf{x}$ will stand for a tuple $(x_1, \dots, x_n)$ of the same length, and such that, for all $i = 1 \dots n$, $x_i \in Ran(X_i)$. In other words, we assume that x_i is a value for X_i. Note also that we write $Ran(\mathbf{X}) := Ran(X_1) \times \dots \times Ran(X_n)$ when $\mathbf{X} = (X_1, \dots, X_n)$.

An **assignment** of signature σ is a mapping $s : Dom \to \bigcup_{X \in Dom} Ran(X)$ such that $s(X) \in Ran(X)$ for each $X \in Dom$. A **team** T of signature σ is a set of such assignments. As a general convention, we will often write $s(\mathbf{X}) = \mathbf{x}$ as an abbreviation for many statements of the form $s(X_1) = x_1, \ldots, s(X_n) = x_n$; notice that we are assuming here that $\mathbf{X}$ and $\mathbf{x}$ are tuples of equal length, that the variables in $\mathbf{X}$ are distinct, and that there is an implicit correspondence between the X_i and the x_i.

We will make use of (directed) graphs. A **graph**, in this note, will be a pair $G = (\mathbf{V}, E)$, where $\mathbf{V}$ is a set of variables and $E \subseteq \mathbf{V} \times \mathbf{V}$. The elements of E will be called **arrows** or **edges**. A **subgraph** of G is a graph $H = (\mathbf{V}', E')$ such that $\mathbf{V}' \subseteq \mathbf{V}$ and $E' \subseteq E$; if, in particular, $\mathbf{V}' = \mathbf{V}$, then H is said to be a **spanning subgraph** of G. Given a graph G and one of its vertices V, we denote as PA_V^G the set of **parents** of V in G (i.e. the set of variables X such that $(X, V) \in E$). We omit the superscript when the graph is clear from the context. If the set of parents of V is empty, V is called an **exogenous** variable; the set of exogenous variables of G is denoted as $Exo(G)$. The remaining variables of G are called **endogenous** and their set is denoted as $End(G)$.

A **(directed) path** of a graph G is a finite sequence $(X_1, Y_1), \ldots, (X_n, Y_n)$ of edges such that $Y_i = X_{i+1}$ for $i = 1, \ldots, n-1$. The index n is called the **length** of the path; if the path is the empty sequence, we say its length is 0. Given a set of nodes $\mathbf{X}$, a node Y is a **strict descendant** of $\mathbf{X}$ if there is an $X \in \mathbf{X}$ and a path $(X, X_1), \ldots, (X_n, Y)$ of length > 0; Y is a **descendant** of $\mathbf{X}$ if either Y is a strict descendant of $\mathbf{X}$ or $Y \in \mathbf{X}$. Y is a **nondescendant** of $\mathbf{X}$ if it is not a descendant of $\mathbf{X}$.

A path $(X_1, X_2), \ldots, (X_{n-1}, X_n)$ where $X_1 = X_n$ while all other variables are distinct is called a **cycle**. A graph without cycles is called **acyclic**, and a graph without cycles of length 1 is said to be **irreflexive**.

2.2. *Causal teams*

A causal team enriches a team by isolating a set of functions which describe the causal mechanisms that link the variables. More precisely, some of the variables (endogenous variables) are associated to functions that describe their behaviour in terms of the other variables; the remaining, exogenous variables are left causally unexplained. A graph is used to keep track of the exogenous/endogenous distinction, and of the domains of the functions.[5]

Definition 2.1 *(Causal team).* A **causal team** T of signature $\sigma = (Dom, Ran)$ with endogenous variables $End(T) \subseteq Dom$ is a triple $T = (T^-, G_T, \mathcal{F}_T)$, where:

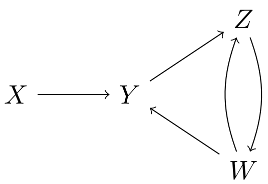
1. T^- is a team of signature σ (*team component* of T).
2. $G_T = (Dom, E)$ is an irreflexive graph over Dom (*graph component* of T) such that $Y \in End(G_T) \iff Y \in End(T)$.
3. $\mathcal{F}_T$ is a function $\{(V_i, f_{V_i}) \mid V_i \in End(T)\}$ (*function component* of T) that assigns to each endogenous variable V a function $f_V : Ran(PA_V) \to Ran(V)$ which satisfies the compatibility constraint:

   (*) For all $s \in T^-$, $s(V) = f_V(s(PA_V))$.

The constraint imposed on the graph G_T makes so that $End(T) = End(G_T)$. We define the set of exogenous variables of T as $Exo(T) = Exo(G_T)$. We also remark that the definition places no upper bound on the cardinalities of Dom, Ran and of the ranges of individual variables, but in some of the later sections we will require that these sets are finite.

[5] The graph component is not strictly necessary for an adequate definition of causal teams, at least at the level of generality that we consider here. See [5] and [3] for two alternative presentations that do not explicitly include a graph in the definition.

Example 2.2. Consider a signature $\sigma = (Dom, Ran)$ with $Dom = \{X, Y, Z, W\}$ and $Ran(V) = \{-1, 0, 1\}$ for each $V \in Dom$. The picture below describes the graph component, function component and team component of a causal team $T = (T^-, G, \mathcal{F})$ with two assignments.



$$\begin{cases} Y := X + W \\ Z := Y - W \\ W := -Z \end{cases}$$

X	Y	Z	W
0	0	0	0
1	0	1	-1

Expressions of the form $Y := X + W$ are called, in the literature on causal inference, *structural equations*. They are just a conventional way of expressing causal laws; in this case, the "equation" describes the function $f_Y = \mathcal{F}(Y)$ which "produces" values for Y in terms of the values of $PA_Y = \{X, W\}$. Or, equivalently, this equation abbreviates a number of counterfactuals of the form "If X, W were fixed to values x, w, then Y would take value $x + w$". The three variables occurring in the left members of the equations are the endogenous variables, while X is the only exogenous one. The team component is represented as a table whose rows are the two assignments; for example, the second row describes an assignment s with $s(X) = 1$, $s(Y) = 0$, $s(Z) = 1$ and $s(W) = -1$. Notice that each of these assignments describes a solution of the system of "equations". The distinction between exogenous and endogenous variables can also be read off from the graph: X is an exogenous variable (the only one) because no arrow ends in X. Notice that the variables Z, W form a cycle of length 2, while Z, W, Y form a cycle of length 3.

We imposed the restriction that G_T be irreflexive; this excludes cases of self-causation (cycles of length 1). This requirement does not, however, exclude cycles in general, as the previous example illustrates. A causal team is said to be **recursive** if its graph is acyclic. The recursive causal teams generalize the recursive causal models from the literature on causal inference. It has sometimes been argued (see [30]) that only the recursive models have a causal interpretation; however, also the nonrecursive models have been studied, especially in relation to interventionist counterfactuals (see e.g. [18,35,19]). The present paper fits in this latter tradition.

2.3. Causal-observational languages

We will consider the following languages (parametrized by a signature $\sigma = (Dom, Ran)$), as they were introduced in [1]:

- $\mathcal{CO}(\sigma) ::= Y = y \mid Y \neq y \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \supset \alpha \mid \mathbf{X} = \mathbf{x} \boxright \alpha$
- $\mathcal{COD}(\sigma) ::= Y = y \mid Y \neq y \mid\ =\!(\mathbf{X}; Y) \mid \psi \wedge \psi \mid \psi \vee \psi \mid \alpha \supset \psi \mid \mathbf{X} = \mathbf{x} \boxright \psi$
- $\mathcal{CO}_{\sqcup}(\sigma) ::= Y = y \mid Y \neq y \mid \psi \wedge \psi \mid \psi \vee \psi \mid \psi \sqcup \psi \mid \alpha \supset \psi \mid \mathbf{X} = \mathbf{x} \boxright \psi$

where $\mathbf{X} \cup \{Y\} \subseteq Dom$, $y \in Ran(Y)$, $\mathbf{x} \in Ran(\mathbf{X})$, $\alpha \in \mathcal{CO}(\sigma)$. $\mathbf{X} = \mathbf{x}$ is an abbreviation for a (finite) conjunction $X_1 = x_1 \wedge \dots \wedge X_n = x_n$.

The language $\mathcal{CO}(\sigma)$ is similar to those used in causal inference; but notice that its formulas are in negation normal form. In principle, we could allow a (dual) negation to occur in front of any $\mathcal{CO}$ formula, but there seems to be no reasonable way to do the same for the larger languages. We will say more about this in section 2.6.

We point out that *propositional dependence logic* ([34]) can be seen as a special case of $\mathcal{COD}(\sigma)$. A language for propositional dependence logic with propositional letters $p_1, \dots, p_n, \dots$ corresponds to a $\mathcal{COD}(\sigma)$

language such that Dom contains some variables $P_1, \dots, P_n, \dots$ with $Ran(P_i) = \{0, 1\}$. Then, the formula p_i can be identified with $P_i = 1$, $\neg p_i$ with $P_i = 0$, and so on.

2.4. Halpern-style interventions

The *interventionist counterfactual* operator $\Box\!\!\rightarrow$ will be given a meaning in terms of *interventions on a causal team*. In this section, we will describe in detail what it means to apply an intervention to a causal team, along the lines suggested in [2]. The definition we give here applies to all causal teams; we will see later how this definition specializes in the case of recursive causal teams. The general idea is that by applying an intervention $do(\mathbf{X} = \mathbf{x})$ to a causal team $T = (T^-, G, \mathcal{F})$ one obtains a new causal team $T_{\mathbf{X}=\mathbf{x}}$. The intervention describes what would happen if we subtracted the variables $\mathbf{X}$ from their current causal mechanisms (i.e. the corresponding functions from $\mathcal{F}$) and forced them to take the constant values $\mathbf{x}$. If the initial team T^- includes all the configurations (of values for the variable in the domain) that are considered possible, the team $(T_{\mathbf{X}=\mathbf{x}})^-$ describes what configurations should be considered possible after the intervention.

Now, as mentioned above, there is no general agreement on whether nonrecursive structural equation models can be given a causal interpretation. A popular interpretation was given by Strotz&Wold ([30]): the solutions to such a system of equations represent all the possible equilibria of the corresponding system of variables. An intervention is not necessarily a deterministic act: we may not know, a priori, which of the possible equilibria will be obtained after the intervention. Along the lines of this interpretation, Halpern ([18]) describes the result of an intervention $do(\mathbf{X} = \mathbf{x})$ $(= do(X_1 = x_1, \dots, X_n = x_n))$ on a causal model as being the set of all vectors of values for the endogenous variables – variables in $\mathbf{X}$ excluded – that satisfy the reduced system of equations. Since the languages considered by Halpern allow no atomic statements concerning the exogenous variables, Halpern is not including the values of the exogenous variables – nor the values of $\mathbf{X}$, which may be thought of as exogenous for the reduced system – in this “solution vector”. It seems natural to adapt Halpern's idea to the context of causal teams, by applying it to each single assignment in the team T^- at hand, and finally taking the union of the sets of solutions obtained from each assignment; but since our assignments cover also the exogenous variables, we must decide what to do about these (and about the variables in $\mathbf{X}$). The most natural option seems to be 1) to keep only those solutions that satisfy $\mathbf{X} = \mathbf{x}$; and, calling $\mathbf{U}$ the set of exogenous variables that are not in $\mathbf{X}$, 2) only admit solutions whose values for $\mathbf{U}$ were already in T^-.

A conjunction $\mathbf{X} = \mathbf{x}$ is said to be **inconsistent** if it contains two conjuncts $X = x$ and $X = x'$ with distinct values x, x'; otherwise it is said to be **consistent**; the intervention $do(\mathbf{X} = \mathbf{x})$ will be defined only in case $\mathbf{X} = \mathbf{x}$ is consistent. Let $\sigma = (Dom, Ran)$ be a signature and $T = (T^-, G, \mathcal{F})$ be a causal team of signature σ. Let $\mathbf{X} \subseteq Dom$ be a tuple of distinct variables and $\mathbf{x} \in Ran(\mathbf{X})$. Write $\mathbf{U}$ for $Exo(T) \setminus \mathbf{X}$. Write $T^-(\mathbf{U}) := \{s(\mathbf{U}) \mid s \in T^-\}$. We will say that an assignment s of signature σ is **compatible** with $(G, \mathcal{F})$ if

$$\text{for all } Y \in End(G),\ s(Y) = \mathcal{F}(Y)(s(PA_Y))$$

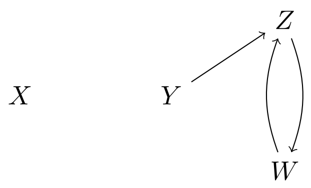
where PA_Y are the parents of Y as encoded in G.

Definition 2.3 *(Intervention, Halpern-style).* Let $\mathbf{X} = \mathbf{x}$ be a consistent conjunction over a signature σ. The result of the intervention $do(\mathbf{X} = \mathbf{x})$ on a causal team $T = (T^-, G, \mathcal{F})$ of signature σ is the causal team $T_{\mathbf{X}=\mathbf{x}} = ((T_{\mathbf{X}=\mathbf{x}})^-, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})$, where:

- $(T_{\mathbf{X}=\mathbf{x}})^- := \{s \text{ compatible with } (G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \mid s(\mathbf{X}) = \mathbf{x} \text{ and } s(\mathbf{U}) \in T^-(\mathbf{U})\}$
- $(Y, Z) \in G_{\mathbf{X}=\mathbf{x}}$ iff $(Y, Z) \in G$ and $Z \notin \mathbf{X}$,
- $\mathcal{F}_{\mathbf{X}=\mathbf{x}}$ is the restriction of $\mathcal{F}$ to $End(T) \setminus \mathbf{X}$.

We point out that the set of endogenous variables for the resulting team $T_{\mathbf{X}=\mathbf{x}}$ is $End(T_{\mathbf{X}=\mathbf{x}}) = End(T) \setminus \mathbf{X}$. That is, we treat the intervened variables as exogenous in the intervened team. In this we slightly differ from the original approach of Halpern: our approach amounts to removing the equations for the (endogenous) variables in $\mathbf{X}$, while Halpern replaces them with equations of the form $X_i = x_i$. These two points of view are equivalent with regards to the languages considered here.

Example 2.4. Consider again the causal team T from Example 2.2. If we apply the intervention $do(Y = 0)$ to it, we obtain the causal team $T_{Y=0}$ described in the picture:



$$\begin{cases} Z := Y - W \\ W := -Z \end{cases}$$

X	Y	Z	W
0	0	0	0
0	0	1	-1
0	0	-1	1
1	0	0	0
1	0	1	-1
1	0	-1	1

In the graph, we have removed all the arrows that pointed to the intervened variable Y; Y has become an exogenous variable in the intervened team. Correspondingly, the function $\mathcal{F}_T(Y)$ describing the causal law for Y has been removed, thus producing a reduced system of equations. The set $\mathbf{U}$ for this intervention is $\{X\}$; each of the values of X appearing in T (0 and 1) has given rise to three solutions of the system compatible with the constraint $Y = 0$. Notice that, when $Y = 0$, the system of equations just says that the value of Z is the opposite of the value of W.

For a comparison, after the intervention $do(Y = 1)$ we obtain the same graph and function component; however, the (reduced) system of equations clearly has no solutions with $Y = 1$, and therefore $(T_{Y=1})^- = \emptyset$.

This definition of intervention can be also thought of as applying an intervention (in the sense of [18]) to each assignment in the team, and then collecting together the outcomes produced by each assignment. In order to express this point precisely, it will be convenient to introduce some auxiliary notations. We say that $S = (S^-, G_S, \mathcal{F}_S)$ is a **causal subteam** of $T = (T^-, G_T, \mathcal{F}_T)$, $S \leq T$, if $S^- \subseteq T^-$, $G_S = G_T$ and $\mathcal{F}_S = \mathcal{F}_T$. We use the improper notation $\{s\}$ for the causal subteam of T of team component $\{s\}$ whenever T is clear from the context.

Proposition 2.5. *Let $T = (T^-, G, \mathcal{F})$ be a causal team of signature σ and $\mathbf{X} = \mathbf{x}$ be a consistent conjunction of signature σ. Then:*

$$T^-_{\mathbf{X}=\mathbf{x}} = \bigcup_{s \in T^-} \{s\}^-_{\mathbf{X}=\mathbf{x}}.$$

Proof. Write $\mathbf{U}$ for $Exo(T) \setminus \mathbf{X}$.

Let $s \in T^-$ and $t \in \{s\}^-_{\mathbf{X}=\mathbf{x}}$. By definition of $\{s\}_{\mathbf{X}=\mathbf{x}}$, we have that t is compatible with $(G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})$, $t(\mathbf{X}) = \mathbf{x}$ and $t(\mathbf{U}) \in \{s\}(\mathbf{U}) \subseteq T^-(\mathbf{U})$. Thus $t \in T^-_{\mathbf{X}=\mathbf{x}}$. Therefore $\{s\}^-_{\mathbf{X}=\mathbf{x}} \subseteq T^-_{\mathbf{X}=\mathbf{x}}$.

Vice versa, let $t \in T^-_{\mathbf{X}=\mathbf{x}}$. By definition of $T^-_{\mathbf{X}=\mathbf{x}}$, $t(\mathbf{U}) \in T^-(\mathbf{U})$. This means there is an $s \in T^-$ such that $t(\mathbf{U}) = s(\mathbf{U})$. Furthermore, the condition $t \in T^-_{\mathbf{X}=\mathbf{x}}$ gives us that $t(\mathbf{X}) = \mathbf{x}$ and that t is compatible with $(G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})$. Therefore $t \in \{s\}^-_{\mathbf{X}=\mathbf{x}}$ for some $s \in T^-$. □

Thus, the definition of intervention considered here generalizes the one proposed by Halpern in [18]. Its adequacy therefore depends on the adequacy of Halpern's definition, in the sense that it would inherit any of its potential issues. We will discuss in a later section whether the adequacy of Halpern's definition is uncontroversial.

We conclude this section by introducing one more special class of causal teams.[6] Let $\sigma = (Dom, Ran)$ be a signature and let $T = (T^-, G, \mathcal{F})$ be a causal team of signature σ with exogenous variables $\mathbf{U}$. T is **unique-solution** (for signature σ) if the following hold:

1. For every tuple $\mathbf{u} \in Ran(\mathbf{U})$, there is exactly one assignment s of signature σ that is compatible with $(G, \mathcal{F})$ and such that $s(\mathbf{U}) = \mathbf{u}$.
2. Similarly, for each causal team $T_{\mathbf{X}=\mathbf{x}}$ ($\mathbf{X} \subseteq Dom$, $\mathbf{x} \in Ran(\mathbf{X})$) and every tuple $\mathbf{z} \in Ran(\mathbf{Z}) = Ran(Exo(T_{\mathbf{X}=\mathbf{x}}))$, there is exactly one assignment s of signature σ that is compatible with $(G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})$ and such that $s(\mathbf{XZ}) = \mathbf{xz}$.

Notice that this definition does not depend on the team component of T. It is also straightforward to prove that every *recursive* causal team is unique-solution, but not vice versa.

2.5. Semantic clauses

We are now ready to define our main semantics for the causal-observational languages. We will define what it means for a formula φ to be satisfied by a causal team T. This relation will be denoted as $T \models \varphi$, or more specifically as $T \models^H \varphi$ if we need to compare it to different semantics.

Satisfaction of a formula by a team, $T \models \varphi$, is defined inductively as:

- $T \models Y = y$ iff, for all $s \in T^-$, $s(Y) = y$.
- $T \models Y \neq y$ iff, for all $s \in T^-$, $s(Y) \neq y$.
- $T \models =\!(\mathbf{X}; Y)$ iff for all $s, s' \in T^-$, $s(\mathbf{X}) = s'(\mathbf{X})$ implies $s(Y) = s'(Y)$.
- $T \models \psi \wedge \chi$ iff $T \models \psi$ and $T \models \chi$.
- $T \models \psi \vee \chi$ iff there are $T_1, T_2 \leq T$ s.t. $T_1^- \cup T_2^- = T^-$, $T_1 \models \psi$ and $T_2 \models \chi$.
- $T \models \psi \sqcup \chi$ iff $T \models \psi$ or $T \models \chi$.
- $T \models \mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \chi$ iff $\mathbf{X} = \mathbf{x}$ is inconsistent or $T_{\mathbf{X}=\mathbf{x}} \models \chi$.
- $T \models \alpha \supset \chi$ iff $T^\alpha \models \chi$, where T^α is the (unique) causal subteam of T with team component $\{s \in T^- \mid \{s\} \models \alpha\}$.

The operators $\vee$ and $\sqcup$ are known, respectively, as *tensor* and *global* disjunction. The *selective implication* $\alpha \supset \chi$ is interpreted similarly as a public announcement operator ([32]); it says that χ holds if we observe/learn that α holds.

This satisfaction relation induces, as usual, notions of semantic entailment and validity. If Γ is a set of formulas, φ a formula, and σ a signature, we write $\Gamma \models_\sigma \varphi$ if, for every causal team T of signature σ, $T \models \Gamma$ implies that $T \models \varphi$. If $\Gamma = \{\theta\}$ we write more simply $\theta \models_\sigma \varphi$. If φ, θ are formulas, we write $\varphi \equiv_\sigma \theta$ if $\theta \models_\sigma \varphi$ and $\varphi \models_\sigma \theta$. We will often omit the subscript σ if the signature is clear from the context.

Our formal languages, interpreted according to the semantics just described, have the following important properties, that we will use throughout the paper.

Proposition 2.6. *Let S, T be causal teams of signature σ.*

- ***Empty team property:*** *if $\varphi \in \mathcal{COD}(\sigma) \cup \mathcal{CO}_\sqcup(\sigma)$ and $T = (\emptyset, G, \mathcal{F})$, then $T \models \varphi$.*

[6] This definition and that of a recursive causal team are straightforward adaptations from [11]. They are slight generalizations, in that we had to decide what to do with interventions on exogenous variables, which are not allowed in the framework of [11].

- ***Flatness:*** *if $\varphi \in \mathcal{CO}(\sigma)$, then $T \models \varphi$ iff, for all $s \in T^-$, $\{s\} \models \varphi$.*
- ***Downward closure:*** *if $\varphi \in \mathcal{COD}(\sigma) \cup \mathcal{CO}_{\sqcup}(\sigma)$, $T \models \varphi$ and $S \leq T$, then $S \models \varphi$.*

Proof. These properties were proved in [2] for the recursive case, by induction on φ. In the argument for flatness, the inductive case differs from [2] when φ is of the form $\mathbf{X} = \mathbf{x} \Box\!\!\rightarrow \psi$, so we give the details.

Suppose first that $T \models \mathbf{X} = \mathbf{x} \Box\!\!\rightarrow \psi$. Then $T_{\mathbf{X}=\mathbf{x}} \models \psi$. By the inductive hypothesis, for every $t \in (T_{\mathbf{X}=\mathbf{x}})^-$, $\{t\} \models \psi$. Now observe that, by Proposition 2.5, for every $s \in T^-$, each element of $(\{s\}_{\mathbf{X}=\mathbf{x}})^-$ is in $(T_{\mathbf{X}=\mathbf{x}})^-$. Thus, by the inductive hypothesis again, for each $s \in T^-$ we have that $\{s\}_{\mathbf{X}=\mathbf{x}} \models \psi$. Thus $\{s\} \models \mathbf{X} = \mathbf{x} \Box\!\!\rightarrow \psi$.

From right to left, assume that, for all $s \in T^-$, $\{s\} \models \mathbf{X} = \mathbf{x} \Box\!\!\rightarrow \psi$. Then $\{s\}_{\mathbf{X}=\mathbf{x}} \models \psi$ for all such s. Then, by the inductive hypothesis, for every $t \in (\{s\}_{\mathbf{X}=\mathbf{x}})^-$, $\{t\} \models \psi$. Since $(T_{\mathbf{X}=\mathbf{x}})^- = \bigcup_{s \in T^-}(\{s\}_{\mathbf{X}=\mathbf{x}})^-$, applying the inductive hypothesis again we obtain $T_{\mathbf{X}=\mathbf{x}} \models \psi$, and thus $T \models \mathbf{X} = \mathbf{x} \Box\!\!\rightarrow \psi$. □

We also remark that each of the operators $\wedge, \vee, \supset$ behaves like the corresponding classical operator on causal teams with singleton team component.

2.6. Negation

The languages $\mathcal{CO}(\sigma)$ can be extended with a (dual) negation that behaves as classical negation on causal teams with singleton team component. The truth clause is:

- $T \models \neg\varphi$ iff, for all $s \in T^-$, $\{s\} \not\models \varphi$.

This operator is already definable in $\mathcal{CO}(\sigma)$. Taking $\perp$ to stand for any contradictory formula (say, $X = x \wedge X \neq x$, for some $X \in Dom$ and $x \in Ran(X)$), we have $\neg\psi \equiv \psi \supset \perp$. Less straightforwardly, in the recursive case the dual negation gives formulas that are equivalent to those that are produced by the following inductive clauses: $(X = x)^d := X \neq x$, $(X \neq x)^d := X = x$, $(\psi \wedge \chi)^d := \psi^d \vee \chi^d$, $(\psi \vee \chi)^d := (\psi^d \wedge \chi^d)$, $(\psi \supset \chi)^d := \psi \wedge \chi^d$, $(\psi \Box\!\!\rightarrow \chi)^d := \psi \Box\!\!\rightarrow \chi^d$.

Lemma 2.7. *Let T be a recursive causal team and $\varphi \in \mathcal{CO}$. Then $T \models \varphi^d$ iff $T \models \neg\varphi$.*

Proof. The proof can be found in [2] (theorem 2.11). □

There seems to be no reasonable extension of this result to $\mathcal{COD}(\sigma)$ or $\mathcal{CO}_{\sqcup}(\sigma)$; see e.g. [25]. Furthermore, the result is false for non-recursive causal teams; in particular, the clause for $(\psi \Box\!\!\rightarrow \chi)^d$ is incorrect, as the following example shows.[7]

Example 2.8. Let $Dom = \{X, Y, Z\}$ and $Ran(X) = Ran(Y) = Ran(Z) = \{0, 1\}$. Let s be the assignment given by $s(X) = s(Y) = s(Z) = 0$. We then consider the causal team T having s as the only assignment, and graph and function components given by the equations $X := Y$ and $Y := X$. Now, as the picture shows,

T:

X↔Y		Z
0	0	0

$\rightsquigarrow$ $T_{Z=0}$:

X↔Y		Z
0	0	0
1	1	0

$T \models \neg(Z = 0 \Box\!\!\rightarrow X = 0)$ (as $\{s\} \not\models Z = 0 \Box\!\!\rightarrow X = 0$), but $T \not\models (Z = 0 \Box\!\!\rightarrow X = 0)^d = Z = 0 \Box\!\!\rightarrow X \neq 0$, since X takes the value 1 in one of the assignments of $T_{Z=0}$.

---

[7] A similar example was given in [18], example 2.3. We thank a reviewer of this paper for pointing out the existence of this counterexample (and others which will appear in the following).

2.7. Unnested form

We end this section with some remarks about the nesting of counterfactuals. The syntax of our languages allows for nesting the counterfactuals *on the right*, as e.g. in $\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow (\mathbf{Y} = \mathbf{y} \,\square\!\!\rightarrow \psi)$ or $\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow (\psi \vee ((\mathbf{Y} = \mathbf{y} \,\square\!\!\rightarrow \chi) \wedge (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \theta)))$. However, when the semantics is restricted to the case of unique-solution causal teams, we can prove that all formulas are equivalent to *un*nested ones.[8] Towards proving this fact, we point out some formula equivalences (distribution rules) that hold in general (and which are easily provable along the lines of [18] and [2]):

$$\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow (\psi \wedge \chi) \equiv (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \psi) \wedge (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \chi)$$
$$\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow (\psi \sqcup \chi) \equiv (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \psi) \sqcup (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \chi)$$

Unfortunately, analogous distribution rules do not hold in general for the operators $\vee$ and $\supset$, as the following example shows.

Example 2.9. Consider again the causal team T from Example 2.8. We have $T \models Z = 0 \,\square\!\!\rightarrow (X = 0 \vee X = 1)$ but $T \not\models (Z = 0 \,\square\!\!\rightarrow X = 0) \vee (Z = 0 \,\square\!\!\rightarrow X = 1)$.

Observe also that, trivially, $T \models (Z = 0 \,\square\!\!\rightarrow X = 0) \supset (Z = 0 \,\square\!\!\rightarrow X = 1)$ (since $T_{Z=0 \square\!\rightarrow X=0}$ is empty). But $T \not\models Z = 0 \,\square\!\!\rightarrow (X = 0 \supset X = 1)$.

The distribution rules for $\vee$ and $\supset$ do hold if the causal team is unique-solution (thus in particular if it is recursive, i.e. the causal graph is acyclic).

Lemma 2.10. *Let ψ, χ be $\mathcal{COD}(\sigma)$ or $\mathcal{CO}_{\sqcup}(\sigma)$ formulas, and α a $\mathcal{CO}(\sigma)$ formula. For all T unique-solution causal teams of signature σ, we have:*

1. $T \models \mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow (\psi \vee \chi) \iff T \models (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \psi) \vee (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \chi)$
2. $T \models \mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow (\alpha \supset \chi) \iff T \models (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \alpha) \supset (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \chi)$.

Without the restriction on the unicity of solutions, we still have:

1. $(\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \psi) \vee (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \chi) \models \mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow (\psi \vee \chi)$
2. $\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow (\alpha \supset \chi) \models (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \alpha) \supset (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \chi)$.

Proof. Let us begin with the statements that hold in general.

3) Suppose $T \models (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \psi) \vee (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \chi)$. Then there are two causal subteams T_1, T_2 of T such that $T_1^- \cup T_2^- = T^-$, $(T_1)_{\mathbf{X}=\mathbf{x}} \models \psi$ and $(T_2)_{\mathbf{X}=\mathbf{x}} \models \chi$. Since by Proposition 2.5 $(T_1)^-_{\mathbf{X}=\mathbf{x}} \cup (T_2)^-_{\mathbf{X}=\mathbf{x}} = T^-_{\mathbf{X}=\mathbf{x}}$, we have $T_{\mathbf{X}=\mathbf{x}} \models \psi \vee \chi$, from which the thesis follows.

4) Let T be a causal team of signature σ. We first prove that $(T^{\mathbf{X}=\mathbf{x} \square\!\rightarrow \alpha})_{\mathbf{X}=\mathbf{x}}$ is a causal subteam of $(T_{\mathbf{X}=\mathbf{x}})^\alpha$; this amounts to proving the inclusion of their respective team components.

Let then $s \in (T^{\mathbf{X}=\mathbf{x} \square\!\rightarrow \alpha})^-_{\mathbf{X}=\mathbf{x}}$. Then by Proposition 2.5 there is a $t \in (T^{\mathbf{X}=\mathbf{x} \square\!\rightarrow \alpha})^-$ such that $s \in \{t\}^-_{\mathbf{X}=\mathbf{x}} \subseteq (T^{\mathbf{X}=\mathbf{x} \square\!\rightarrow \alpha})^-_{\mathbf{X}=\mathbf{x}} \subseteq T^-_{\mathbf{X}=\mathbf{x}}$. Since $t \in (T^{\mathbf{X}=\mathbf{x} \square\!\rightarrow \alpha})^-$, we have $\{t\} \models \mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \alpha$, thus $\{t\}_{\mathbf{X}=\mathbf{x}} \models \alpha$ and, by downward closure, $\{s\} \models \alpha$. Since $s \in T^-_{\mathbf{X}=\mathbf{x}}$, then, we conclude $s \in ((T_{\mathbf{X}=\mathbf{x}})^\alpha)^-$.

Suppose now that $T \models \mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow (\alpha \supset \chi)$. Thus $(T_{\mathbf{X}=\mathbf{x}})^\alpha \models \chi$. By downward closure, $(T^{\mathbf{X}=\mathbf{x} \square\!\rightarrow \alpha})_{\mathbf{X}=\mathbf{x}} \models \chi$. Thus $T \models (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \alpha) \supset (\mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow \chi)$.

1) Assume $T \models \mathbf{X} = \mathbf{x} \,\square\!\!\rightarrow (\psi \vee \chi)$. Then, by the semantic clauses, there are S_1, S_2 causal subteams of $T_{\mathbf{X}=\mathbf{x}}$ such that $S_1^- \cup S_2^- = T^-_{\mathbf{X}=\mathbf{x}}$, $S_1 \models \psi$ and $S_2 \models \chi$. Let $T_i := \{s \in T^- \mid (t \in \{s\}^-_{\mathbf{X}=\mathbf{x}} \Rightarrow t \in S_i^-)\}$.

[8] We leave it as an open problem whether this result can be extended to the general case.

Since T is unique-solution (ensuring that $\{s\}^-_{\mathbf{X}=\mathbf{x}}$ is a singleton for each $s \in T^-$), we have $T_1^- \cup T_2^- = T^-$. Furthermore, it is easy to see that $(T_i)_{\mathbf{X}=\mathbf{x}} = S_i$: indeed, if $t \in (T_i)^-_{\mathbf{X}=\mathbf{x}}$ then by Proposition 2.5 there exists some $s \in T_i^-$ such that $t \in \{s\}^-_{\mathbf{X}=\mathbf{x}}$, but by definition $\{s\}^-_{\mathbf{X}=\mathbf{x}} \subseteq S_i^-$ for all $s \in T_i^-$ and so in particular $t \in \{s\}^-_{\mathbf{X}=\mathbf{x}} \subseteq S_i^-$; and conversely, if $t \in S_i^-$ then $t \in T^-_{\mathbf{X}=\mathbf{x}}$, and so by Proposition 2.5 there exists some $s \in T^-$ such that $t \in \{s\}^-_{\mathbf{X}=\mathbf{x}}$; but then since T is unique-solution we have that $\{s\}^-_{\mathbf{X}=\mathbf{x}} = \{t\} \subseteq S_i^-$ and so $s \in T_i^-$ and finally – again by Proposition 2.5 – $t \in \{s\}^-_{\mathbf{X}=\mathbf{x}} \subseteq (T_i)^-_{\mathbf{X}=\mathbf{x}}$. Thus, $(T_1)_{\mathbf{X}=\mathbf{x}} \models \psi$ and $(T_2)_{\mathbf{X}=\mathbf{x}} \models \chi$, from which the thesis follows by the semantic clauses. For the right-to-left direction, see the proof of 3).

2) We prove that $((T_{\mathbf{X}=\mathbf{x}})^\alpha)^- = (T^{\mathbf{X}=\mathbf{x} \Box\!\!\rightarrow \alpha})^-_{\mathbf{X}=\mathbf{x}}$. The left-to-right inclusion was proved in 4). For the opposite direction, assume $s \in ((T_{\mathbf{X}=\mathbf{x}})^\alpha)^-$. Then $\{s\} \models \alpha$ and by Proposition 2.5 there is a $t \in T^-$ such that $s \in \{t\}^-_{\mathbf{X}=\mathbf{x}}$. Since T is unique-solution, $\{t\}^-_{\mathbf{X}=\mathbf{x}} = \{s\}$. Thus $\{t\}_{\mathbf{X}=\mathbf{x}} \models \alpha$, i.e. $\{t\} \models \mathbf{X} = \mathbf{x} \Box\!\!\rightarrow \alpha$. So, $t \in (T^{\mathbf{X}=\mathbf{x} \Box\!\!\rightarrow \alpha})^-$, and therefore $s \in (T^{\mathbf{X}=\mathbf{x} \Box\!\!\rightarrow \alpha})^-_{\mathbf{X}=\mathbf{x}}$; thus, $((T_{\mathbf{X}=\mathbf{x}})^\alpha)^- \subseteq (T^{\mathbf{X}=\mathbf{x} \Box\!\!\rightarrow \alpha})^-_{\mathbf{X}=\mathbf{x}}$.

Then $T \models (\mathbf{X} = \mathbf{x} \Box\!\!\rightarrow \alpha) \supset (\mathbf{X} = \mathbf{x} \Box\!\!\rightarrow \chi)$ if and only if $(T^{\mathbf{X}=\mathbf{x} \Box\!\!\rightarrow \alpha})_{\mathbf{X}=\mathbf{x}} \models \chi$ if and only if $(T_{\mathbf{X}=\mathbf{x}})^\alpha \models \chi$ if and only if $T \models \mathbf{X} = \mathbf{x} \Box\!\!\rightarrow (\alpha \supset \chi)$. $\square$

The paper [2] presents also a rule for reducing the depth of nesting of counterfactuals (*Overwriting* rule). Its proof was given under the assumption of recursivity. We show here that it can be extended to a much larger class. Following the terminology of [35], we say a causal team $T = (T^-, G, \mathcal{F})$ of signature $\sigma = (Dom, Ran)$ is **solutionful** if, for every assignment s of signature σ, and every consistent conjunction $\mathbf{X} = \mathbf{x}$ of signature σ, $(\{s\}, G, \mathcal{F})^-_{\mathbf{X}=\mathbf{x}} \neq \emptyset$. (Notice that the unique-solution causal teams are a special case of solutionful ones.) We then have:

Theorem 2.11. *Let $\sigma = (Dom, Ran)$ be a signature, $\mathbf{X} \subseteq Dom$ be distinct variables and $\mathbf{x} \in Ran(\mathbf{X})$ (therefore $\mathbf{X} = \mathbf{x}$ is a consistent conjunction). Let $\mathbf{Y} \subseteq Dom$ and $\mathbf{y} \in Ran(\mathbf{Y})$ and ψ a $\mathcal{COD}$ or $\mathcal{CO}_\sqcup$ formula of signature σ. Then, for every solutionful causal team T of signature σ:*

$$T \models \mathbf{X} = \mathbf{x} \Box\!\!\rightarrow (\mathbf{Y} = \mathbf{y} \Box\!\!\rightarrow \psi) \iff T \models (\mathbf{X}' = \mathbf{x}' \wedge \mathbf{Y} = \mathbf{y}) \Box\!\!\rightarrow \psi$$

where $\mathbf{X}' = \mathbf{X} \setminus \mathbf{Y}$ and $\mathbf{x}' = \mathbf{x} \setminus \mathbf{y}$.

In particular, this holds whenever T is unique-solution (resp. recursive).

Proof. Let us fix some notation:

$$\mathbf{U} := Exo(T) \setminus (\mathbf{X} \cup \mathbf{Y})$$
$$\mathbf{V} := Exo(T_{\mathbf{X}=\mathbf{x}}) \setminus \mathbf{Y} = (Exo(T) \cup \mathbf{X}) \setminus \mathbf{Y}$$

Furthermore, write

$$A := ((T_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}})^- = \{s \mid s(\mathbf{Y}) = \mathbf{y}, s \text{ comp. with } (\mathcal{F}_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}} \text{ and } s(\mathbf{V}) \in T^-_{\mathbf{X}=\mathbf{x}}(\mathbf{V})\}$$
$$B := (T_{\mathbf{X}'=\mathbf{x}' \wedge \mathbf{Y}=\mathbf{y}})^- = \{s \mid s(\mathbf{X}'\mathbf{Y}) = \mathbf{x}'\mathbf{y}, s \text{ comp. with } \mathcal{F}_{\mathbf{X}'=\mathbf{x}' \wedge \mathbf{Y}=\mathbf{y}} \text{ and } s(\mathbf{U}) \in T^-(\mathbf{U})\}.$$

Notice also that $(\mathcal{F}_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}} = \mathcal{F}_{\mathbf{X}'=\mathbf{x}' \wedge \mathbf{Y}=\mathbf{y}}$. The thesis will follow if we prove $A = B$.

$A \subseteq B$) Assume $s \in A$. Then $s(\mathbf{Y}) = \mathbf{y}$, s is compatible with $\mathcal{F}_{\mathbf{X}'=\mathbf{x}' \wedge \mathbf{Y}=\mathbf{y}}$, and $s(\mathbf{V}) \in T^-_{\mathbf{X}=\mathbf{x}}(\mathbf{V})$. In particular, since $\mathbf{X}' \subseteq \mathbf{V}$, we have by definition of $T_{\mathbf{X}=\mathbf{x}}$ that $s(\mathbf{X}') = \mathbf{x}'$. Thus, in order to prove that $s \in B$, we only need to prove that $s(\mathbf{U}) \in T^-(\mathbf{U})$. But $\mathbf{U} \subseteq \mathbf{V}$; thus $s(\mathbf{U}) \in T^-_{\mathbf{X}=\mathbf{x}}(\mathbf{U})$. By the definition of $T_{\mathbf{X}=\mathbf{x}}$, since $\mathbf{U} \subseteq Exo(T) \setminus \mathbf{X}$, we conclude $s(\mathbf{U}) \in T^-(\mathbf{U})$.

$B \subseteq A$) Assume $s \in B$. Then $s(\mathbf{Y}) = \mathbf{y}$ and s compatible with $(\mathcal{F}_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}}$. If we prove that $s(\mathbf{V}) \in T^-_{\mathbf{X}=\mathbf{x}}(\mathbf{V})$, we are done. Now if $X \in \mathbf{X} \setminus \mathbf{Y}$, we have (by definition of $s \in B$) that $s(X) = x \in T^-_{\mathbf{X}=\mathbf{x}}(X)$ (the last inclusion is by the definition of $T^-_{\mathbf{X}=\mathbf{x}}$). If instead $V \in \mathbf{V} \setminus \mathbf{X}$, we have $V \in Exo(T) \setminus (\mathbf{X} \cup \mathbf{Y}) = \mathbf{U}$. Thus $s(V) \in T^-(V)$ by definition of B. Since V is exogenous and $V \notin \mathbf{X}$, we have that any $t \in \{s\}^-_{\mathbf{X}=\mathbf{x}}$ is such that $t(V) = s(V)$. Since T is solutionful, there is such a t. Therefore, $s(V) \in T^-_{\mathbf{X}=\mathbf{x}}(V)$. □

Actually the proof of the equivalence above (for each fixed T and fixed formulas) requires weaker assumptions on T: that for every assignment s *in* T^-, $(\{s\}, G, \mathcal{F})_{\mathbf{X}=\mathbf{x}} \neq \emptyset$ (for the specific intervention $do(\mathbf{X} = \mathbf{x})$ corresponding to the antecedent of the formula $\mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow (\mathbf{Y} = \mathbf{y} \,\Box\!\!\rightarrow \psi)$). It is instead false in general, as the following example shows:

Example 2.12. Consider a causal team T of variables X, Y, Z, equations $Y := X + Z$, $Z := Y$, and with a single assignment $s(X) = s(Y) = s(Z) = 0$. Note that the system of equations has no solution with $X = 1$; thus $T^-_{X=1} = \emptyset$. By the empty team property, then, we obtain $T \models X = 1 \,\Box\!\!\rightarrow (Y = 1 \,\Box\!\!\rightarrow Z = 0)$. On the other hand, $T^-_{X=1 \wedge Y=1} = \{t\}$, where t is the assignment such that $t(X) = t(Y) = 1$ and $t(Z) = 1$. $T_{X=1 \wedge Y=1} \not\models Z = 0$, and thus $T \not\models (X = 1 \wedge Y = 1) \,\Box\!\!\rightarrow Z = 0$.

If we restrict the semantics to the case of unique-solution causal teams, it should be clear that the equivalence rules described above in this section allow transforming any formula of the form $\mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi$ into an equivalent formula in which only "unnested" counterfactuals occur, i.e. counterfactuals of the form $\mathbf{Z} = \mathbf{z} \,\Box\!\!\rightarrow \chi$, where χ is a formula without occurrences of $\Box\!\!\rightarrow$. In order to extend this observation to all formulas, we also need a substitution principle; a weaker version of it was proved in [2]. In its statement, we write $\varphi[\theta]$ to highlight a specific occurrence of θ as a subformula of $\varphi[\theta]$; and $\varphi[\theta']$ for the formula obtained by replacing this occurrence with another formula θ' (if θ is in an antecedent of $\supset$ and $\theta' \notin \mathcal{CO}$, or if θ is in an antecedent of $\Box\!\!\rightarrow$ and θ' is not of the form $\mathbf{X} = \mathbf{x}$, we consider the substitution $\varphi[\theta']$ to be undefined). If the occurrence θ in $\varphi[\theta]$ is in the antecedent of a counterfactual, we say it is a **neutral** occurrence; if it is not neutral and it occurs in an even number of antecedents of $\supset$ (possibly none), we say it is **positive**; and otherwise we say it is **negative**.

Lemma 2.13. *Let $\theta, \theta', \varphi[\theta]$ be $\mathcal{COD}(\sigma)$ or $\mathcal{CO}_{\sqcup}(\sigma)$ formulas. Suppose that the substitution formula $\varphi[\theta']$ is well-defined and that $\theta \models \theta'$. If the occurrence of θ in $\varphi[\theta]$ is positive, then $\varphi[\theta] \models \varphi[\theta']$; if it is negative, then $\varphi[\theta'] \models \varphi[\theta]$.*

Proof. We proceed by induction on the syntax of $\varphi[\theta]$. The cases for atoms $Y = y, =(X; Y)$ and for $\wedge, \vee, \sqcup, \Box\!\!\rightarrow$ are easy.

The only case left is that $\varphi[\theta]$ is $\alpha[\theta] \supset \chi[\theta]$. Suppose first that θ occurs positively in $\varphi[\theta]$. There are two subcases; suppose first that the occurrence of θ is in $\chi[\theta]$. Then $\alpha[\theta] = \alpha[\theta']$ and, by the inductive hypothesis, $\chi[\theta] \models \chi[\theta']$. Now assume $T \models \varphi[\theta]$. Then $T^{\alpha[\theta']} = T^{\alpha[\theta]} \models \chi[\theta] \models \chi[\theta']$, i.e. $T \models \varphi[\theta']$. Thus $\varphi[\theta] \models \varphi[\theta']$.

Suppose instead that the occurrence of θ is in $\alpha[\theta]$ (thus $\chi[\theta] = \chi[\theta']$). Since the occurrence of θ is positive in $\varphi[\theta]$, it is negative in $\alpha[\theta]$; thus by inductive hypothesis $\alpha[\theta'] \models \alpha[\theta]$. Now assume $T \models \varphi[\theta]$. Then $T^{\alpha[\theta]} \models \chi[\theta] = \chi[\theta']$. Since $\alpha[\theta'] \models \alpha[\theta]$, $T^{\alpha[\theta']}$ is a causal subteam of $T^{\alpha[\theta]}$; thus, by downward closure, $T^{\alpha[\theta']} \models \chi[\theta']$. Then $T \models \varphi[\theta']$; we have proved $\varphi[\theta] \models \varphi[\theta']$.

In the case that θ is a negative occurrence in $\varphi[\theta]$, we have to prove that $\varphi[\theta'] \models \varphi[\theta]$; the proof is completely symmetrical to the positive case. □

Corollary 2.14. *Let $\theta, \theta', \psi[\theta]$ be $\mathcal{COD}$ or $\mathcal{CO}_{\sqcup}$ formulas, and assume that the substitution formula $\psi[\theta']$ is well-defined. If $\theta \equiv \theta'$, then $\psi[\theta] \equiv \psi[\theta']$.*

Proof. Lemma 2.13 entails that the statement holds when θ is not a neutral occurrence. The case for neutral occurrences follows immediately from the observation that equivalent conjunctions of atomic formulas either define one and the same intervention, or they are both inconsistent. $\square$

Now, if we have any formula, we can replace each counterfactual occurring in it with an unnested formula that is equivalent to it (for unique-solution causal teams), using Lemma 2.10, 1.-2., Theorem 2.11 and Corollary 2.14. We have then proved:

Theorem 2.15. *For every $\mathcal{COD}(\sigma)$ (resp. $\mathcal{CO}_{\sqcup}(\sigma)$) formula φ there is an unnested $\mathcal{COD}(\sigma)$ (resp. $\mathcal{CO}_{\sqcup}(\sigma)$) formula φ' such that, for all unique-solution causal teams T:*

$$T \models \varphi \iff T \models \varphi'.$$

3. The main embedding results

3.1. First-order dependence logic

In this section, we will go through the basic definitions of first-order dependence logic. We will not discuss in any detail the properties of this logic or its connection with other logics based on team semantics, for which we refer to [31] and [13], but we will merely give the definitions that will be necessary for the rest of this work. In accordance with the conventions of causal team semantics, we will use capital letters $X, Y, Z, \ldots$ for variable symbols, and use non-capital letters $x, y, z, \ldots$ for *constant* symbols. It will be useful to think that the variables (resp. the values) used for the causal-observational languages come from this same pool of variables symbols (resp. constant symbols).

Definition 3.1 *(Team).* Let M be a first-order structure with at least two elements,[9] and let $\mathbf{V}$ be a finite set of variables. A team T over M with domain $\mathbf{V}$ is a set of variable assignments $s : \mathbf{V} \to M$.[10]

Definition 3.2 *(Team supplementation).* Let T be a team over some first-order structure M, and let $F : T \to M$ be a choice function selecting one element of M for each assignment of T.[11] Also, let X be a variable symbol. Then we write $T[F/X]$ for the team $\{s[F(s)/X] \mid s \in T\}$, where $s[F(s)/X]$ is the assignment obtained from s by fixing the value of the variable X to $F(s)$.

Definition 3.3 *(Team duplication).* Let T be a team over some first-order structure M. Also, let X be a variable symbol. Then we write $T[M/X]$ for the team $\{s(a/X) \mid s \in T, a \in M\}$.

Definition 3.4 *(Dependence logic, syntax).* An expression φ is a formula of dependence logic $\mathbf{FO}(=(\cdot;\cdot))$ if it is produced by the following grammar:

$$\varphi ::= \lambda \mid =(X_1, \ldots, X_n; Y) \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \forall V \varphi \mid \exists V \varphi$$

where λ is a first-order literal (in the usual sense) and $X_1, \ldots, X_n, Y, V$ are variable symbols.

[9] In this work, we will only consider first-order structures with at least two elements. This is a common assumption in the literature on team semantics.

[10] More precisely, we should say: $s : V \to \mathbf{Dom}(M)$, where $\mathbf{Dom}(M)$ is the set of all elements of M. For simplicity, we will write M both for a first-order structure and for the set of its elements.

[11] Here we give the so-called "strict" form of supplementation, rather than the "lax" one in which an assignment may be extended by picking more than one element. Since we will consider dependence logic proper, rather than other logics based on team semantics, these two forms of supplementation are expressively equivalent; and this form has the advantage of simplicity and was the originally given one.

A dependence logic formula without occurrences of dependence atoms is called a first-order formula. Given a dependence logic formula φ, the sets of its *free* and *bound* variables are defined as in the case of first-order logic (with the additional stipulation that the variables in a dependence atom are considered free). A dependence logic formula without free variables is called a *sentence*.

As before, we may abbreviate a dependence atom $=(X_1, \ldots, X_n; Y)$ as $=(\mathbf{X}; Y)$, and write a constancy atom $=(; Y)$ as $=(Y)$.

Definition 3.5 *(Dependence logic, semantics).* Let M be a first-order structure, let T be a team over M with domain $\mathbf{V}$, and let φ be a dependence logic formula over the signature of M with free variables in $\mathbf{V}$. Then we say that T satisfies φ in M, and we write $M, T \models \varphi$, if this can be derived by the following rules:

- If λ is a first-order literal then $M, T \models \lambda$ iff for all $s \in T$, it holds that $M, s \models \lambda$ in the sense of ordinary Tarskian semantics;
- $M, T \models =(X_1 \ldots X_n; Y)$ iff any two $s, s' \in T$ that agree over $X_1 \ldots X_n$ also agree over Y;
- $M, T \models \varphi_1 \vee \varphi_2$ iff $T = T_1 \cup T_2$ for some $T_1, T_2 \subseteq T$ such that $M, T_1 \models \varphi_1$ and $M, T_2 \models \varphi_2$;
- $M, T \models \varphi_1 \wedge \varphi_2$ iff $M, T \models \varphi_1$ and $M, T \models \varphi_2$;
- $M, T \models \exists X \varphi$ iff there exists a choice function F such that $M, T[F/X] \models \varphi$;
- $M, T \models \forall X \varphi$ iff $M, T[M/X] \models \varphi$.

If φ is a dependence logic sentence, we say that φ is true in a model M (and we write $M \models \varphi$) if $M, \{\emptyset\} \models \varphi$, where $\emptyset$ is the empty assignment.

A few well-known basic facts about dependence logic that will be of use for the rest of the work (and for whose proofs we refer to [31] or to simple application of the rules of team semantics) are the following:

Proposition 3.6 *(Flatness). Let φ be a dependence logic formula in which no dependence atom $=(X_1 \ldots X_n, Y)$ appears (i.e., let φ be a first-order formula). Then $M, T \models \varphi$ if and only if, for all $s \in T$, $M, s \models \varphi$ in the ordinary sense of Tarski's semantics.*

Proposition 3.7 *(Downwards closure). Let φ be a dependence logic formula and let M and T be a structure and a team such that $M, T \models \varphi$. Then, for all teams $T' \subseteq T$, we have that $M, T' \models \varphi$.*

Proposition 3.8 *(Empty team property). Let φ be a dependence logic formula and let M be a structure whose signature contains the signature of φ. Then $M, \emptyset \models \varphi$.*

Proposition 3.9 *(Locality). Let φ be a dependence logic formula, let M be a structure whose signature contains the signature of φ, and let T, T' be two teams over M whose restriction*[12] *to the free variables of φ is the same. Then $M, T \models \varphi$ if and only if $M, T' \models \varphi$.*

Let us now recall (and sketch the proofs of) some statements that are well-known in the dependence logic research community.

Proposition 3.10. *If X and Y are distinct variables and φ is any dependence logic formula, $\exists X \exists Y \varphi$ is equivalent to $\exists Y \exists X \varphi$; and furthermore, $\exists X \exists X \varphi$ is equivalent to $\exists X \varphi$.*

[12] The restriction of a team T to a set of variables $\mathbf{V}$ is the team $T_{\upharpoonright \mathbf{V}} := \{s_{\upharpoonright \mathbf{V}} \mid s \in T\}$.

Proof. Suppose that $M, T \models \exists X \exists Y \varphi$, where by locality (Proposition 3.9) we can assume that T contains neither X nor Y in its domain.

Then there exists a choice function F with domain T and a choice function G with domain $T[F/X]$ such that $M, T[F/X][G/Y] \models \varphi$. Now define the choice function G', with domain T, as

$$G'(s) = G(s[F(s)/X]).$$

Furthermore, define the choice function F', with domain $T[G'/Y]$, as

$$F'(s[y/Y]) = F(s).$$

Notice that G' and F' are well-defined by our assumption that X, Y are not in the domain of T (and thus of s), together with the fact that X and Y are distinct. Then $T[G'/Y][F'/X] = T[F/X][G/Y]$: indeed, $s[y/Y][x/X] \in T[G'/Y][F'/X]$ if and only if $s \in T$, $y = G(s[F(s)/X])$, and $x = F(s)$, i.e. if and only if $s \in T$, $x = F(s)$, and $y = G(s[x/X])$, i.e. if and only if $s[x/X][y/Y] \in T[F/X][G/Y]$; and since X and Y are distinct we also have that $s[y/Y][x/X] = s[x/X][y/Y]$, and therefore $s[y/Y][x/X] \in T[G'/Y][F'/X] \iff s[x/X][y/Y] \in T[F/X][G/Y] \iff s[y/Y][x/X] \in T[F/X][G/Y]$.

Then in particular, since by assumption we have that $M, T[F/X][G/Y] \models \varphi$, we also have that $M, T[G'/Y][F'/X] \models \varphi$ and hence that $M, T \models \exists Y \exists X \varphi$.

Conversely, suppose that $M, T \models \exists Y \exists X \varphi$: then by the same argument (with X and Y swapped) we can conclude that $M, T \models \exists X \exists Y \varphi$.

Finally, let us consider the case that $X = Y$: $M, T \models \exists X \exists X \varphi$ if and only if there exists some F such that $M, T[F/X] \models \exists X \varphi$. But T and $T[F/X]$ agree over the free variables of $\exists X \varphi$, since X is not free in $\exists X \varphi$, and so by locality this is the case if and only if $M, T \models \exists X \varphi$, as required. □

Corollary 3.11. *Let $\mathbf{X} = (X_1 \ldots X_n)$ be a tuple of – possibly repeated – variables, and let $\mathbf{Z} = (Z_1 \ldots Z_k)$ list the same variables without repetitions (and possibly in a different order). Then, for all φ, the formulas $\exists X_1 \ldots X_n \varphi$ and $\exists Z_1 \ldots Z_k \varphi$ are equivalent.*

Proof. Repeatedly apply Proposition 3.10 to reorder the X_i in the order of the $\mathbf{Z}$ and eliminate doubles. □

The next two statements are also well known in the dependence logic research community. We provide a proof sketch for the first (and less obvious) one:

Proposition 3.12. *For any first-order structure M, team T, dependence logic formula φ and sequence of variables $\mathbf{X} = X_1 \ldots X_n$, $M, T \models \exists X_1 \ldots \exists X_n \varphi$ if and only if there exists a tuple $\mathbf{F} = (F_1 \ldots F_n)$ of functions $F_i : T \to M$ such that $M, T[\mathbf{F}/\mathbf{X}] \models \varphi$, where $T[\mathbf{F}/\mathbf{X}] := \{s[F_1(s)/X_1] \ldots [F_n(s)/X_n] \mid s \in T\}$.*

Proof. By locality, we can assume that the variables of $\mathbf{X}$ are not in the domain of T.

For the left-to-right direction, we proceed by induction on n. The base case $n = 1$ is simply the rule for existential quantification in first-order team semantics. Suppose now that the statement holds for n, and that $M, T \models \exists X_1 \ldots \exists X_n \exists X_{n+1} \varphi$. Let $\mathbf{Z} = Z_1 \ldots Z_k$ enumerate, without repetition, the variables of $\{X_1 \ldots X_n\} \setminus \{X_{n+1}\}$: then $k \leq n$ and, by Corollary 3.11, $M, T \models \exists Z_1 \ldots \exists Z_k \exists X_{n+1} \varphi$.

Now, by induction hypothesis, there exists a tuple $\mathbf{G} = (G_1 \ldots G_k)$ of functions $T \to M$ such that $M, T[\mathbf{G}/Z_1 \ldots Z_k] \models \exists X_{n+1} \varphi$. Therefore, there exists some function $G'_{k+1} : T[\mathbf{G}/Z_1 \ldots Z_k] \to M$ such that $M, T[\mathbf{G}/Z_1 \ldots Z_k][G'_{k+1}/X_{n+1}] \models \varphi$. But now, if we define $G_{k+1} : T \to M$ as $G_{k+1}(s) = G'_{k+1}(s[G_1(s)/Z_1] ... [G_k(s)/Z_k])$ and we let $(\mathbf{G}; G_{k+1})$ be the tuple $(G_1 \ldots G_k, G_{k+1})$, we have that $T[\mathbf{G}/Z_1 \ldots Z_k][G'_{k+1}/X_{n+1}] = T[(\mathbf{G}; G_{k+1})/Z_1 \ldots Z_k X_{n+1}]$, and hence that $M, T[(\mathbf{G}; G_{k+1})/Z_1 \ldots Z_k X_{n+1}] \models \varphi$. Finally, define the tuple of functions $\mathbf{F} = (F_1 \ldots F_{n+1})$ so that, whenever $X_i = Z_j$, $F_i = G_j$ and, whenever $X_i = X_{n+1}$, $F_i = G_{k+1}$: then $T[(\mathbf{G}; G_{k+1})/Z_1 \dots Z_k X_{k+1}] = T[\mathbf{F}/X_1 \dots X_n X_{n+1}]$ and finally $M, T[\mathbf{F}/X_1 \dots X_n X_{n+1}] \models \varphi$, as required.

For the right-to-left direction, we also proceed by induction. Again, the base case is simply the rule for existential quantification. Suppose now that the statement holds for n and that $M, T[\mathbf{F}/\mathbf{X}] \models \varphi$ for $\mathbf{F} = (F_1 \dots F_{n+1})$ and $\mathbf{X} = (X_1 \dots X_{n+1})$. Now let $\mathbf{F}' = (F_1 \dots F_n)$, let $\mathbf{X}' = (X_1 \dots X_n)$, and let $F'_{n+1} : T[\mathbf{F}'/\mathbf{X}'] \to M$ be such that $F'_{n+1}(s') = F_{n+1}(s)$ for some $s \in T$ such that $s[F_1(s)/X_1] \dots [F_n(s)/X_n] = s'$.[13]

Then $T[\mathbf{F}'/\mathbf{X}'][F'_{n+1}/X_{n+1}] \subseteq T[\mathbf{F}/\mathbf{X}]$, and so by downwards closure (Proposition 3.7) we have that $M, T[\mathbf{F}'/\mathbf{X}'][F'_{n+1}/X_{n+1}] \models \varphi$. But then $M, T[\mathbf{F}'/\mathbf{X}'] \models \exists X_{n+1}\varphi$, and so by induction hypothesis $M, T \models \exists X \dots \exists X_n \exists X_{n+1} \varphi$. □

Proposition 3.13. *For all first-order structures M, teams T, dependence logic formulas φ and sequences of variables $\mathbf{X} = X_1 \dots X_n$, $M, T \models \forall X_1 \dots \forall X_n \varphi$ if and only if $M, T[\mathbf{M}/\mathbf{X}] \models \varphi$, where $T[\mathbf{M}/\mathbf{X}] := \{s[a_1/X_1] \dots [a_n/X_n] \mid s \in T, a_1 \dots a_n \in M\}$.*

Two additional connectives that have been studied in the context of dependence logic are the *Constant Quantification* $\exists^1$ and the *Global* (or *Boolean*) Disjunction $\sqcup$, whose semantic rules are given by

- $M, T \models \exists^1 X \psi$ iff there exists some fixed element x of M such that $M, T[x/X] \models \psi$, where $T[x/X] = \{s(x/X) : s \in T\}$;
- $M, T \models \psi_1 \sqcup \psi_2$ iff $M, T \models \psi_1$ or $M, T \models \psi_2$.

It is easy to see that the previous results (3.10-3.11-3.12-3.13) also apply to the language $\mathbf{FO}(=(\cdot;\cdot), \sqcup, \exists^1)$ that extends dependence logic with these two additional connectives (the results work for any local, downward closed language). The logic $\mathbf{FO}(=(\cdot;\cdot), \sqcup, \exists^1)$ is known to be equivalent to dependence logic. In fact, we can say more, as every expression of this logic is reducible to an expression of $\mathbf{FO}(=(\cdot;\cdot))$ with some extra *initial* existential quantifiers:

Proposition 3.14. *$\sqcup$ and $\exists^1$ commute with all connectives of dependence logic, up to renamings, i.e.*

- *For $\circ = \wedge$ or $\circ = \vee$, $(\varphi \sqcup \psi) \circ \theta \equiv (\varphi \circ \theta) \sqcup (\psi \circ \theta)$, and likewise $\varphi \circ (\psi \sqcup \theta) \equiv (\varphi \circ \psi) \sqcup (\varphi \circ \theta)$*
- *$\exists X(\varphi \sqcup \psi) \equiv (\exists X \varphi) \sqcup (\exists X \psi)$;*
- *$\forall X(\varphi \sqcup \psi) \equiv (\forall X \varphi) \sqcup (\forall X \psi)$*

and

- *For $\circ = \wedge$ or $\circ = \vee$, $(\exists^1 X \varphi) \circ \psi \equiv \exists^1 Y(\varphi[Y/X] \circ \psi)$, where Y is a new variable not occurring in φ or in ψ and $\varphi[Y/X]$ is obtained by changing all occurrences of X in φ that were bound by $\exists^1 X$ into occurrences of Y;*
- *$\exists X(\exists^1 Y \varphi) \equiv \exists^1 Y(\exists X \varphi)$ if X and Y are distinct variables, and $\exists X(\exists^1 X \varphi) \equiv \exists^1 X \varphi$ otherwise;*
- *$\forall X(\exists^1 Y \varphi) \equiv \exists^1 Y(\forall X \varphi)$ if X and Y are distinct variables, and $\forall X(\exists^1 X \varphi) \equiv \exists^1 X \varphi$ otherwise.*

Furthermore, $\exists^1$ distributes over $\sqcup$, in the sense that $\exists^1 Y(\psi \sqcup \theta) \equiv (\exists^1 Y \psi) \sqcup (\exists^1 Y \theta)$.

Proof. Straightforward by examining the semantic rules. □

[13] At least one such s exists, since $s' \in T[\mathbf{F}'/\mathbf{X}']$. It is possible for multiple such s to exist; in this case, any choice may be taken.

In the formulation of the following result we call "non-constant existential quantifiers" the occurrences of the symbol $\exists$, as opposed to occurrences of $\exists^1$.

Theorem 3.15. *Every formula φ of $\mathbf{FO}(=(\cdot;\cdot),\sqcup,\exists^1)$ is equivalent to some formula φ' of dependence logic $\mathbf{FO}(=(\cdot;\cdot))$, which can be assumed to be first order in case φ is flat and has no occurrences of dependence atoms; and if no non-constant existential quantifier appears in the scope of some universal quantifier in φ, then no non-constant existential quantifier appears in the scope of some universal quantifier in φ'. If φ has a purely relational vocabulary, the same holds for φ'.*

Proof. By the previous result, φ is equivalent to some expression of the form

$$\bigsqcup_i (\exists^1 \mathbf{Y}_i \varphi_i) \tag{1}$$

where each φ_i is a dependence logic formula; and it is easy to see that if no non-constant existential quantifier appeared in the scope of a universal quantifier in φ, that does not happen in any φ_i either.

Then observe that $\exists^1 Y\psi \equiv \exists Y(=(Y)\wedge\psi)$ and that $\varphi_1 \sqcup \varphi_2 \equiv \exists Z \exists W(=(Z)\wedge =(W)\wedge((Z = W\wedge\varphi_1)\vee(Z \neq W \wedge \varphi_2)))$,[14] where Z and W are new variables not appearing in φ_1 and φ_2. Thus, φ is equivalent to some expression in the required form.

Now suppose that no dependence atoms appear in φ and that this formula is flat (in the sense of Proposition 3.6). Since flatness is preserved by logical equivalence, and since the formula transformation just described does not add dependence atoms, also (1) has the same properties. Thus in particular all the φ_i are first order and therefore flat. Then, as we will now show, (1) is logically equivalent to

$$\bigvee_i (\exists \mathbf{Y}_i \varphi_i) \tag{2}$$

which is a first-order formula.

Indeed, every team that satisfies (1) satisfies (2), since any team that satisfies $\exists^1 X\psi$ also satisfies $\exists X\psi$ (you can always pick a constant value for X) and any team that satisfies $\psi_1 \sqcup \psi_2$ satisfies $\psi_1 \vee \psi_2$ by the empty team property.

Conversely, every team that satisfies (2) is the union of teams that satisfy (1). Indeed, suppose that $M, T \models \bigvee_i(\exists \mathbf{Y}_i\varphi_i)$: then for every assignment $s \in T$, we have (by downwards closure) that $(M, \{s\}) \models \bigvee_i(\exists \mathbf{Y}_i\varphi_i)$, and hence there exists an index i and some tuple of elements $\mathbf{m}$ such that $M, \{s(\mathbf{m}/\mathbf{Y}_i)\} \models \varphi_i$. Therefore, for all such s we have that $\{s\}$ satisfies (1) in M; and since (1) is flat, we can conclude that T satisfies (1) in M. $\square$

3.2. The embedding result, in the general case

We will now show how the causal-observational languages can be embedded into the first-order (dependence) languages just described.

The embedding will be specified in two phases: 1) associating to each causal team an appropriate first-order structure and a team over the structure; 2) defining a truth-preserving translation from causal-observational languages to first-order languages. Concerning 2), we will consider a number of alternative translations. Those we describe in the present section are meant to work for all causal teams (without the restriction of recursiveness) under the Halpern-style semantics described in sections 2.4-2.5.

[14] The latter equivalence holds due to our stipulation that all structure have at least two elements. A formula covering also the case of one-element structures can be obtained by adding a disjunct of the form $\forall X \forall Y (X = Y) \wedge (\varphi_1 \vee \varphi_2)$ and observing that, if there is only one element, all teams are empty or singletons.

We now present the first phase of the embedding. To each causal team T we associate a team (its team component T^-) and a first-order structure M_T, so that the pair (M_T, T^-) encodes in a natural way the content of T.

Definition 3.16. To each causal team $T = (T^-, G_T, \mathcal{F}_T)$ of signature $\sigma = (Dom, Ran)$ and endogenous variables $End(T)$ we associate a structure $M_T = (\mathbf{Dom}(M_T), (c^{M_T})_{c\in|M_T|}, (f_V^{M_T})_{V\in End(T)})$, where:

- $\mathbf{Dom}(M_T) = \bigcup_{V\in Dom} Ran(V)$
- $c^{M_T} = c$
- $f_V^{M_T}(\mathbf{c}) = \begin{cases} \mathcal{F}_T(V)(\mathbf{c}) & \text{if } \mathbf{c} \in Ran(PA_V) \\ \text{an arbitrary } d \in Ran(V) & \text{otherwise.} \end{cases}$

The arity of each function symbol f_V is $card(PA_V)$.

We remark that in this definition we are again using the ambiguity between a value c and the constant symbol c that denotes it; so each constant symbol is denoted by itself in M_T. Notice also that the team component of T is not involved in the definition of M_T; therefore, a causal subteam S of T will have the same associated first-order structure $M_S = M_T$. Notice also that, for any consistent $\mathbf{X} = \mathbf{x}$, $M_{T_{\mathbf{X}=\mathbf{x}}}$ is a reduct of M_T.

Now to the second step: we will show that, for an appropriate translation tr of $\mathcal{COD} \cup \mathcal{CO}_{\sqcup}$ formulas into formulas of first-order (dependence) logic plus $\sqcup$, we have: $T \models \varphi \iff M_T, T^- \models tr(\varphi)$. More precisely, we define a distinct translation $tr(\varphi, G)$ for each spanning subgraph G of G_T, together with an auxiliary family of “dual translations” $tr^d(\varphi, G)$. (The latter are needed in order to translate the antecedents of selective implications.) The definition will proceed by induction on the syntax of φ. The idea here is that, for each intervention $do(\mathbf{X} = \mathbf{x})$, the formula $tr(\psi, G_{T_{\mathbf{X}=\mathbf{x}}})$ will encode the fact that ψ holds in the modified causal team $T_{\mathbf{X}=\mathbf{x}}$; we need this step in order to define the translation of a counterfactual formula $\mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi$. For each spanning subgraph G of G_T, we define a quantifier-free formula

$$Eq(G) := \bigwedge_{V\in End(G)} V = f_V(PA_V),$$

where PA_V is the list of the parent variables of V.[15] By this definition, $Eq(G_T)$ asserts that the system of equations $V = \mathcal{F}(V)(PA_V)$ $(V \in End(T))$ associated to T holds; and $Eq(G_{T_{\mathbf{X}=\mathbf{x}}})$ will similarly describe the reduced system of equations that is obtained after applying the intervention $do(\mathbf{X} = \mathbf{x})$ to T. Notice that $Eq(G)$ is a first-order formula in case Dom is finite; otherwise, $Eq(G)$ can be either an infinitary conjunction or it may have occurrences of functions of infinite arity. It is worth emphasizing here that $Eq(G)$ depends only on the graph G, not on the functions $\mathcal{F}$ associated to it in a given causal team. The interpretation of the symbols f_V according to Definition 3.16 will vary depending on $\mathcal{F}$, but the expression $Eq(G)$ itself will not.

We now define by simultaneous induction the two (relativized) translations $tr(\varphi, G)$ and $tr^d(\varphi, G)$, where (provided G is a finite graph) $tr(\varphi, G)$ goes from $\mathcal{COD}$ or $\mathcal{CO}_{\sqcup}$ to the language of dependence logic and $tr^d(\varphi, G)$ goes from $\mathcal{CO}$ to the language of first-order logic. Write $\mathbf{V}$ for the set of variables that are endogenous according to graph $G_{\mathbf{X}=\mathbf{x}}$, and define

- $tr(\eta, G) = \eta$ if η is $X = x$ or $X \neq x$ or $=(\mathbf{X}; Y)$.

[15] Remember that we are taking the variables from the causal-observational languages to be subsets of the set of first-order variables. Thus, $V = f_V(PA_V)$ is a first-order formula.

- $tr(\psi \circ \chi, G) := tr(\psi, G) \circ tr(\chi, G)$ for $\circ = \wedge, \vee$ or $\sqcup$.
- $tr(\alpha \supset \chi, G) := tr^d(\alpha, G) \vee tr(\chi, G)$.
- $tr(\mathbf{X} = \mathbf{x} \boxright \psi, G) :=$

$$\begin{cases} \exists^1 \mathbf{X}(\mathbf{X} = \mathbf{x} \wedge \forall \mathbf{V}(Eq(G_{\mathbf{X}=\mathbf{x}})^d \vee tr(\psi, G_{\mathbf{X}=\mathbf{x}}))) & \text{if } \mathbf{X} = \mathbf{x} \text{ is consistent} \\ \top & \text{if } \mathbf{X} = \mathbf{x} \text{ is inconsistent.} \end{cases}$$ [16]

and

- $tr^d(\eta, G) = X \neq x$ if η is $X = x$, and vice versa.
- $tr^d(\alpha \circ \beta, G) := tr^d(\alpha, G) \circ' tr^d(\beta, G)$, where $\wedge' = \vee$ and $\vee' = \wedge$.
- $tr^d(\alpha \supset \beta, G) := tr(\alpha, G) \wedge tr^d(\beta, G)$.
- $tr^d(\mathbf{X} = \mathbf{x} \boxright \beta, G) :=$

$$\begin{cases} \exists^1 \mathbf{X} \exists \mathbf{V}(Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr^d(\beta, G_{\mathbf{X}=\mathbf{x}})) & \text{if } \mathbf{X} = \mathbf{x} \text{ is consistent} \\ \bot & \text{if } \mathbf{X} = \mathbf{x} \text{ is inconsistent.} \end{cases}$$

Let us now prove the correctness of the translation.

Remark 3.17. The translations $tr(\varphi, G)$ and $tr^d(\varphi, G)$ depend only on the expression φ and the graph G, and not on any team component T^- or any function component $\mathcal{F}$.

Theorem 3.18. *Let $T = (T^-, G, \mathcal{F})$ be a causal team and φ a formula of $\mathcal{COD}$ or $\mathcal{CO}_\sqcup$. Then:*

$$M_T, T^- \models tr(\varphi, G) \iff T \models \varphi.$$

Furthermore, if φ is in $\mathcal{CO}$,

$$M_T, T^- \models tr^d(\varphi, G) \iff \forall s \in T^-, (\{s\}, G, \mathcal{F}) \not\models \varphi.$$

Proof. The proof is, for the most part, a straightforward induction on φ,[17] with only the cases of $\supset$ and $\boxright$ being nontrivial. Let us show all of them anyway:

- $M_T, T^- \models X = x$ if and only if $s(X) = x$ for all $s \in T^-$, that is, if and only if $T \models X = x$. On the other hand, $M_T, T^- \models X \neq x$ if and only if for all $s \in T^-$, $s(X) \neq x$, as required.
  As for the case of $=(\mathbf{X}; Y)$, since this is expression is not in $\mathcal{CO}$ the second part of the theorem is trivially true. As for the first part, $M_T, T^- \models =(\mathbf{X}; Y)$ if and only if any two $s, s' \in T^-$ that agree wrt $\mathbf{X}$ also agree wrt Y, i.e. if and only if $T \models =(\mathbf{X}; Y)$.
- $M_T, T^- \models tr(\psi, G) \wedge tr(\chi, G)$ if and only if $M_T, T^- \models tr(\psi, G)$ and $M_T, T^- \models tr(\chi, G)$, that is, by induction hypothesis, if and only if $T \models \psi$ and $T \models \chi$ and thus $T \models \psi \wedge \chi$.
  Suppose instead that $M_T, T^- \models tr^d(\psi \wedge \chi, G)$, that is, $T^- = T_1^- \cup T_2^-$ for two subteams T_1^-, T_2^- such that $M_T, T_1^- \models tr^d(\psi, G)$ and $M_T, T_2^- \models tr^d(\chi, G)$. Then by induction hypothesis, for all $s \in T^-$ we have that $(\{s\}, G, \mathcal{F}) \not\models \psi$ or $(\{s\}, G, \mathcal{F}) \not\models \chi$, depending on whether s is in T_1^- or in T_2^-; and in either case we thus have that $(\{s\}, G, \mathcal{F}) \not\models \psi \wedge \chi$, as required.

[16] Note that $Eq(G_{\mathbf{X}=\mathbf{x}})$ is a first-order formula and that $Eq(G_{\mathbf{X}=\mathbf{x}})^d$ is simply its dual in the usual first-order sense: the only application of the dual translation tr^d is in the rule for selective implication $\supset$.
[17] The statement must be proved simultaneously for all T.

Conversely, suppose that for all $s \in T^-$, $(\{s\}, G, \mathcal{F}) \not\models \psi \wedge \chi$. Then $T^- = T_1^- \cup T_2^-$, where $T_1^- = \{s \in T^- : (\{s\}, G, \mathcal{F}) \not\models \psi\}$ and $T_2^- = \{s \in T^- : (\{s\}, G, \mathcal{F}) \not\models \chi\}$; and, by induction hypothesis, $M_T, T_1^- \models tr^d(\psi, G)$ and $M_T, T_2^- \models tr^d(\chi, G)$; so in conclusion $M_T, T^- \models tr^d(\psi, G) \vee tr^d(\chi, G)$.

- $M_T, T^- \models tr(\psi, G) \vee tr(\chi, G)$ if and only if $T^- = T_1^- \cup T_2^-$ for two subteams T_1^-, T_2^- such that $M_T, T_1^- \models tr(\psi, G)$ and $M_T, T_2^- \models tr(\chi, G)$. By induction hypothesis, this is the case if and only if $(T_1^-, G, \mathcal{F}) \models \psi$ and $(T_2^-, G, \mathcal{F}) \models \chi$, that is, if and only if $(T^-, G, \mathcal{F}) \models \psi \vee \chi$.
Suppose instead that $M_T, T^- \models tr^d(\psi, G) \wedge tr^d(\chi, G)$. Then by induction hypothesis, for all $s \in T^-$ we have that $(\{s\}, G, \mathcal{F}) \not\models \psi$ and $(\{s\}, G, \mathcal{F}) \not\models \chi$, and thus $(\{s\}, G, \mathcal{F}) \not\models \psi \vee \chi$, as required.
Conversely, if $(\{s\}, G, \mathcal{F}) \not\models \psi \vee \chi$ for all $s \in T^-$ then, for all such s, $(\{s\}, G, \mathcal{F}) \not\models \psi$ and $(\{s\}, G, \mathcal{F}) \not\models \chi$, and so by induction hypothesis $M_T, T^- \models tr^d(\psi, G)$ and $M_T, T^- \models tr^d(\chi, G)$; thus, $M_T, T^- \models tr^d(\psi \vee \chi, G)$, as required.
- $M_T, T^- \models tr(\psi, G) \sqcup tr(\chi, G)$ if and only if $(M_T, T^-) \models tr(\psi, G)$ or $(M_T, T^-) \models tr(\chi, G)$, that is, by induction hypothesis, if and only if $(T^-, G, \mathcal{F}) \models \psi$ or $(T^-, G, \mathcal{F}) \models \chi$, i.e. $(T^-, G, \mathcal{F}) \models \psi \sqcup \chi$.
Since $\psi \sqcup \chi$ is not in $\mathcal{CO}$, the second part of the theorem is trivially true.
- Suppose that $M_T, T^- \models tr^d(\alpha, G) \vee tr(\psi, G)$, where $\alpha \in \mathcal{CO}$ by the definition of our syntax. This is the case if and only if $T^- = T_1^- \cup T_2^-$ for two T_1^-, T_2^- such that $M_T, T_1^- \models tr^d(\alpha, G)$ and $M_T, T_2^- \models tr(\psi, G)$, that is, by induction hypothesis, if $T^- = T_1^- \cup T_2^-$ for two T_1^-, T_2^- such that $(\{s\}, G, \mathcal{F}) \not\models \alpha$ for all $s \in T_1^-$ and $(T_2^-, G, \mathcal{F}) \models \psi$. If this is the case, then since $(T^\alpha)^- = \{s \in T^- : (\{s\}, G, \mathcal{F}) \models \alpha\} \subseteq T_2$, by downwards closure we have that $((T^\alpha)^-, G, \mathcal{F}) \models \psi$ and hence that $(T^-, G, \mathcal{F}) \models \alpha \supset \psi$, as required; and conversely, if this is the case then $((T^\alpha)^-, G, \mathcal{F}) \models \psi$, and so we can let $T_2^- = (T^\alpha)^-$ and $T_1^- = T^- \backslash T_2^-$.
Suppose instead that $M_T, T^- \models tr(\alpha, G)$ and $M_T, T^- \models tr^d(\psi, G)$, where $\alpha, \psi \in \mathcal{CO}$. Then by induction hypothesis we have that $(T^-, G, \mathcal{F}) \models \alpha$ – and hence, by Downwards Closure, $(\{s\}, G, \mathcal{F}) \models \alpha$ for all $s \in T^-$ – and that, for all $s \in T^-$, $(\{s\}, G, \mathcal{F}) \not\models \psi$. So for all $s \in T^-$ we have that $(\{s\}, G, \mathcal{F}) \not\models \alpha \supset \psi$, as required.
Conversely, suppose that for all $s \in T^-$, $(\{s\}, G, \mathcal{F}) \not\models \alpha \supset \psi$. Then all such s must satisfy α but not ψ, and therefore, by flatness of α and induction hypothesis, $M_T, T^- \models tr(\alpha, G)$ and $M_T, T^- \models tr^d(\psi, G)$.
- The case for $\mathbf{X} = \mathbf{x}$ inconsistent is easy to prove, so let us assume that $\mathbf{X} = \mathbf{x}$ is consistent.
Suppose that $M_T, T^- \models \exists^1 \mathbf{X}(\mathbf{X} = \mathbf{x} \wedge \forall \mathbf{V}(Eq(G_{\mathbf{X}=\mathbf{x}})^d \vee tr(\psi, G_{\mathbf{X}=\mathbf{x}})))$. Then $T^-[\mathbf{x}/\mathbf{X}][\mathbf{M}/\mathbf{V}]$ can be split into two subteams T_1^- and T_2^- such that

$$M_T, T_1^- \models Eq(G_{\mathbf{X}=\mathbf{x}})^d \tag{3}$$
$$M_T, T_2^- \models \mathbf{X} = \mathbf{x} \wedge tr(\psi, G_{\mathbf{X}=\mathbf{x}}). \tag{4}$$

Since $Eq(G_{\mathbf{X}=\mathbf{x}})^d$ is flat, we can assume without loss of generality that T_1^- contains *all* assignments satisfying $Eq(G_{\mathbf{X}=\mathbf{x}})^d$; thus, since the formula $\mathbf{X} = \mathbf{x} \wedge tr(\psi, G_{\mathbf{X}=\mathbf{x}})$ is downward closed, we can also assume that no assignment in T_2^- satisfies $Eq(G_{\mathbf{X}=\mathbf{x}})^d$. Thus, since M, T_2^- satisfies both $\mathbf{X} = \mathbf{x}$ and $Eq(G_{\mathbf{X}=\mathbf{x}})$, we have $T_2^- \subseteq T_{\mathbf{X}=\mathbf{x}}^-$. On the other hand, since clearly $T_{\mathbf{X}=\mathbf{x}}^- \subseteq T^-[\mathbf{x}/\mathbf{X}][\mathbf{M}/\mathbf{V}]$ and the assignments in T_1^- do not satisfy $Eq(G_{\mathbf{X}=\mathbf{x}})$, we must also have $T_{\mathbf{X}=\mathbf{x}}^- \subseteq T_2^-$. Thus $T_{\mathbf{X}=\mathbf{x}}^- = T_2^-$.
From (4) we also obtain

$$M_{T_{\mathbf{X}=\mathbf{x}}}, T_2^- \models tr(\psi, G_{\mathbf{X}=\mathbf{x}}), \tag{5}$$

since $M_{T_{\mathbf{X}=\mathbf{x}}}$ is the reduct of M_T to the vocabulary of $tr(\psi, G_{\mathbf{X}=\mathbf{x}})$. Since $T_{\mathbf{X}=\mathbf{x}}^- = T_2^-$, we can apply the inductive hypothesis and obtain:

$$(T_{\mathbf{X}=\mathbf{x}}^-, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \models \psi. \tag{6}$$

This implies that $(T^-, G, \mathcal{F}) \models \mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi$, as required.

Conversely, suppose that $(T^-, G, \mathcal{F}) \models \mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi$. Then we have that $(T^-_{\mathbf{X}=\mathbf{x}}, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \models \psi$, and thus – by induction hypothesis – $M_{T_{\mathbf{X}=\mathbf{x}}} T^-_{\mathbf{X}=\mathbf{x}} \models tr(\psi, G_{\mathbf{X}=\mathbf{x}})$. Then, since $M_{T_{\mathbf{X}=\mathbf{x}}}$ is a reduct of M_T, $M_T, T^-_{\mathbf{X}=\mathbf{x}} \models tr(\psi, G_{\mathbf{X}=\mathbf{x}})$. Additionally, $T^-_{\mathbf{X}=\mathbf{x}} \subseteq T^-[\mathbf{x}/\mathbf{X}][\mathbf{M}/\mathbf{V}]$, and all assignments in $T^-_0 = T^-[\mathbf{x}/\mathbf{X}][\mathbf{M}/\mathbf{V}] \setminus T^-_{\mathbf{X}=\mathbf{x}}$ violate some structural equation of $G_{\mathbf{X}=\mathbf{x}}$ (and therefore satisfy $Eq(G_{\mathbf{X}=\mathbf{x}})^d$). Thus, $M_T, T^-_0 \models Eq(G_{\mathbf{X}=\mathbf{x}})^d$ and $M_T, T^-[\mathbf{x}/\mathbf{X}] \models \forall \mathbf{V}(Eq(G_{\mathbf{X}=\mathbf{x}})^d \vee tr(\psi, G_{\mathbf{X}=\mathbf{x}}))$. Therefore, $M_T, T^- \models \exists^1 \mathbf{X}(\mathbf{X} = \mathbf{x} \wedge \forall \mathbf{V}(Eq(G_{\mathbf{X}=\mathbf{x}})^d \vee tr(\psi, G_{\mathbf{X}=\mathbf{x}})))$, as required.

Now suppose that $M_T, T^- \models \exists^1 \mathbf{X} \exists \mathbf{V}(Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr^d(\psi, G_{\mathbf{X}=\mathbf{x}}))$ and $s \in T^-$. By downward closure, $M_T, \{s\} \models \exists^1 \mathbf{X} \exists \mathbf{V}(Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr^d(\psi, G_{\mathbf{X}=\mathbf{x}}))$. Then there are functions $\mathbf{F}$ such that $M_T, \{s\}[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{V}] \models Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr^d(\psi, G_{\mathbf{X}=\mathbf{x}})$. By the first two conjuncts, we have then that $(\emptyset \neq)\{s\}[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{V}] \subseteq \{s\}^-_{\mathbf{X}=\mathbf{x}}$, the set of possible assignments resulting from assigning $\mathbf{x}$ to $\mathbf{X}$ in s by intervention. Since $M_{T_{\mathbf{X}=\mathbf{x}}}$ is the reduct of M_T to the vocabulary of $tr^d(\psi, G_{\mathbf{X}=\mathbf{x}})$, we also have $M_{T_{\mathbf{X}=\mathbf{x}}}, \{s\}[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{V}] \models tr^d(\psi, G_{\mathbf{X}=\mathbf{x}})$. Thus, by inductive assumption, for the single assignment s' in $\{s\}[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{V}]$, $(\{s'\}, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \not\models \psi$. Thus in particular there exists an $s' \in \{s\}^-_{\mathbf{X}=\mathbf{x}}$ such that $(\{s'\}, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \not\models \psi$. Thus $(\{s\}, G, \mathcal{F}) \not\models \mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi$, as required.

Conversely, suppose that for all $s \in T^-$, $(\{s\}, G, \mathcal{F}) \not\models \mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi$, where $\psi \in \mathcal{CO}$. By definition, this means that $(\{s\}^-_{\mathbf{X}=\mathbf{x}}, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \not\models \psi$, that is, by the flatness of $\mathcal{CO}$, there exists at least one $s' \in \{s\}_{\mathbf{X}=\mathbf{x}}$ such that $(\{s'\}, G, \mathcal{F}) \not\models \psi$. Now let the tuple of choice functions $\mathbf{F}$ pick $\mathbf{F}(s[\mathbf{x}/\mathbf{X}]) = s'(\mathbf{V})$ for any such s, so that $s[\mathbf{x}/\mathbf{X}][\mathbf{F}(s[\mathbf{x}/\mathbf{X}])/\mathbf{V}] = s'$ and that $T^-[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{V}] = \{s' : s \in T^-\}$. Notice that $\mathbf{F}$ is well-defined, because if $s[\mathbf{x}/\mathbf{X}] = t[\mathbf{x}/\mathbf{X}]$ then $\{s\}_{\mathbf{X}=\mathbf{x}} = \{t\}_{\mathbf{X}=\mathbf{x}}$, so that the same s' can be associated to both s and t. Now, by construction, all the assignments in $\{s' : s \in T^-\}$ satisfy $Eq(G_{\mathbf{X}=\mathbf{x}})$ and $\mathbf{X} = \mathbf{x}$, and furthermore they do not satisfy ψ; thus, by induction hypothesis and the usual considerations about reducts, $M_T, T^-[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{V}] \models Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr^d(\psi, G_{\mathbf{X}=\mathbf{x}})$ and the conclusion follows. $\square$

For the purpose of some applications, it might be preferable to translate the causal-observational formulas into statements about a purely relational structure. The transition is not difficult. We need to consider special structures in which the relation symbols are interpreted by functional relations, as follows. To a causal team $T = (T^-, G, \mathcal{F})$ we associate the structure:

$$M^*_T = (|M_T|, (P_c^{M^*_T})_{c \in |M_T|}, (R_V^{M^*_T})_{V \in End(T)})$$

where the unary predicate symbol P_c is interpreted by the set $\{c\}$, and the $card(PA_V) + 1$-ary relation symbol R_V is interpreted by $R_V^{M^*_T} = \{(a_1, \ldots, a_n, b) \mid f_V^{M_T}(a_1, \ldots, a_n) = b\}$, where $f_V^{M_T}$ is as in the definition of M_T. For any spanning subgraph H of G, we can then express the fact that the relations R_V (for $V \in End(H)$) correctly interpret corresponding functions $\mathcal{F}(V)$ by the formula:

$$Eq_*(H) := \bigwedge_{V \in End(H)} R_V(PA_V, V).$$

We can then define a "relational" translation tr_* (and its dual tr^d_*) simply by replacing, in the clauses of tr (resp. tr^d), the subformula $Eq(G_{\mathbf{X}=\mathbf{x}})$ with $Eq_*(G_{\mathbf{X}=\mathbf{x}})$. The following result immediately entails the correctness of this new translation.

Theorem 3.19. *Let $T = (T^-, G, \mathcal{F})$ be a causal team of signature σ and M^*_T as above. Then, for any $\mathcal{COD}(\sigma)$ or $\mathcal{CO}_{\sqcup}(\sigma)$ formula φ, and any spanning subgraph H of G:*

$$M^*_T, T^- \models tr_*(\varphi, H) \iff M_T, T^- \models tr(\varphi, H).$$

In particular (by Theorem 3.18), $M^*_T, T^- \models tr_*(\varphi, G) \iff T \models \varphi$.

Proof. By induction on φ; the nontrivial case is when φ is of the form $\mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi$. The statement follows immediately from the fact that, for each $V \in End(G)$, $M^*_T, T^- \models R_V(PA_V, V)$ iff $M_T, T^- \models V = f_V(PA_V)$, and thus $M^*_T, T^- \models Eq_*(H) \iff M_T, T^- \models Eq(H)$. □

Theorems 3.18 and 3.19 tell us how to embed $\mathcal{CO}$, $\mathcal{COD}$ and $\mathcal{CO}_\sqcup$ into appropriate fragments of $\mathbf{FO}(\exists^1, \sqcup, =\!(\cdot;\cdot))$ and thus, by Theorem 3.15, into dependence logic, for any fixed choice of graph G.

By theorem 6.2 of [31], this implies that these formulas can be translated into existential second-order logic; but for some rather general fragments of these logics, as we will now prove, we can show much more - i.e. that they can be embedded into the Bernays-Schönfinkel-Ramsey fragment of dependence logic, i.e. the set of prenex formulas (of relational vocabulary) with prefix of the form $\exists^*\forall^*$. Some properties of this fragment have been studied in [27]; as its first-order counterpart, it has a decidable satisfaction problem. These embeddings are possible only if we appropriately restrict the use of the selective implication. More precisely, the fragments to which the result applies will be called $\mathcal{CO}^0$, $\mathcal{COD}^0$ and $\mathcal{CO}^0_\sqcup$; they consist of those formulas of $\mathcal{CO}$, resp. $\mathcal{COD}, \mathcal{CO}_\sqcup$ which do not contain subformulas of the form:

- $\mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi$, where ψ contains a selective implication $\alpha \supset \chi$ and α contains another counterfactual.

A simple example of a formula that does *not* belong to these fragments is $\mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow ((\mathbf{Z} = \mathbf{z} \,\Box\!\!\rightarrow W = w) \supset Y = y)$. It is worth pointing out that these fragments still allow for quite expressive uses of the selective implication, such as formulas of the form $(\mathbf{Z} = \mathbf{z} \,\Box\!\!\rightarrow \psi) \supset \chi$. This kind of formula does not simply state what follows from observing the current state of the system; rather, it describes what we can deduce from the outcome of an "experiment" (setting $\mathbf{X}$ to $\mathbf{x}$). Notice also that $\mathcal{CO}^0$, resp. $\mathcal{COD}^0$ and $\mathcal{CO}^0_\sqcup$ properly contain the fragments $\mathcal{CO}^{un}$, $\mathcal{COD}^{un}$ and $\mathcal{CO}^{un}_\sqcup$ of *unnested* formulas of $\mathcal{CO}$, resp. $\mathcal{COD}$ and $\mathcal{CO}_\sqcup$, i.e. formulas in which a counterfactual cannot occur in the consequent of a counterfactual. Many papers in the literature on causation, such as [11] and [18], embrace this restriction; therefore our following result fully covers many commonly used languages for causation.

If C is a class of causal teams, we will say that a causal language $\mathcal{L}$ **C-embeds** into a fragment L of first-order dependence logic if for every formula $\varphi \in \mathcal{L}$ and every graph G over its variables there is a formula $\varphi' \in L$ such that, for all causal teams $T = (T^-, G, \mathcal{F}) \in \mathsf{C}$, $T \models \varphi$ if and only if $M_T, T^- \models \varphi'$. As a special case, if C is the class of all causal teams of the signature of $\mathcal{L}$, we simply say that $\mathcal{L}$ **embeds** into L.

Corollary 3.20. *Let* $\sigma = (Dom, Ran)$ *be a signature with finite* Dom. *Then:*

a) $\mathcal{CO}^0(\sigma)$ *embeds into the Bernays-Schönfinkel-Ramsey fragment of first-order logic.*
b) $\mathcal{COD}^0(\sigma)$ *embeds into the Bernays-Schönfinkel-Ramsey fragment of dependence logic.*
c) $\mathcal{CO}^0_\sqcup(\sigma)$ *embeds into the Bernays-Schönfinkel-Ramsey fragment of dependence logic.*

Proof. Observe that, for $\varphi \in \mathcal{CO}^0, \mathcal{COD}^0$ or $\mathcal{CO}^0_\sqcup$, $tr_*(\varphi, G)$ is a purely relational formula of $\mathbf{FO}(\exists^1, \sqcup, =\!(\cdot;\cdot))$ in which no non-constant existential quantifier $\exists\mathbf{V}$ appears in the scope of any universal quantifier. Thus, by Theorem 3.15, this expression is equivalent to some expression in the Bernays-Schönfinkel-Ramsey fragment of dependence logic.

Now for the case of $\mathcal{CO}^0$, note that this expression will be flat: indeed, since formulas in $\mathcal{CO}^0$ are flat in the sense of causal team semantics $M^*_T, T^- \models tr_*(\varphi, G)$ iff $(T^-, G, \mathcal{F}) \models \varphi$ iff $(\{s\}, G, \mathcal{F}) \models \varphi$ for all $s \in T^-$ iff $M^*_T, \{s\} \models tr_*(\varphi, G)$ for all such s. Additionally, $tr_*(\varphi, G)$ will not contain dependence atoms; and therefore, again by Theorem 3.15, it will be equivalent to some first-order formula in the Bernays-Schönfinkel-Ramsey fragment of first-order logic. □

These embedding results can be extended to the unrestricted causal languages, at the cost of restricting attention to the class of unique-solution causal teams (and we will see later that the restriction to recursive causal teams allows for sharper embedding results).

Calling U_σ the class of unique-solution causal teams of signature σ, we have the following:

Corollary 3.21. *Let $\sigma = (Dom, Ran)$ be a signature with finite Dom. Then:*

a) $\mathcal{CO}(\sigma)$ U_σ-embeds into the Bernays-Schönfinkel-Ramsey fragment of first-order logic.
b) $\mathcal{COD}(\sigma)$ U_σ-embeds into the Bernays-Schönfinkel-Ramsey fragment of dependence logic.
c) $\mathcal{CO}_\sqcup(\sigma)$ U_σ-embeds into the Bernays-Schönfinkel-Ramsey fragment of dependence logic.

Proof. Let φ be any formula in $\mathcal{CO}(\sigma)$. By Theorem 2.15, we can find an unnested formula φ' which is equivalent to it over all unique-solution causal teams. Now notice that φ' is in $\mathcal{CO}^0(\sigma)$: indeed, since it is unnested it is never the case that a counterfactual appears in the consequent of another counterfactual.

Therefore, by point a) of Corollary 3.20 the formula $tr_*(\varphi', G)$ – which is in the Bernays-Schönfinkel-Ramsey fragment of first-order logic – is such that $(T^-, G, \mathcal{F}) \models \varphi'$ if and only if $M_T^*, T^- \models tr_*(\varphi', G)$.

Since for all unique-solution causal teams, $(T^-, G, \mathcal{F}) \models \varphi$ if and only if $(T^-, G, \mathcal{F}) \models \varphi'$, for all such causal teams we have that $(T^-, G, \mathcal{F}) \models \varphi$ if and only if $M_T^*, T^- \models tr_*(\varphi', G)$, as required.

The cases of $\varphi \in \mathcal{COD}(\sigma)$ or $\varphi \in \mathcal{CO}_\sqcup(\sigma)$ are analogous: by Theorem 2.15 we can find a $\varphi' \in \mathcal{COD}^0(\sigma)$ (respectively $\mathcal{CO}^0_\sqcup(\sigma)$) which is equivalent to φ over unique-solution causal teams, and by point b) (respectively c)) of Corollary 3.20 this translates into the Bernays-Schönfinkel-Ramsey fragment of dependence logic. □

4. Alternative semantics

4.1. An alternative definition of intervention

The definition of intervention that we have been following until now is the natural extension of the definition of intervention on causal models proposed by Halpern ([18]) to the more general case of causal teams. While we see no doubt about the correctness of the generalization procedure, Halpern's rule itself seems vulnerable to criticism. The following example should make this clear.

Consider a signature with four variables X, Y, W and Z, all with range $\{0, 1\}$. We then assume that Y, W, Z are endogenous variables, generated by the identity functions $\mathcal{F}(Y)(X) = X$, $\mathcal{F}(W)(Z) = Z$ and $\mathcal{F}(Z)(W) = W$, so that W, Z form a cycle; furthermore, this cycle is completely separated from the rest of the system.[18] We assume that in the current state of the system all variables have value 0. These assumptions are represented by the following causal team:

T:

X→Y		W↔Z	
0	0	0	0

We should expect the intervention $do(X = 0)$ not to alter the system in any way. However, if we evaluate it according to the Halpern-style definition, we obtain a two-row causal team:

$T_{X=0}$:

X→Y		W↔Z	
0	0	0	0
0	0	1	1

[18] The cycle also happens not to causally depend on any exogenous variable; this point is not crucial for our example.

It would seem then that intervening on X has some causal effect on W and Z, *even though there is no directed path from X to W, resp. Z* in the causal graph.[19] Thus, the Halpern-style definition of intervention contradicts a usual assumption of causal inference - that the lack of directed paths among two variables indicates the lack of causal connections between the two. The outcome just described would be reasonable only if the model were an incorrect representation of reality, which omits some causal connections going from the variable X to variables W and Z. Observe also that T has a singleton team component, and therefore it can be identified with a causal model; thus, the antinomy does not arise from our generalized definition for causal teams, but it is already present in Halpern's definition of intervention for causal models.

We want then to propose a more coherent definition of intervention, and see whether the causal-observational languages can still be embedded into dependence logic under this more refined semantics. We have seen that the problem with Halpern's definition lies in the fact that an intervention on $\mathbf{X}$ may produce solutions (to the system of equations) that differ from the initial ones over variables that are not descendants of $\mathbf{X}$. The lack of paths from $\mathbf{X}$ to such variables should entail that $\mathbf{X}$ has no causal import at all for such variables, if we want to take seriously the causal meaning of the graph. It would then seem more natural to us to accept only solutions which do not differ from the initial state(s) on variables (distinct from $\mathbf{X}$) that are not descendants of $\mathbf{X}$. Call $\mathbf{N_X}$ the set of nondescendants of $\mathbf{X}$ that are, furthermore, endogenous in the graph under consideration (note that $\mathbf{X} \cap \mathbf{N_X} = \emptyset$!). As before, we denote by $\mathbf{U}$ the set $Exo(T) \setminus \mathbf{X}$. Using the notations of section 2.4, we redefine:

$$(T^A_{\mathbf{X}=\mathbf{x}})^- := \{s \text{ compatible with } (G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \mid s(\mathbf{X}) = \mathbf{x} \text{ and } s(\mathbf{UN_X}) \in T^-(\mathbf{UN_X})\}.$$

We then define $G^A_{\mathbf{X}=\mathbf{x}} := G_{\mathbf{X}=\mathbf{x}}$, $\mathcal{F}^A_{\mathbf{X}=\mathbf{x}} := \mathcal{F}_{\mathbf{X}=\mathbf{x}}$; the result of the intervention $do(\mathbf{X} = \mathbf{x})$ applied to T is then the causal team $T^A_{\mathbf{X}=\mathbf{x}} = ((T^A_{\mathbf{X}=\mathbf{x}})^-, G^A_{\mathbf{X}=\mathbf{x}}, \mathcal{F}^A_{\mathbf{X}=\mathbf{x}})$. Clearly, with this definition of intervention, in the example above we simply obtain $T^A_{X=0} = T$, as should be expected.

Having modified the definition of intervention, we obtain a new notion of satisfaction $T \models^A \varphi$ by keeping the same semantic clauses as before, with the exception that in the clause for counterfactuals we refer to this alternative definition of intervened causal team. Also the semantic clause for $\supset$ must be modified (in an obvious way):

- $T \models \alpha \supset \psi \iff T^\alpha_A \models \psi$, where T^α_A is the causal subteam of T with team component $\{s \in T^- \mid \{s\} \models^A \alpha\}$.

Let us state some basic properties of this alternative semantics. The first shows that, as before, the intervention on a causal team could in principle be defined on an assignment-by-assignment basis.

Proposition 4.1. *Let $T = (T^-, G, \mathcal{F})$ be a causal team of signature σ and $\mathbf{X} = \mathbf{x}$ be a consistent conjunction of signature σ. Then:*

$$(T^A_{\mathbf{X}=\mathbf{x}})^- = \bigcup_{s \in T^-} (\{s\}^A_{\mathbf{X}=\mathbf{x}})^-.$$

Proof. Write $\mathbf{U}$ for $Exo(T) \setminus \mathbf{X}$ and $\mathbf{N_X}$ for the set of nondescendants of $\mathbf{X}$ that are endogenous in G.

Let $s \in T^-$ and $t \in (\{s\}^A_{\mathbf{X}=\mathbf{x}})^-$. By definition of $\{s\}^A_{\mathbf{X}=\mathbf{x}}$, we have that t is compatible with $(G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})$, $t(\mathbf{X}) = \mathbf{x}$ and $t(\mathbf{UN_X}) \in \{s\}(\mathbf{UN_X}) \subseteq T^-(\mathbf{UN_X})$. Thus $t \in (T^A_{\mathbf{X}=\mathbf{x}})^-$. Therefore $(\{s\}^A_{\mathbf{X}=\mathbf{x}})^- \subseteq T^-_{\mathbf{X}=\mathbf{x}}$.

[19] This immediately affects the truth values of formulas: by the observations above, the Halpern-style semantics entails that $T \not\models [X = 0]Z = 0$, while we should expect that $T_{X=0} = T$ and thus $T \models [X = 0]Z = 0$.

Vice versa, let $t \in (T^A_{\mathbf{X}=\mathbf{x}})^-$. By definition, $t(\mathbf{UN_X}) \in T^-(\mathbf{UN_X})$. This means there is an $s \in T^-$ such that $t(\mathbf{UN_X}) = s(\mathbf{UN_X})$, i.e. $t(\mathbf{UN_X}) \in \{s\}(\mathbf{UN_X})$. Furthermore, the assumption that $t \in (T^A_{\mathbf{X}=\mathbf{x}})^-$ gives us that $t(\mathbf{X}) = \mathbf{x}$ and that t is compatible with $(G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})$. Therefore $t \in \{s\}^-_{\mathbf{X}=\mathbf{x}}$. □

Proposition 4.2. *Let S, T be causal teams of signature σ.*

- ***Empty team property:*** *if $\varphi \in \mathcal{COD}(\sigma) \cup \mathcal{CO}_{\sqcup}(\sigma)$ and $T = (\emptyset, G, \mathcal{F})$, then $T \models^A \varphi$.*
- ***Flatness:*** *if $\varphi \in \mathcal{CO}(\sigma)$, then $T \models^A \varphi$ iff, for all $s \in T^-$, $\{s\} \models^A \varphi$.*
- ***Downward closure:*** *if $\varphi \in \mathcal{COD}(\sigma) \cup \mathcal{CO}_{\sqcup}(\sigma)$, $T \models^A \varphi$ and $S \leq T$, then $S \models^A \varphi$.*

The proofs are analogous to those for $\models^H$; in the case of downward closure, one needs to use the easily provable fact that, if S is a causal subteam of T, then $S^A_{\mathbf{X}=\mathbf{x}}$ is a causal subteam of $T^A_{\mathbf{X}=\mathbf{x}}$.

On a negative side, the alternative semantics shares with the Halpern-style semantics the failure of some *Distribution* properties of the connectives that hold in the recursive case.

Example 4.3. Consider the causal team T depicted below (together with the result of the intervention $do(Z = 0)$):

T:

Z→X↔Y		
0	0	0

$\rightsquigarrow$ $T_{Z=0}$:

Z→X↔Y		
0	0	0
0	1	1

with a single assignment s, where $X := \max(\mathrm{Z}, \mathrm{Y})$ and $Y := X$. Now $T \models \neg(Z = 0 \,\Box\!\!\rightarrow X = 0)$, but $T \not\models (Z = 0 \,\Box\!\!\rightarrow X = 0)^d$ i.e. $T \not\models Z = 0 \,\Box\!\!\rightarrow X \neq 0$. Thus, the dualization procedure for the recursive case is not correct in general, even under the alternative semantics.

Furthermore, $T \models Z = 0 \,\Box\!\!\rightarrow (X = 0 \vee X = 1)$ but $T \not\models (Z = 0 \,\Box\!\!\rightarrow X = 0) \vee (Z = 0 \,\Box\!\!\rightarrow X = 1)$; the distribution rule for $\Box\!\!\rightarrow$ over $\vee$ fails.

Finally, observe that $T \models (Z = 0 \,\Box\!\!\rightarrow X = 0) \supset (Z = 0 \,\Box\!\!\rightarrow X = 1)$ (trivially, since the antecedent is not satisfied by s), while $T \not\models Z = 0 \,\Box\!\!\rightarrow (X = 0 \supset X = 1)$. Thus also the distributivity of $\Box\!\!\rightarrow$ over $\supset$ is falsified.

We can say that a causal team is ***A*-unique-solution** if the definition of "unique-solution" holds when we replace the Halpern-style definition of intervention with the new one. An analogue of Lemma 2.10 can then be proved (by the same method), showing in particular that the *Distribution* rules do hold in the A-unique-solution case.

Concerning the *Overwriting* rule (the equivalence of $\mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow (\mathbf{Y} = \mathbf{y} \,\Box\!\!\rightarrow \psi)$ and $(\mathbf{X}' = \mathbf{x}' \wedge \mathbf{Y} = \mathbf{y}) \,\Box\!\!\rightarrow \psi$, under the assumption of the consistency of $\mathbf{X} = \mathbf{x}$), we saw that, with the H-semantics, it is valid on teams in which every assignment s satisfies $\{s\}^-_{\mathbf{X}=\mathbf{x}} \neq \emptyset$; the following counterexample shows that this condition does not suffice with the A-semantics.

Example 4.4. Consider a causal team T of four variables, with $Ran(X) = Ran(Y) = \{0, 1\}$ and $Ran(V) = Ran(W) = \mathbb{Z}$, satisfying $V := X - W$ and $W := Y - V$. Thus X, Y are exogenous, while V, W are endogenous and form a 2-cycle. It is easy to see that the solutions to this system of equations are all and only the assignments of the form $t(X) = t(Y) = k, t(V) = n, t(W) = k - n$ $(k, n \in \mathbb{Z})$. Let $T^- = \{s\}$, where $s(X) = s(V) = s(W) = s(Y) = 0$. Now, intervening according to the Halpern-style definition, we obtain $T^-_{X=1} = \{s\}^-_{X=1} \neq \emptyset$: this team contains all assignments of the form $t(X) = t(Y) = 1, t(V) = n, t(W) = 1 - n$ $(n \in \mathbb{Z})$. On the other hand, $(T^A_{X=1})^- = (\{s\}^A_{X=1})^- = \emptyset$: Y is not a descendant of X, and therefore it keeps its value 0; but there is no solution t to the system having both $t(X) = 1$ and $t(Y) = 0$. Notice also that $(T^A_{X=1 \wedge Y=1})^- = T^-_{X=1}$. Therefore, $T \models^A X = 1 \,\Box\!\!\rightarrow (Y = 1 \,\Box\!\!\rightarrow \bot)$ (by the empty team property), while $T \not\models^A (X = 1 \wedge Y = 1) \,\Box\!\!\rightarrow \bot$.

Let us say instead that a causal team $T = (T^-, G, \mathcal{F})$ of signature $\sigma = (Dom, Ran)$ is **A-solutionful** if, for every assignment s of signature σ, and every consistent conjunction $\mathbf{X} = \mathbf{x}$ of signature σ, $((\{s\}, G, \mathcal{F})^A_{\mathbf{X}=\mathbf{x}})^- \neq \emptyset$. This assumption suffices to validate the *Overwriting* rule.

Theorem 4.5. *Let $\sigma = (Dom, Ran)$ be a signature, $\mathbf{X} \subseteq Dom$ be distinct variables and $\mathbf{x} \in Ran(\mathbf{X})$ (therefore $\mathbf{X} = \mathbf{x}$ is a consistent conjunction). Let $\mathbf{Y} \subseteq Dom$, $\mathbf{y} \in Ran(\mathbf{Y})$ and ψ be a $\mathcal{COD}$ or $\mathcal{CO}_{\sqcup}$ formula of signature σ. Then, for every A-solutionful causal team T of signature σ:*

$$T \models^A \mathbf{X} = \mathbf{x} \;\Box\!\!\rightarrow (\mathbf{Y} = \mathbf{y} \;\Box\!\!\rightarrow \psi) \iff T \models^A (\mathbf{X}' = \mathbf{x}' \wedge \mathbf{Y} = \mathbf{y}) \;\Box\!\!\rightarrow \psi$$

where $\mathbf{X}' = \mathbf{X} \setminus \mathbf{Y}$ and $\mathbf{x}' = \mathbf{x} \setminus \mathbf{y}$.

Proof. Let $I^{G_{\mathbf{X}=\mathbf{x}}}_{\mathbf{Y}}$ and $I^G_{\mathbf{X}'\mathbf{Y}}$ be the sets of non-descendants of $\mathbf{Y}$ in graph $G_{\mathbf{X}=\mathbf{x}}$ and of $\mathbf{X}'\mathbf{Y}$ in graph G, and let us write

$$\begin{aligned} A &:= ((T^A_{\mathbf{X}=\mathbf{x}})^A_{\mathbf{Y}=\mathbf{y}})^- \\ &= \{s \mid s(\mathbf{Y}) = \mathbf{y}, s \text{ comp. with } ((G_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}}, (\mathcal{F}_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}}) \text{ and } s(I^{G_{\mathbf{X}=\mathbf{x}}}_{\mathbf{Y}}) \in (T^A_{\mathbf{X}=\mathbf{x}})^-(I^{G_{\mathbf{X}=\mathbf{x}}}_{\mathbf{Y}})\} \\ B &:= (T^A_{\mathbf{X}'=\mathbf{x}' \wedge \mathbf{Y}=\mathbf{y}})^- \\ &= \{s \mid s(\mathbf{X}'\mathbf{Y}) = \mathbf{x}'\mathbf{y}, s \text{ comp. with } (G_{\mathbf{X}'=\mathbf{x}' \wedge \mathbf{Y}=\mathbf{y}}, \mathcal{F}_{\mathbf{X}'=\mathbf{x}' \wedge \mathbf{Y}=\mathbf{y}}) \text{ and } s(I^G_{\mathbf{X}'\mathbf{Y}}) \in T^-(I^G_{\mathbf{X}'\mathbf{Y}})\}. \end{aligned}$$

Proving $A = B$ leads easily to the conclusion.

- $A \subseteq B$) Assume $s \in A$. Then there exists some $s_1 \in (T^A_{\mathbf{X}=\mathbf{x}})^-$ such that $s(I^{G_{\mathbf{X}=\mathbf{x}}}_{\mathbf{Y}}) = s_1(I^{G_{\mathbf{X}=\mathbf{x}}}_{\mathbf{Y}})$.
  Furthermore, since $s_1 \in (T^A_{\mathbf{X}=\mathbf{x}})^-$, there exists some $s_0 \in T^-$ such that $s_1(I^G_{\mathbf{X}}) = s_0(I^G_{\mathbf{X}})$ and s_1 is compatible with $\mathcal{F}_{\mathbf{X}=\mathbf{x}}$.
  Now observe that $I^G_{\mathbf{X}'\mathbf{Y}} \subseteq I^{G_{\mathbf{X}=\mathbf{x}}}_{\mathbf{Y}} \cap I^G_{\mathbf{X}}$: indeed, if there existed a causal path from variables in $\mathbf{Y}$ to V in $G_{\mathbf{X}=\mathbf{x}}$ or from variables in $\mathbf{X}$ to V in G, the same path would go from variables in $\mathbf{X}'\mathbf{Y}$ to V in G.
  Therefore, $s(I^G_{\mathbf{X}'\mathbf{Y}}) = s_1(I^G_{\mathbf{X}'\mathbf{Y}}) = s_0(I^G_{\mathbf{X}'\mathbf{Y}}) \in T^-(I^G_{\mathbf{X}'\mathbf{Y}})$; $s(\mathbf{Y}) = \mathbf{y}$; since $\mathbf{X}' \subseteq I^{G_{\mathbf{X}=\mathbf{x}}}_{\mathbf{Y}}$, $s(\mathbf{X}') = s_1(\mathbf{X}') = \mathbf{x}'$; and finally s is compatible with $(\mathcal{F}_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}} = \mathcal{F}_{\mathbf{X}'=\mathbf{x}' \wedge \mathbf{Y}=\mathbf{y}}$ and the corresponding graph. Thus $s \in B$, as required.
- $B \subseteq A$) Assume $s \in B$. Then there exists some $s_0 \in T^-$ such that $s(I^G_{\mathbf{X}'\mathbf{Y}}) = s_0(I^G_{\mathbf{X}'\mathbf{Y}})$.
  Now let $\mathbf{D}$ be the set of variables that strictly descend from $\mathbf{X}$ *but do not descend from* $\mathbf{Y}$ in $G_{\mathbf{X}=\mathbf{x}}$, and let $\mathbf{d} = s(\mathbf{D})$.
  Write $\mathbf{x}$ for $s(\mathbf{X})$; the conjunction $\mathbf{X} = \mathbf{x} \wedge \mathbf{D} = \mathbf{d}$ is consistent, since $s(\mathbf{X}) = \mathbf{x}$ and $s(\mathbf{D}) = \mathbf{d}$. Because T is A-solutionful, there exists some $s_1 \in (\{s_0\}^A_{\mathbf{X}=\mathbf{x} \wedge \mathbf{D}=\mathbf{d}})^-$.
  We state that $s_1 \in (\{s_0\}^A_{\mathbf{X}=\mathbf{x}})^-$ as well. Indeed,

  1. $s_1(\mathbf{X}) = \mathbf{x}$;
  2. s_1 is compatible with $(G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})$.
     Indeed, s_1 is compatible with $\mathcal{F}_{\mathbf{X}=\mathbf{x} \wedge \mathbf{D}=\mathbf{d}}$; thus for all variables $V \notin \mathbf{X} \cup \mathbf{D}$ we have that

     $$\mathcal{F}_{\mathbf{X}=\mathbf{x} \wedge \mathbf{D}=\mathbf{d}}(V) = \mathcal{F}_{\mathbf{X}=\mathbf{x}}(V);$$

     furthermore,

     $$PA_V^{G_{\mathbf{X}=\mathbf{x} \wedge \mathbf{D}=\mathbf{d}}} = PA_V^{G_{\mathbf{X}=\mathbf{x}}}$$

and thus $s_1(V) = \mathcal{F}_{\mathbf{X}=\mathbf{x}}(V)(s_1(PA_V^{G_{\mathbf{X}=\mathbf{x}}}))$.
If $V \in \mathbf{X}$, there is nothing to show because the variables in $\mathbf{X}$ are exogenous in $G_{\mathbf{X}=\mathbf{x}}$. For $V \in \mathbf{D}$ we have instead that

$$s_1(V) = s(V) = \mathcal{F}_{\mathbf{X}'=\mathbf{x}'\wedge\mathbf{Y}=\mathbf{y}}(V)(s(PA_V^{G_{\mathbf{X}'=\mathbf{x}'\wedge\mathbf{Y}=\mathbf{y}}})) = \mathcal{F}_{\mathbf{X}=\mathbf{x}}(V)(s(PA_V^{G_{\mathbf{X}=\mathbf{x}}}))$$

where we used the fact that, since $V \notin \mathbf{Y}$, $PA_V^{G_{\mathbf{X}'=\mathbf{x}'\wedge\mathbf{Y}=\mathbf{y}}} = PA_V^{G_{\mathbf{X}=\mathbf{x}}}$ and $\mathcal{F}_{\mathbf{X}'=\mathbf{x}'\wedge\mathbf{Y}=\mathbf{y}}(V) = \mathcal{F}_{\mathbf{X}=\mathbf{x}}(V)$.
If we can prove that s_1 and s agree over the parents of V, we are done, because then

$$s_1(V) = \mathcal{F}_{\mathbf{X}=\mathbf{x}}(V)(s(PA_V^{G_{\mathbf{X}=\mathbf{x}}})) = \mathcal{F}_{\mathbf{X}=\mathbf{x}}(V)(s_1(PA_V^{G_{\mathbf{X}=\mathbf{x}}}))$$

as required.
Let then W be a parent of V: since $V \in \mathbf{D}$ (i.e., in $G_{\mathbf{X}=\mathbf{x}}$, V strictly descends from $\mathbf{X}$ but is not in $\mathbf{Y}$ nor it is a descendant of $\mathbf{Y}$) its parents do not descend from $\mathbf{Y}$, and so they must either descend from $\mathbf{X}$ but not from $\mathbf{Y}$ (i.e. they must be in $\mathbf{X}$ or in $\mathbf{D}$) or from neither $\mathbf{X}$ nor $\mathbf{Y}$ (i.e. they must be in $I^G_{\mathbf{X}'\mathbf{Y}}$).
In the first case, $s_1(W) = s(W)$ since $s_1(\mathbf{X}) = \mathbf{x} = s(\mathbf{X})$ and $s_1(\mathbf{D}) = \mathbf{d} = s(\mathbf{D})$; and in the second case, W in particular is a non-descendant of $\mathbf{X}$ and of $\mathbf{D}$ (since $\mathbf{D}$ contains only descendants of $\mathbf{X}$) and so $s_1(W) = s_0(W)$, and moreover since s_0 and s agree over the non-descendants of $\mathbf{X}'\mathbf{Y}$ we have that $s_0(W) = s(W)$.
3. For all $V \in I^G_{\mathbf{X}}$, we need to prove that $s_1(V) = s_0(V)$. Now, by definition if $V \in I^G_{\mathbf{X}}$ there is no path from $\mathbf{X}$ to V in G. But then there is no path from $\mathbf{D}$ to V either, since $\mathbf{D}$ contains only descendants of $\mathbf{X}$, and so $V \in I^G_{\mathbf{X}\cup\mathbf{D}}$ and $s_1(V) = s_0(V)$ as required.

It remains only to show that $s \in (\{s_1\}^A_{\mathbf{Y}=\mathbf{y}})^- \subseteq ((T^A_{\mathbf{X}=\mathbf{x}})^A_{\mathbf{Y}=\mathbf{y}})^-$. This is indeed the case:

1. $s(\mathbf{Y}) = \mathbf{y}$, as required.
2. We already know that s is compatible with $(\mathcal{F}_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}} = \mathcal{F}_{\mathbf{X}'=\mathbf{x}'\wedge\mathbf{Y}=\mathbf{y}}$ and the corresponding graph.
3. Let $V \in I_{\mathbf{Y}}^{G_{\mathbf{X}=\mathbf{x}}}$ be a non-descendant of $\mathbf{Y}$ in $G_{\mathbf{X}=\mathbf{x}}$. We need to prove that $s(V) = s_1(V)$.
If V descends from $\mathbf{X}$ in $G_{\mathbf{X}=\mathbf{x}}$ then $V \in \mathbf{X} \cup \mathbf{D}$, and so by definition $s_1(V) = s(V)$.
If V descends from neither $\mathbf{X}$ nor $\mathbf{Y}$ in $G_{\mathbf{X}=\mathbf{x}}$ then it descends from neither $\mathbf{X}$ nor $\mathbf{Y}$ in G as well: indeed, $G_{\mathbf{X}=\mathbf{x}}$ differs from G only in that $\mathbf{X}$ is now exogenous, and so the only paths that are available in G but not in $G_{\mathbf{X}=\mathbf{x}}$ pass through $\mathbf{X}$.
Therefore $V \in I^G_{\mathbf{X}'\mathbf{Y}} \subseteq I^G_{\mathbf{X}}$ and $s(V) = s_0(V)$. Moreover, since $s_1 \in (\{s_0\}^A_{\mathbf{X}=\mathbf{x}})^-$ we have that $s_0(V) = s_1(V)$. Thus,

$$s(V) = s_0(V) = s_1(V).$$

Therefore we have that $s(I_{\mathbf{Y}}^{G_{\mathbf{X}=\mathbf{x}}}) = s_1(I_{\mathbf{Y}}^{G_{\mathbf{X}=\mathbf{x}}})$, as required.

Thus, we showed that there exists some $s_0 \in T^-$ and some $s_1 \in (\{s_0\}^A_{\mathbf{X}=\mathbf{x}})^- \subseteq (T^A_{\mathbf{X}=\mathbf{x}})^-$ such that $s \in (\{s_1\}^A_{\mathbf{Y}=\mathbf{y}})^-$; therefore, s is in $((T^A_{\mathbf{X}=\mathbf{x}})^A_{\mathbf{Y}=\mathbf{y}})^-$, as required. □

Being A-unique-solution entails being A-solutionful. The observations above, then, allow us to prove that, over A-unique-solution causal teams, every formula is equivalent to an unnested one under the A-semantics, exactly as was done in section 2.7.

Notice that the new definition of intervention produces a causal subteam of what is produced by the corresponding Halpern-style intervention. We should then expect formulas to be easier to satisfy in the new semantics, as long as counterfactuals do not occur in antecedents of $\supset$. The following results prove and generalize this observation. Remember the definition of positive and negative occurrence from section 2.7.[20]

Lemma 4.6. *Let $\alpha \in \mathcal{CO}(\sigma)$, and T a causal team of signature σ. Then:*

1. *If all occurrences of $\Box\!\!\rightarrow$ in α are positive, then $T^\alpha \leq T^\alpha_A$.*
2. *If all occurrences of $\Box\!\!\rightarrow$ in α are negative, then $T^\alpha_A \leq T^\alpha$.*

Proof. We prove 1. and 2. simultaneously, for all causal teams T of signature σ, by induction on the syntax of α.

- It is straightforward to see that, if α is an atom, then $T^\alpha = T^\alpha_A$. So α satisfies both 1. and 2.
- Suppose $\alpha = \beta \wedge \gamma$ only has positive occurrences of $\Box\!\!\rightarrow$. Then the same holds for β and γ. Thus, by inductive hypothesis, $T^\beta \leq T^\beta_A$ and $T^\gamma \leq T^\gamma_A$. Thus $(T^\alpha)^- = (T^\beta)^- \cap (T^\gamma)^- \subseteq (T^\beta_A)^- \cap (T^\gamma_A)^- = (T^\alpha_A)^-$. The cases for α with negative occurrences of $\Box\!\!\rightarrow$ and for $\alpha = \beta \vee \gamma$ are analogous.
- Suppose $\alpha = \beta \supset \gamma$ only has positive occurrences of $\Box\!\!\rightarrow$. Then the same holds for γ, and thus by inductive assumption $T^\gamma \leq T^\gamma_A$. Instead, β only has negative occurrences of $\Box\!\!\rightarrow$, and thus the inductive assumption yields $T^\beta_A \leq T^\beta$. From $T^\gamma \leq T^\gamma_A$ we obtain $T^- \setminus (T^\gamma_A)^- \subseteq T^- \setminus (T^\gamma)^-$. This, together with $T^\beta_A \leq T^\beta$, yields $(T^\beta_A)^- \cap (T^- \setminus (T^\gamma_A)^-) \subseteq (T^\beta)^- \cap (T^- \setminus (T^\gamma)^-)$. Thus $(T^\alpha)^- = T^- \setminus ((T^\beta)^- \cap (T^- \setminus (T^\gamma)^-)) \subseteq T^- \setminus ((T^\beta_A)^- \cap (T^- \setminus (T^\gamma_A)^-)) = (T^\alpha_A)^-$.
  Suppose instead that $\alpha = \beta \supset \gamma$ only has negative occurrences of $\Box\!\!\rightarrow$. Then also γ is such, and then (i.h.) $T^\gamma_A \leq T^\gamma$, from which we obtain $T^- \setminus (T^\gamma)^- \subseteq T^- \setminus (T^\gamma_A)^-$. Instead, β only has positive occurrences of $\Box\!\!\rightarrow$, and the inductive hypothesis yields $T^\beta \leq T^\beta_A$. We then have $(T^\beta)^- \cap (T^- \setminus (T^\gamma)^-) \subseteq (T^\beta_A)^- \cap (T^- \setminus (T^\gamma_A)^-)$. Thus $(T^\alpha_A)^- = T^- \setminus ((T^\beta_A)^- \cap (T^- \setminus (T^\gamma_A)^-)) \subseteq T^- \setminus ((T^\beta)^- \cap (T^- \setminus (T^\gamma)^-)) = (T^\alpha)^-$, as required.
- Suppose $\alpha = \mathbf{X} = \mathbf{x} \Box\!\!\rightarrow \gamma$. Notice that here the assumption of 2. (that all occurrences of $\Box\!\!\rightarrow$ are negative) is false, so we have nothing to prove. Let us assume instead that γ only has positive occurrences of $\Box\!\!\rightarrow$. Then, by induction hypothesis, $S^\gamma \leq S^\gamma_A$ for any causal team S of signature σ; in particular, $(T_{\mathbf{X}=\mathbf{x}})^\gamma \leq (T_{\mathbf{X}=\mathbf{x}})^\gamma_A$. Then $(T^\alpha)^- = \{s \in T^- \mid \{s\}_{\mathbf{X}=\mathbf{x}} \models \gamma\} = \{s \in T^- \mid \{s\}^-_{\mathbf{X}=\mathbf{x}} \subseteq ((T_{\mathbf{X}=\mathbf{x}})^\gamma)^-\} \subseteq \{s \in T^- \mid \{s\}^-_{\mathbf{X}=\mathbf{x}} \subseteq ((T_{\mathbf{X}=\mathbf{x}})^\gamma_A)^-\} \subseteq \{s \in T^- \mid (\{s\}^A_{\mathbf{X}=\mathbf{x}})^- \subseteq ((T_{\mathbf{X}=\mathbf{x}})^\gamma_A)^-\} = \{s \in T^- \mid \{s\}^A_{\mathbf{X}=\mathbf{x}} \models^A \gamma\} = (T^\alpha_A)^-$, where the second inclusion holds because, by the two definitions of intervention, $\{s\}^A_{\mathbf{X}=\mathbf{x}} \subseteq \{s\}_{\mathbf{X}=\mathbf{x}}$ for any assignment s. □

Theorem 4.7. *Let $\varphi \in \mathcal{COD}(\sigma) \cup \mathcal{CO}_\sqcup(\sigma)$ and T a team of signature σ. Then:*

1. *If all occurrences of $\Box\!\!\rightarrow$ in φ are positive, then $T \models^H \varphi \Rightarrow T \models^A \varphi$.*
2. *If all occurrences of $\Box\!\!\rightarrow$ in φ are negative, then $T \models^A \varphi \Rightarrow T \models^H \varphi$.*

Proof. By induction on φ. The nontrivial cases are those for $\Box\!\!\rightarrow$ and $\supset$.

- Case φ is $\mathbf{X} = \mathbf{x} \Box\!\!\rightarrow \psi$. Assume that it contains only positive occurrences of $\Box\!\!\rightarrow$. If $T \models^H \varphi$, then $T_{\mathbf{X}=\mathbf{x}} \models^H \psi$. Now all occurrences of $\Box\!\!\rightarrow$ in ψ are positive; so, by inductive hypothesis, $T_{\mathbf{X}=\mathbf{x}} \models^A \psi$.

[20] The definition of positive and negative occurrences was given for subformulas, but it can be adapted in an obvious way to other tokens (in this case, the occurrences of $\Box\!\!\rightarrow$).

Since $T^A_{\mathbf{X}=\mathbf{x}}$ is a causal subteam of $T_{\mathbf{X}=\mathbf{x}}$, by downward closure of $\models^A$ (Proposition 4.2), $T^A_{\mathbf{X}=\mathbf{x}} \models^A \psi$. Thus $T \models^A \varphi$.

On the other hand, notice that such a formula cannot have only negative occurrences of $\boxright$.

- Case φ is $\alpha \supset \psi$. Assume that it contains only positive occurrences of $\boxright$. If $T \models^H \varphi$, $T^\alpha \models^H \psi$ (where T^α is the causal subteam of T with team component $\{s \in T^- \mid \{s\} \models^H \alpha\}$). Now all occurrences of $\boxright$ in ψ are positive; so, by inductive hypothesis, $T^\alpha \models^A \psi$. Let us write T^α_A, instead, for the causal subteam of T with team component $\{s \in T^- \mid \{s\} \models^A \alpha\}$. Now observe that all occurrences of $\boxright$ in α are negative. Thus, by Lemma 4.6, 2., $T^\alpha_A \subseteq T^\alpha$. So, by downward closure, $T^\alpha_A \models^A \psi$. Thus $T \models^A \alpha \supset \psi$.
  Assume instead that φ contains only negative occurrences of $\boxright$. If $T \models^A \varphi$, $T^\alpha_A \models^A \psi$. Now all occurrences of $\boxright$ in ψ are negative; so, by inductive hypothesis, $T^\alpha_A \models^H \psi$. Now observe that all occurrences of $\boxright$ in α are positive. Thus, by Lemma 4.6, 1., $T^\alpha \subseteq T^\alpha_A$. So, by downward closure, $T^\alpha \models^H \psi$. Thus $T \models^H \alpha \supset \psi$. $\square$

Let us now see how the translation of counterfactuals into dependence logic should be modified when the counterfactuals are interpreted under this alternative semantics. Given a fixed graph G, we write $\mathbf{D}_\mathbf{X}$ for the set of strict descendants of $\mathbf{X}$ in G; notice that this is a set of endogenous variables. We replace the translation clauses for counterfactuals with consistent antecedent with the following:

- $tr_A(\mathbf{X} = \mathbf{x} \boxright \psi, G) := \exists^1 \mathbf{X}(\mathbf{X} = \mathbf{x} \wedge \forall \mathbf{D}_\mathbf{X}(Eq(G_{\mathbf{X}=\mathbf{x}})^d \vee tr_A(\psi, G_{\mathbf{X}=\mathbf{x}})))$
- $tr^d_A(\mathbf{X} = \mathbf{x} \boxright \psi, G) := \exists^1 \mathbf{X} \exists \mathbf{D}_\mathbf{X}(Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr^d_A(\psi, G_{\mathbf{X}=\mathbf{x}}))$

and maintain the other clauses as in $tr(\varphi)$.

We now prove the correctness of this translation.

Theorem 4.8. *Let $T = (T^-, G, \mathcal{F})$ be a causal team and φ a formula of $\mathcal{COD}$ or $\mathcal{CO}_\sqcup$. Then:*

$$M_T, T^- \models tr_A(\varphi, G) \iff T \models^A \varphi.$$

Furthermore, if φ is in $\mathcal{CO}$,

$$M_T, T^- \models tr^d_A(\varphi, G) \iff \forall s \in T^-, (\{s\}, G, \mathcal{F}) \not\models^A \varphi.$$

Proof. The proof is by induction on φ along the lines of the proof of 3.18. We describe the only case that differs significantly, the case for $\boxright$.

Suppose that $M_T, T^- \models \exists^1 \mathbf{X}(\mathbf{X} = \mathbf{x} \wedge \forall \mathbf{D}_\mathbf{X}(Eq(G_{\mathbf{X}=\mathbf{x}})^d \vee tr_A(\psi, G_{\mathbf{X}=\mathbf{x}})))$. Then $T^-[\mathbf{x}/\mathbf{X}][\mathbf{M}/\mathbf{D}_\mathbf{X}]$ can be split into two subteams T^-_1 and T^-_2 such that $M_T, T^-_1 \models Eq(G_{\mathbf{X}=\mathbf{x}})^d$ and $M_T, T^-_2 \models tr_A(\psi, G_{\mathbf{X}=\mathbf{x}})$. As in Theorem 3.18, because of the flatness of $Eq(G_{\mathbf{X}=\mathbf{x}})^d$ and of the downwards closure property we can assume that T^-_2 satisfies $Eq(G_{\mathbf{X}=\mathbf{x}})$, and therefore – since it is the set of all assignments obtained from assignments in T^- by fixing the value of $\mathbf{X}$ to $\mathbf{x}$ and letting the descendants of $\mathbf{X}$ take values compatible with the equations – that it is exactly $(T^A_{\mathbf{X}=\mathbf{x}})^-$.

Then, since $M_{T^A_{\mathbf{X}=\mathbf{x}}}$ is the reduct of M_T to the vocabulary of $tr_A(\psi, G_{\mathbf{X}=\mathbf{x}})$ and $M_T, (T^A_{\mathbf{X}=\mathbf{x}})^- \models tr_A(\psi, G_{\mathbf{X}=\mathbf{x}})$, by induction hypothesis we have $((T^A_{\mathbf{X}=\mathbf{x}})^-, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \models^A \psi$. This implies that $(T^-, G, \mathcal{F}) \models^A \mathbf{X} = \mathbf{x} \boxright \psi$, as required.

Conversely, suppose that $(T^-, G, \mathcal{F}) \models^A \mathbf{X} = \mathbf{x} \boxright \psi$. Then we have that $((T^A_{\mathbf{X}=\mathbf{x}})^-, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \models^A \psi$, and thus – by induction hypothesis – $M_{T^A_{\mathbf{X}=\mathbf{x}}}, (T^A_{\mathbf{X}=\mathbf{x}})^- \models tr_A(\psi, G_{\mathbf{X}=\mathbf{x}})$. Then, since $M_{T^A_{\mathbf{X}=\mathbf{x}}}$ is a reduct of M_T, we have $M_T, (T^A_{\mathbf{X}=\mathbf{x}})^- \models tr_A(\psi, G_{\mathbf{X}=\mathbf{x}})$. Additionally, $(T^A_{\mathbf{X}=\mathbf{x}})^- \subseteq T^-[\mathbf{x}/\mathbf{X}][\mathbf{M}/\mathbf{D}_\mathbf{X}]$, and all assignments in $T^-_0 = T^-[\mathbf{x}/\mathbf{X}][\mathbf{M}/\mathbf{D}_\mathbf{X}] \setminus (T^A_{\mathbf{X}=\mathbf{x}})^-$ violate some structural equation of $G_{\mathbf{X}=\mathbf{x}}$ (and therefore satisfy $Eq(G_{\mathbf{X}=\mathbf{x}})^d$). Thus, $M_T, T_0^- \models Eq(G_{\mathbf{X}=\mathbf{x}})^d$ and $M_T, T^-[\mathbf{x}/\mathbf{X}] \models \mathbf{X} = \mathbf{x} \wedge \forall \mathbf{D_X}(Eq(G_{\mathbf{X}=\mathbf{x}})^d \vee tr_A(\psi, G_{\mathbf{X}=\mathbf{x}}))$. Thus, $M_T, T^- \models \exists^1 \mathbf{X}(\mathbf{X} = \mathbf{x} \wedge \forall \mathbf{D_X}(Eq(G_{\mathbf{X}=\mathbf{x}})^d \vee tr_A(\psi, G_{\mathbf{X}=\mathbf{x}})))$.

Now suppose that $M_T, T^- \models \exists^1 \mathbf{X} \exists \mathbf{D_X}(Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr_A^d(\psi, G_{\mathbf{X}=\mathbf{x}}))$. By downward closure, $M_T, \{s\} \models \exists^1 \mathbf{X} \exists \mathbf{D_X}(Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr_A^d(\psi, G_{\mathbf{X}=\mathbf{x}}))$ for each $s \in T^-$. Then there are functions $\mathbf{F}$ such that $M_T, \{s\}[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{D_X}] \models Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr_A^d(\psi, G_{\mathbf{X}=\mathbf{x}})$. From the first two conjuncts, we infer that $(\emptyset \neq)\{s\}[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{D_X}] \subseteq (\{s\}^A_{\mathbf{X}=\mathbf{x}})^-$, the set of possible assignments resulting from assigning $\mathbf{x}$ to $\mathbf{X}$ in s by intervention. Write M' for $M_{(\{s\}[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{D_X}], G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})}$. Since M' is the reduct of M_T to the vocabulary of $tr_A^d(\psi, G_{\mathbf{X}=\mathbf{x}})$, we also have $M', \{s\}[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{D_X}] \models tr_A^d(\psi, G_{\mathbf{X}=\mathbf{x}})$. Thus, by inductive assumption, for all assignments s' in $\{s\}[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{D_X}]$, $(\{s'\}, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \not\models^A \psi$. Thus in particular there exists an $s' \in (\{s\}^A_{\mathbf{X}=\mathbf{x}})^-$ such that $(\{s'\}, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \not\models^A \psi$. Thus $(\{s\}, G, \mathcal{F}) \not\models^A \mathbf{X} = \mathbf{x} \;\Box\!\!\rightarrow \psi$, as required.

Conversely, suppose that for all $s \in T^-$, $(\{s\}, G, \mathcal{F}) \not\models^A \mathbf{X} = \mathbf{x} \;\Box\!\!\rightarrow \psi$, where $\psi \in \mathcal{CO}$. By definition, this means that $((\{s\}^A_{\mathbf{X}=\mathbf{x}})^-, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \not\models^A \psi$, that is, by the flatness of $\mathcal{CO}$, there exists at least one $s' \in (\{s\}^A_{\mathbf{X}=\mathbf{x}})^-$ such that $(\{s'\}, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \not\models^A \psi$. Now let the tuple of choice functions $\mathbf{F}$ pick $\mathbf{F}(s[\mathbf{x}/\mathbf{X}]) = s'(\mathbf{D_X})$ for any such s, so that $s[\mathbf{x}/\mathbf{X}][\mathbf{F}(s[\mathbf{x}/\mathbf{X}])/\mathbf{D_X}] = s'$ and $T^-[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{D_X}] = \{s' : s \in T^-\}$. By construction, all the assignments in $\{s' : s \in T^-\}$ satisfy $Eq(G_{\mathbf{X}=\mathbf{x}})$ and $\mathbf{X} = \mathbf{x}$, and furthermore they do not satisfy ψ; thus, by induction hypothesis and the usual considerations about reducts, $M_T, T^-[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{D_X}] \models Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr_A^d(\psi, G_{\mathbf{X}=\mathbf{x}})$ and the conclusion follows. □

We can also obtain a relational version of the translation (as in Theorem 3.19) and use it to produce, exactly as in Corollary 3.20, embeddings of the causal-observational languages $\mathcal{CO}^0, \mathcal{COD}^0$ and $\mathcal{CO}^0_{\sqcup}$ (over finite variable domains) into the Bernays-Schönfinkel-Ramsey fragments of first-order and dependence logic. For the unrestricted languages we can also obtain embeddings into the Bernays-Schönfinkel-Ramsey if we restrict the semantics to A-unique-solution causal teams (analogously to Corollary 3.21).

4.2. The recursive case

The recursive causal teams have acyclic causal graphs; thus they are models of non-circular causation. In the literature on causal inference it is then not uncommon to restrict attention to the recursive case; the recursive models are technically easier to treat and have a less controversial causal interpretation (see [30] for a discussion). Also the literature on causal teams has until now mostly favoured the investigation of this case. We then think it convenient to see how the approach of the present paper can be refined in this special case.

First of all, in the recursive case the definition of intervention can be reduced to a more concrete presentation, which will be seen to agree with *both* the Halpern-style definition (section 2.4) and the alternative one (section 4.1) over recursive causal teams.

Definition 4.9. [Intervention, recursive case]

Let $T = (T^-, G, \mathcal{F})$ be a causal team over some signature $\sigma = (Dom, Ran)$. Let $\mathbf{X} = \mathbf{x}$ stand for a consistent conjunction $X_1 = x_1 \wedge \cdots \wedge X_n = x_n$ over σ. The **intervention** $do(\mathbf{X} = \mathbf{x})$ on T is the procedure that generates a new causal team $T^R_{\mathbf{X}=\mathbf{x}} = ((T^R_{\mathbf{X}=\mathbf{x}})^-, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})$ over σ, where $G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}$ are as in Definition 2.3, and $(T^R_{\mathbf{X}=\mathbf{x}})^-$ is defined as follows:

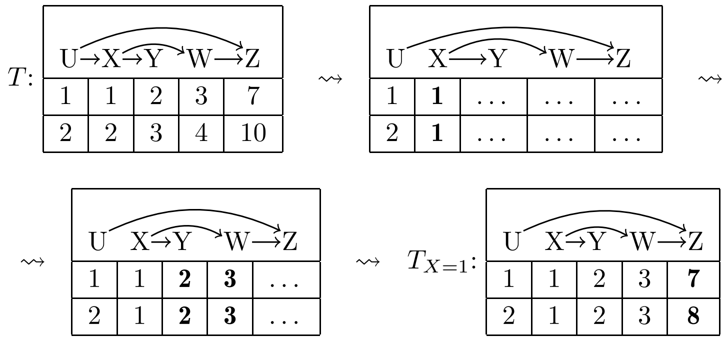
- $(T^R_{\mathbf{X}=\mathbf{x}})^- = \{s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}} \mid s \in T^-\}$, where each $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}$ is the unique assignment compatible with $\mathcal{F}_{\mathbf{X}=\mathbf{x}}$ defined (recursively) as

$$s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(V) = \begin{cases} x_i & \text{if } V = X_i \in \mathbf{X} \\ s(V) & \text{if } V \in Exo(T) \setminus \mathbf{X} \\ \mathcal{F}(V)(s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(PA_V)) & \text{if } V \in End(T) \setminus \mathbf{X} \end{cases}$$

We will simply write $T_{\mathbf{X}=\mathbf{x}}$ for $T^R_{\mathbf{X}=\mathbf{x}}$ when we do not need to distinguish this object from those obtained by other definitions of intervention.

The result of an intervention on a recursive causal team can then be computed via an intuitive iterative procedure; we illustrate this with an example.

Example 4.10. Consider a causal team T with $Dom = \{U, X, Y, W, Z\}$, function component $\mathcal{F}(X) = U, \mathcal{F}(Y) = X + 1, \mathcal{F}(W) = X + 2, \mathcal{F}(Z) = U + 2 * W$ and team and graph component as in the picture below; we apply to T the intervention $do(X = 1)$.



The team component $T^-_{X=1}$ is produced by first rewriting with 1 the values in the X column, and then recomputing the columns corresponding to each descendant V of X as soon as the new values for the parents of V are available. $\mathcal{F}_{X=1}$ is as $\mathcal{F}$, but without a function for X; $G_{X=1}$ is obtained by removing from G_T all arrows that point to X (in this case, the arrow (U, X)).

As the example illustrates, the structure of the causal graph determines a specific (pre-)order in which the variables need to be updated. We can understand what pre-order this is by introducing a notion of "distance" (from a set of variables $\mathbf{X}$ to a variable Y). This was done in [2] for the case of finite variable domains; we show here that this observation can be extended also to causal teams with infinitely many variables. Unless otherwise specified, in the rest of the section we allow the ranges of variables to have arbitrary (set-sized) cardinality. We will also assume that the reader is familiar with the basics of the theory of ordinals. The symbols sup, max, + will denote the corresponding ordinal operations. Now write $\mathbf{D_X}$ for the set of strict descendants of $\mathbf{X}$ in $G_{\mathbf{X}=\mathbf{x}}$ and $\mathbf{D}^+_{\mathbf{X}}$ for $\mathbf{D_X} \cup \mathbf{X}$.

Definition 4.11 *(distance).* Let $G = (\mathbf{V}, E)$ be an acyclic directed graph, $\mathbf{X}$ a finite subset of $\mathbf{V}$. Then, for every $Y \in \mathbf{D}^+_{\mathbf{X}}$ we define:

$$d(\mathbf{X}, Y) := \begin{cases} 0 & \text{if } Y \in \mathbf{X} \\ \sup\{d(\mathbf{X}, Z) + 1 \mid Z \in PA_Y \cap \mathbf{D}^+_{\mathbf{X}}\} & \text{if } Y \in \mathbf{D_X}. \end{cases}$$

An intervention $do(\mathbf{X} = \mathbf{x})$ is then computed by replacing all values in the $\mathbf{X}$-columns of the team with $\mathbf{x}$, then updating all columns corresponding to variables at distance 1 from $\mathbf{X}$, then all those at distance 2, and so on. This point will be formalized in Lemma 4.17. For now, we observe that, by the next lemma, each variable in $\mathbf{D_X}$ is "further away" from $\mathbf{X}$ than its parents; and if Y is a strict descendant of $\mathbf{X}$, then at least one of the parents of Y is a descendant of $\mathbf{X}$; thus, it is indeed possible to calculate the new values for a variable at distance $n+1$ from $\mathbf{X}$ only after calculating the new values for some variables at distance $\leq n$. Furthermore, it should be clear from the definition of the distance that there are no "gaps", that is, if there is a variable at distance $n+1$, all distances $\leq n$ are obtained by some variable and if there is a variable at distance λ for some limit ordinal λ then there are variables at distances n for all $n < \lambda$.

Lemma 4.12. *Let $G = (\mathbf{V}, E)$ be an acyclic directed graph. Let $\mathbf{X}$ be a finite subset of $\mathbf{V}$, $Y \in \mathbf{D_X}$, and d be the distance from Definition 4.11. If $W \in PA_Y^G \cap \mathbf{D_X^+}$, then $d(\mathbf{X}, W) < d(\mathbf{X}, Y)$.*

Proof. Since $Y \in \mathbf{D_X}$ and G is acyclic, we have $Y \notin \mathbf{X}$. Then $d(\mathbf{X}, Y) \geq 1$, and thus

$$\begin{aligned} d(\mathbf{X}, Y) &= \sup\{d(\mathbf{X}, Z) + 1 \mid Z \in PA_Y \cap \mathbf{D_X^+}\} \\ &\geq d(\mathbf{X}, W) + 1 > d(\mathbf{X}, W). \quad \square \end{aligned}$$

We can now prove that, in the finite case, our notion of distance coincides with that which was suggested in [2].

Theorem 4.13. *Let $G = (\mathbf{V}, E)$ be an acyclic directed graph. If $\mathbf{V}$ is finite, and $\mathbf{X} \subseteq \mathbf{V}$, then for each $Y \in \mathbf{D_X}$:*

$$d(\mathbf{X}, Y) = \max\{\textit{lengths of paths in } G_{\mathbf{X}=\mathbf{x}} \textit{ from some } X \in \mathbf{X} \textit{ to } Y\}.$$

Proof. Let κ be the smallest infinite value taken by the distance (say $d(\mathbf{X}, W) = \kappa$ for some W). By Lemma 4.12, for every $Z \in PA_W \cap \mathbf{D_X^+}$, $d(\mathbf{X}, Z) < d(\mathbf{X}, W)$ is then a finite number. Furthermore, since G is finite and acyclic, $\{d(\mathbf{X}, Z) \mid Z \in PA_W \cap \mathbf{D_X^+}\}$ is a finite set. Thus $d(\mathbf{X}, W) = \sup\{d(\mathbf{X}, Z) + 1 \mid Z \in PA_W \cap \mathbf{D_X^+}\}$ is finite, a contradiction. We then conclude that $d(\mathbf{X}, Y)$ is finite for every $Y \in \mathbf{D_X^+}$. Therefore, we can prove the statement by ordinary induction on $d(\mathbf{X}, Y) < \omega$.

Since $Y \in \mathbf{D_X}$ and G is acyclic, we have $Y \notin \mathbf{X}$. Thus by definition $d(\mathbf{X}, Y) > 0$.

If $d(\mathbf{X}, Y) = 1$, by definition we have that $d(\mathbf{X}, Z) = 0$ for all $Z \in PA_Y \cap \mathbf{D_X^+}$; therefore, $PA_Y \cap \mathbf{D_X^+} \subseteq \mathbf{X}$ and (since in $G_{\mathbf{X}=\mathbf{x}}$ there are no arrows connecting distinct variables of $\mathbf{X}$) all paths in $G_{\mathbf{X}=\mathbf{x}}$ from $\mathbf{X}$ to Y have length 1, as required.

Now suppose that $d(\mathbf{X}, Y) = k+1$. By the inductive assumption the statement holds for all variables W with $d(\mathbf{X}, W) \leq k$; in particular, for all $Z \in PA_Y \cap \mathbf{D_X^+}$ the inductive hypothesis yields

$$d(\mathbf{X}, Z) = \max\{\text{lengths of paths in } G_{\mathbf{X}=\mathbf{x}} \text{ from some } X \in \mathbf{X} \text{ to } Z\}.$$

But then $d(\mathbf{X}, Y) =$

$$\begin{aligned} &= \sup\{\max\{\text{lengths of paths in } G_{\mathbf{X}=\mathbf{x}} \text{ from some } X \in \mathbf{X} \text{ to } Z\} + 1 \mid Z \in PA_Y \cap \mathbf{D_X^+}\} \\ &= \max\{\max\{\text{lengths of paths in } G_{\mathbf{X}=\mathbf{x}} \text{ from some } X \in \mathbf{X} \text{ to } Z\} + 1 \mid Z \in PA_Y \cap \mathbf{D_X^+}\} \\ &= \max\{\text{lengths of paths in } G_{\mathbf{X}=\mathbf{x}} \text{ from some } X \in \mathbf{X} \text{ to some } Z \in PA_Y \cap \mathbf{D_X^+}\} + 1. \end{aligned}$$

On the other hand, notice that every path P from some $X \in \mathbf{X}$ to Y ends with an edge of the form (Z, Y) for some $Z \in PA_Y \cap \mathbf{D_X^+}$, thus $\text{length}(P) = \text{length}(P') + 1$ where P' is some path from X to Z (the path obtained removing the last edge of P). Thus

$$\begin{aligned} &\max\{ \text{ lengths of paths in } G_{\mathbf{X}=\mathbf{x}} \text{ from some } X \in \mathbf{X} \text{ to } Y \ \} \\ &= \max\{ \text{ length } (P') + 1 \text{ for } P' \text{ path in } G_{\mathbf{X}=\mathbf{x}} \text{ from an } X \in \mathbf{X} \text{ to a } Z \in PA_Y \cap \mathbf{D_X^+} \ \} \end{aligned}$$

$$= \max\{ \text{ lengths of paths in } G_{\mathbf{X}=\mathbf{x}} \text{ from some } X \in \mathbf{X} \text{ to some } Z \in PA_Y \cap \mathbf{D}^+_{\mathbf{X}} \} + 1,$$

where the second equality used the finitude and acyclicity of G. Putting all equations together, one obtains the desired result. □

The new definition of intervention induces as before a semantic clause for the interventionist counterfactuals; and, keeping the same clauses for the other connectives,[21] we obtain a semantics that we may denote, where there is risk of confusion, by the symbol $\models^R$. However, as the next results show, there is *no* risk of confusion: the relation $\models^R$ is just the restriction of either $\models^H$ or $\models^A$ to recursive causal teams.

Lemma 4.14. *Let $T = (T^-, G, \mathcal{F})$ be a recursive causal team of signature σ with $T^- = \{s\}$, and let $\mathbf{X} = \mathbf{x}$ be a consistent conjunction of signature σ. Write $\mathbf{U}$ for $Exo(T) \setminus \mathbf{X}$. Then there is a unique assignment t of signature σ which is compatible with $(G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})$ and such that $t(\mathbf{U}) = s(\mathbf{U})$ and $t(\mathbf{X}) = \mathbf{x}$; t is $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}$.*

Proof. Notice first that such a t exists: given values for $\mathbf{X}$ and $\mathbf{U}$, the values of all variables can be recomputed using the procedure described in Definition 4.10, yielding the assignment $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}$.

Suppose now that some t satisfies the required constraints; we will prove that $t = s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}$. Suppose first $V = X_i \in \mathbf{X}$; then, since $t(\mathbf{X}) = \mathbf{x}$, we have $t(V) = x_i = s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(V)$. If $V \in \mathbf{U} = Exo(T) \setminus \mathbf{X}$, then since $t(\mathbf{U}) = s(\mathbf{U})$ we have $t(V) = s(V) = s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(V)$. Finally, if $V \in End(T) \setminus \mathbf{X}$, we prove that $t(V) = s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(V)$ by induction on the distance $d(\mathbf{XU}, V)$ (as per Definition 4.11; this distance is well-defined because, by the acyclicity of G, $V \in \mathbf{D}^+_{\mathbf{XU}}$ for all variables V).

Observe that $Y \in \mathbf{D}_{\mathbf{XU}}$, and thus by Lemma 4.12 we have (*): for each $Y \in PA_V$, $d(\mathbf{XU}, Y) < d(\mathbf{XU}, V)$. Now, since t is compatible with $(G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})$ we have $t(V) = \mathcal{F}_{\mathbf{X}=\mathbf{x}}(t(PA_V)) = \mathcal{F}_{\mathbf{X}=\mathbf{x}}(s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(PA_V)) = s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(V)$, where in the second equality we used (*) and the inductive hypothesis.

Thus we conclude that $t = s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}$, and therefore that t is unique. □

Lemma 4.15. *For any recursive causal team T, $T^R_{\mathbf{X}=\mathbf{x}} = T_{\mathbf{X}=\mathbf{x}} = T^A_{\mathbf{X}=\mathbf{x}}$.*

Proof. It suffices to prove that $(T^R_{\mathbf{X}=\mathbf{x}})^- = (T_{\mathbf{X}=\mathbf{x}})^- = (T^A_{\mathbf{X}=\mathbf{x}})^-$.

Write $T = (T^-, G, \mathcal{F})$. According to the Halpern-style definition (2.3), $(T_{\mathbf{X}=\mathbf{x}})^-$ is the set of assignments t that are compatible with $(G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}})$ and such that $t(\mathbf{X}) = \mathbf{x}$ and $t(\mathbf{U}) \in T^-(\mathbf{U})$ (where $\mathbf{U} = Exo(T) \setminus \mathbf{X}$). Since T is recursive, by Lemma 4.14 we have $(T_{\mathbf{X}=\mathbf{x}})^- = \{s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}} \mid s \in T^-\} = (T^R_{\mathbf{X}=\mathbf{x}})^-$.

$(T^A_{\mathbf{X}=\mathbf{x}})^-$ is the set of assignments t of $(T_{\mathbf{X}=\mathbf{x}})^-$ that satisfy the further condition that $t(\mathbf{N_X}) \in T^-(\mathbf{N_X})$ (where $\mathbf{N_X}$ is the set of endogenous nondescendants of $\mathbf{X}$). Now we know that each $t \in (T_{\mathbf{X}=\mathbf{x}})^-$ is of the form $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}$ for some $s \in T^-$. If we show that, for each such s, $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(\mathbf{N_X}) = s(\mathbf{N_X})$, then we conclude that $(T^A_{\mathbf{X}=\mathbf{x}})^- = (T_{\mathbf{X}=\mathbf{x}})^-$.

We proceed by induction on $d(Exo(T) \setminus \mathbf{X}, V)$. For the purposes of induction, we need to prove a stronger statement: that, for each $V \in (Exo(T) \setminus \mathbf{X}) \cup \mathbf{N_X}$, $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(V) = s(V)$. By the acyclicity of $G_{\mathbf{X}=\mathbf{x}}$, there is always at least a variable $U \in Exo(T) \setminus \mathbf{X}$ with a path from U to V; therefore $d(Exo(T) \setminus \mathbf{X}, V)$ is always well-defined. It should also be clear that, for each $V \in \mathbf{N_X}$, we have $PA_V \subseteq (Exo(T) \setminus \mathbf{X}) \cup \mathbf{N_X} \subseteq \mathbf{D}^+_{Exo(T)\setminus\mathbf{X}}$, and thus, by Lemma 4.12 we have $d(Exo(T) \setminus \mathbf{X}, Y) < d(Exo(T) \setminus \mathbf{X}, V)$ for each such V and each $Y \in PA_V$.

Now, if $V \in Exo(T) \setminus \mathbf{X}$, by the definition of $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}$ we have $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(V) = s(V)$, as required. If instead $V \in \mathbf{N_X} \subseteq End(T) \setminus \mathbf{X}$, by definition of $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}$ we have $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(V) = \mathcal{F}(V)(s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(PA_V))$. Since $PA_V \subseteq (Exo(T) \setminus \mathbf{X}) \cup \mathbf{N_X}$, the inductive hypothesis gives us $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(Y) = s(Y)$ for each $Y \in PA_V$. Therefore $s^{\mathcal{F}}_{\mathbf{X}=\mathbf{x}}(V) = \mathcal{F}(V)(s(PA_V)) = s(V)$, as required. □

[21] With a straightforward change in the clause for $\supset$; cp. with the definition of $\models^A$.

Theorem 4.16. *Let T be a recursive causal team of signature σ, and let φ be a $\mathcal{COD}(\sigma)$ or $\mathcal{CO}_{\sqcup}(\sigma)$ formula. Then*

$$T \models^R \varphi \iff T \models^H \varphi \iff T \models^A \varphi.$$

Proof. This is proved by a straightforward induction. It suffices to consider the case for φ of the form $\mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi$.

We have $T \models^R \mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi \iff T^R_{\mathbf{X}=\mathbf{x}} \models^R \psi$, $T \models^H \mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi \iff T_{\mathbf{X}=\mathbf{x}} \models^H \psi$ and $T \models^A \mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi \iff T^A_{\mathbf{X}=\mathbf{x}} \models^A \psi$. Since $T^R_{\mathbf{X}=\mathbf{x}} = T_{\mathbf{X}=\mathbf{x}} = T^A_{\mathbf{X}=\mathbf{x}}$ by Lemma 4.15, the statement amounts to the equivalence $T_{\mathbf{X}=\mathbf{x}} \models^R \psi \iff T_{\mathbf{X}=\mathbf{x}} \models^H \psi \iff T_{\mathbf{X}=\mathbf{x}} \models^A \psi$, which immediately follows from the inductive hypothesis. □

This theorem tells us that both the Halpern-style and our alternative definition of intervention do not contradict the commonly accepted notion of intervention on a recursive causal model. Therefore, in the following of this section we will simply write $\models$ for $\models^R$.

Let us now discuss the translations into dependence logic in the recursive case. Of course, since recursive causal teams are only a special case of causal teams, by Theorem 4.16 the translations tr, tr^d, tr_A and tr^d_A are still correct in this restricted case. But we may wonder whether on this restricted class of structures it is possible to obtain a more informative translation. This is indeed the case. We define a translation tr_R by replacing the translation of selective implications and counterfactuals with the following clauses:

- $tr_R(\alpha \supset \chi, G) := tr_R(\alpha^d, G) \vee tr_R(\chi, G)$, where α^d is the dual of the $\mathcal{CO}$ formula α discussed in section 2.6;
- $tr_R(\mathbf{X} = \mathbf{x} \,\Box\!\!\rightarrow \psi, G) := \exists \mathbf{X} \exists \mathbf{D_X}(\mathbf{X} = \mathbf{x} \wedge Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge tr_R(\psi, G_{\mathbf{X}=\mathbf{x}}))$.[22]

The most evident advantage of this specialized translation is that it completely dispenses with universal quantifiers and requires no dual translation tr^d_R, due to the different treatment of selective implication. In the following, we prove its correctness and we use it to show that, under the recursivity restriction, the causal-observational languages can be embedded into the *existential* fragment of first-order (dependence) logic. As before, this will be possible only for finite variable domains.

We see in the next lemma that in the recursive case the effect of an intervention on a causal team T can be "simulated" by a sequence of supplementations; this result is *not* limited to finite signatures. Since we assume causal teams to be recursive, the variables in $\mathbf{D_X}$ can be partitioned according to their distance from $\mathbf{X}$. So we may write $\mathbf{D}^1_{\mathbf{X}}, \dots \mathbf{D}^\kappa_{\mathbf{X}} \dots$ for the set of variables from $\mathbf{D_X}$ which are at distance $1 \dots \kappa \dots$ from $\mathbf{X}$. We omit the suffix $\mathbf{X}$ when clear from the context. As mentioned before, when performing an intervention $do(\mathbf{X} = \mathbf{x})$, we are updating first the variables in $\mathbf{X}$, then those in $\mathbf{D}^1$, in $\mathbf{D}^2$, and so on. Write n_j for $card(\mathbf{D}^j)$; n_j might in general be an infinite cardinal.

Given a causal team $T = (T^-, G, \mathcal{F})$, let M_T be the first-order structure given by Definition 3.16. Given a tuple of distinct variables $\mathbf{X}$ and values $\mathbf{x} \in \mathbf{X}$, we define simultaneously a sequence of extended teams $T_0, \dots T_\kappa, \dots$ and a sequence of functions $\mathbf{F}^j : T_{j-1} \to M_T^{n_j}$ (for $j > 0$), as follows:

- $T_0 := T[\mathbf{x}/\mathbf{X}]$
- $T_{j+1} := T_j[\mathbf{F}^{j+1}/\mathbf{D}^{j+1}]$
- $\mathbf{F}^{j+1}(s) := (f^{M_T}_{V_1}(s(PA_{V_1})), \dots, f^{M_T}_{V_{n_j}}(s(PA_{V_{n_{j+1}}})))$, where $s \in T_j$ and $\{V_1, \dots V_{n_{j+1}}\} = \mathbf{D}^{j+1}$.

[22] $\mathbf{D_X}$ is the set of strict descendants of $\mathbf{X}$, listed in a fixed alphabetical order.

The following lemma ensures that, in M_T, the effect of an intervention on T can be simulated by applying a sequence of supplementations to T^-; the sequence is determined by the signature and by the function component of T.

Lemma 4.17. *Let $T = (T^-, G, \mathcal{F})$ be a recursive causal team of signature $\sigma = (Dom, Ran)$, and $M_T = (|M_T|, (c^{M_T})_{c\in|M_T|}, (f_V^{M_T})_{V\in End(T)})$ as in Definition 3.16. Then, for every $\mathbf{X} \subseteq Dom$ and every $\mathbf{x} \in Ran(\mathbf{X})$:*

$$(T_{\mathbf{X}=\mathbf{x}})^- = T^-[\mathbf{x}/\mathbf{X}, \mathbf{F}^1/\mathbf{D}^1, \dots, \mathbf{F}^\kappa/\mathbf{D}^\kappa, \dots]$$

where $\kappa < sup\{d(\mathbf{X}, V) \mid V \in Dom\}$ if there is no variable V with $d(\mathbf{X}, V) = sup\{d(\mathbf{X}, V) \mid V \in Dom\}$, and $\kappa \le sup\{d(\mathbf{X}, V) \mid V \in Dom\}$ otherwise.

Proof. Straightforward from the definition of intervention in the recursive case, the definition of the interpretations $f_V^{M_T}$ and Proposition 3.12. $\square$

Theorem 4.18. *Let $T = (T^-, G, \mathcal{F})$ be a recursive causal team and let φ be a formula of $\mathcal{COD}$ or $\mathcal{CO}_\sqcup$. Then:*

$$M_T, T^- \models tr_R(\varphi, G) \iff T \models \varphi.$$

Proof. All cases are treated as in the proof of Theorem 3.18, with the exception of the cases for $\supset$ and $\Box\!\!\rightarrow$.

- Suppose that $M_T, T^- \models tr_R(\alpha^d, G) \vee tr_R(\chi, G)$. Then $T^- = T_1^- \cup T_2^-$, where $M_T, T_1^- \models tr_R(\alpha^d, G)$ and $M_T, T_2^- \models tr_R(\chi, G)$. Since $(T^-, G, \mathcal{F})$ is a recursive causal team, so are $(T_1^-, G, \mathcal{F})$ and $(T_2^-, G, \mathcal{F})$; and so, by induction hypothesis,[23] $(T_1^-, G, \mathcal{F}) \models \alpha^d$ and $(T_2^-, G, \mathcal{F}) \models \chi$.
  By Lemma 2.7, we then have that $(T_1^-, G, \mathcal{F}) \models \neg\alpha$, i.e. $(\{s\}, G, \mathcal{F}) \not\models \alpha$ for all $s \in T_1^-$. This implies that the set $(T^-)^\alpha = \{s \in T^- : (\{s\}, G, \mathcal{F}) \models \alpha\}$ is contained in T_2^-; and thus, by downwards closure, $(T^-, G, \mathcal{F})^\alpha = ((T^-)^\alpha, G, \mathcal{F}) \models \chi$. Therefore, $(T^-, G, \mathcal{F}) \models \alpha \supset \chi$, as required.
  Conversely, suppose that $(T^-, G, \mathcal{F}) \models \alpha \supset \chi$. Then $(T_2^-, G, \mathcal{F}) \models \chi$, where $T_2^- = (T^-)^\alpha = \{s \in T^- : (\{s\}, G, \mathcal{F}) \models \alpha\}$. Now let $T_1^- = T^- \setminus T_2^- = \{s \in T^- : (\{s\}, G, \mathcal{F}) \not\models \alpha\}$.
  By definition and the flatness of $\neg\alpha$, $(T_1^-, G, \mathcal{F}) \models \neg\alpha$, and therefore by Lemma 2.7 $(T_1^-, G, \mathcal{F}) \models \alpha^d$, which by induction hypothesis implies that $M_T, T_1^- \models tr(\alpha^d, G)$; and furthermore, by induction hypothesis we also have that $M_T, T_2^- \models tr(\chi, G)$. But $T^- = T_1^- \cup T_2^-$, and so $M_T, T^- \models tr(\alpha^d, G) \vee tr(\chi, G)$, as required.
- First, assume that $M_T, T^- \models \exists\mathbf{X}\exists\mathbf{D_X}(\mathbf{X} = \mathbf{x} \wedge Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge tr_R(\psi, G_{\mathbf{X}=\mathbf{x}}))$. Then there is a sequence of functions $\mathbf{F} = (F_V)_{V\in\mathbf{D_X}}$ such that $M_T, T^-[\mathbf{x}/\mathbf{X}, \mathbf{F}/\mathbf{D_X}] \models \mathbf{X} = \mathbf{x} \wedge Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge tr_R(\psi, G_{\mathbf{X}=\mathbf{x}})$. The second conjunct entails that $F_V(s) = f_V(s(PA_V))$ for all $V \in \mathbf{D_X}$, and therefore that

  $$T^-[\mathbf{x}/\mathbf{X}, \mathbf{F}/\mathbf{D_X}] = T^-[\mathbf{x}/\mathbf{X}, \mathbf{F}^1/\mathbf{D}^1, \dots, \mathbf{F}^\kappa/\mathbf{D}^\kappa, \dots]$$

  for the $\mathbf{F}^j$ defined, as above, by the condition $\mathbf{F}^j(s) := (f_{V_1}^{M_T}(s(PA_{V_1})), \dots, f_{V_{n_j}}^{M_T}(s(PA_{V_{n_j}})))$, where $\{V_1, \dots V_{n_j}\} = \mathbf{D}_j$.
  The third conjunct, therefore, entails that $M_T, T^-[\mathbf{x}/\mathbf{X}, \mathbf{F}^1/\mathbf{D}^1, \dots, \mathbf{F}^\kappa/\mathbf{D}^\kappa, \dots] \models tr_R(\psi, G_{\mathbf{X}=\mathbf{x}})$. Thus, by Lemma 4.17 plus taking reducts, we obtain $M_{T_{\mathbf{X}=\mathbf{x}}}, (T_{\mathbf{X}=\mathbf{x}})^- \models tr_R(\psi, G_{\mathbf{X}=\mathbf{x}})$. The inductive hypothesis then gives $T_{\mathbf{X}=\mathbf{x}} \models \psi$; and finally, $T \models \mathbf{X} = \mathbf{x} \Box\!\!\rightarrow \psi$.

[23] Recall that the model M_T is determined only by the graph and function components of T, and so $M_T = M_{T_1} = M_{T_2}$.

Conversely, assume that $T \models \mathbf{X} = \mathbf{x} \;\Box\!\!\rightarrow\; \psi$. Then $T_{\mathbf{X}=\mathbf{x}} \models \psi$. $T_{\mathbf{X}=\mathbf{x}}$ has graph $G_{\mathbf{X}=\mathbf{x}}$; so, by the inductive hypothesis, $M_{T_{\mathbf{X}=\mathbf{x}}}, (T_{\mathbf{X}=\mathbf{x}})^- \models tr_R(\psi, G_{\mathbf{X}=\mathbf{x}})$. Since $M_{T_{\mathbf{X}=\mathbf{x}}}$ is a reduct of M_T, we obtain $M_T, (T_{\mathbf{X}=\mathbf{x}})^- \models tr_R(\psi, G_{\mathbf{X}=\mathbf{x}})$. Now by Lemma 4.17 we also have $(T_{\mathbf{X}=\mathbf{x}})^- = T^-[\mathbf{x}/\mathbf{X}, \mathbf{F}^1/\mathbf{D}^1, \ldots, \mathbf{F}^\kappa/\mathbf{D}^\kappa, \ldots]$. Furthermore, it is straightforward to see that $M_T, T^-[\mathbf{x}/\mathbf{X}, \mathbf{F}^1/\mathbf{D}^1, \ldots, \mathbf{F}^\kappa/\mathbf{D}^\kappa, \ldots] \models \mathbf{X} = \mathbf{x} \wedge Eq(G_{\mathbf{X}=\mathbf{x}})$. By the clause for existential quantifiers and Proposition 3.12, $M_T, T^- \models \exists \mathbf{X} \exists \mathbf{D}_{\mathbf{X}} (\mathbf{X} = \mathbf{x} \wedge Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge tr_R(\psi, G_{\mathbf{X}=\mathbf{x}}))$, that is, $M_T, T^- \models tr_R(\varphi, G)$. □

Let us denote the class of recursive causal teams as R_σ. Using the terminology introduced immediately before Corollary 3.20, we have the following.

Corollary 4.19. *Let $\sigma = (Dom, Ran)$ be a signature with finite Dom. Then:*

a) $\mathcal{CO}(\sigma)$ R_σ-embeds into the existential fragment of first-order logic

b) $\mathcal{COD}(\sigma)$ R_σ-embeds into the existential fragment of dependence logic

c) $\mathcal{CO}_\sqcup(\sigma)$ R_σ-embeds into the existential fragment of dependence logic.

Proof. Observe that, for $\varphi \in \mathcal{CO}$, $\mathcal{COD}$ or $\mathcal{CO}_\sqcup$, if G is a finite graph $tr_R(\varphi, G)$ is an expression in $\mathbf{FO}(\sqcup, =(\cdot;\cdot))$ in which no universal quantifier appears. Thus, by Theorem 3.15, this expression is equivalent to some expression in the existential fragment of dependence logic.

Now for the case of $\mathcal{CO}$, note that this expression will be flat: indeed, since formulas in $\mathcal{CO}$ are flat in the sense of causal team semantics $M_T, T^- \models tr_R(\varphi, G)$ iff $(T^-, G, \mathcal{F}) \models \varphi$ iff $(\{s\}, G, \mathcal{F}) \models \varphi$ for all $s \in T^-$ iff $M_T, \{s\} \models tr_R(\varphi, G)$ for all such s.

Additionally, $tr_R(\varphi, G)$ will not contain dependence atoms; and therefore, again by Theorem 3.15, it will be equivalent to some existential first-order formula. □

It is well-known that the existential sentences of dependence logic are equivalent to first-order (existential) sentences. However, $tr_R(\varphi, G)$ will typically *not* be a sentence. For example, the formula $(=(X;Y) \vee =(X;Y)) \vee =(Z;V)$ is in the range of the translations of $\mathcal{COD}$ formulas; and it is known that its model-checking problem is NP-complete with respect to data complexity, while first-order formulas always have $\mathrm{AC}_0 \subset \mathrm{LOGSPACE} \subseteq \mathrm{NP}$ model-checking problems.[24] Thus, $\mathcal{COD}(\sigma)$ cannot be embedded into first-order logic.

5. Applications of the embedding to the satisfiability problem

In this section we will show an example of how our embeddings can help in transferring known results for dependence logic to the causal-observational languages. We will use the decidability of existential dependence logic to prove the decidability of a particular satisfiability problem: given a $\mathcal{COD}$ (or $\mathcal{CO}_\sqcup$) formula φ and a finite set of variables Dom, is there a recursive causal team of domain Dom which satisfies φ?[25]

We will need to slightly generalize our embedding results in order to prove this. The embeddings as defined until now are too concrete: they link truth over a causal team T (picked in complete generality) to truth over a very specific pair (M_T, T^-); that is, to a very special case of truth in first-order team semantics. Therefore, information about the satisfiability problem in general (i.e. over all possible pairs (M, S)) cannot

[24] Both results are proved in [26]. The model checking problem for logics based on team semantics asks, given a formula $\varphi(\mathbf{V})$ with free variables $\mathbf{V}$, to decide whether a structure (M, R) is of the form $(M, T(\mathbf{V}))$ for a given team T such that $M, T \models \varphi$, where $T(\mathbf{V})$ is the relation that the team T induces over the sequence of variables $\mathbf{V} = V_1 \ldots V_n$, i.e. $\{(s(V_1) \ldots s(V_n)) \mid s \in T\}$.

[25] As pointed out by one of the reviewers of this paper, it would be easy to obtain an answer to this question if the answer to the analogous question for language $\mathcal{CO}$ were known. Halpern ([18], p. 328) states that a more general version of the problem (for a language not too dissimilar from $\mathcal{CO}$) is NP-complete, and it would not be difficult to convert this into a result for $\mathcal{CO}$. However, Halpern provides no proof of his claim, and we are not aware of any published proof. Thus, our proof will not go through this claim. Notice also that the argument given in this section proves the decidability of $\mathcal{CO}$ as a special case.

tell us much about satisfiability in causal team semantics. In the following section, we explain how to improve the flow of information in this direction.

5.1. Going backwards: from teams to causal teams

The embedding results we have presented until now have the following form: for all causal teams T,

$$T \models \varphi \Leftrightarrow M_T, T^- \models tr(\varphi, G),$$

where M_T and T^- are appropriately constructed from T. We want now to proceed in the opposite direction: if we are given an arbitrary structure M and an arbitrary team T on M, can we build a causal team T^* such that

$$T^* \models \varphi \Leftrightarrow M, T \models tr(\varphi, G)?$$

We do not know the full answer to this question, but we show here that the right-to-left implication can be guaranteed - provided we pick T^* as a function of G. We now describe in detail this construction.

When discussing a graph, the **indegree** of a node V is the cardinality of the set of parents of V.

Definition 5.1. We say that a *finite* graph $G = (\mathbf{V}, E)$ (where $\mathbf{V}$ is a finite set of variables) is:

- **compatible with a first-order structure** M in case (for all $m, n \in \mathbb{N}$) if G has m nodes of indegree n, then M interprets at least m function symbols of arity n
- **compatible with a team** T in case $dom(T) = \mathbf{V}$ and, for all $V \in \mathbf{V}$, if $PA_V = \{X_1, \dots, X_n\}$, then $T \models =(X_1, \dots, X_n; V)$.

We then abide to the following:

Convention 5.2. *For every pair G, M, where M is a first-order structure and $G = (\mathbf{V}, E)$ a graph compatible with M, we fix a correspondence between the endogenous variables in $\mathbf{V}$ and (distinct) function symbols interpreted by M. The chosen correspondence is such that to each $V \in \mathbf{V}$ of indegree $n > 0$ is associated a function symbol f_V of arity n. (Such a correspondence exists by the definition of compatibility with M.)*

Notice the following important fact.

Fact 5.3. *Let $G = (\mathbf{V}, E)$ be a graph, and M a first-order structure. Let $\sigma = (Dom, Ran)$ be the signature given by $Dom = \mathbf{V}$, $Ran(V) = M$ for each $V \in \mathbf{V}$.*

If G is compatible with M, then, for every formula $\varphi \in \mathcal{COD}(\sigma) \cup \mathcal{CO}_\sqcup(\sigma)$ the translations $tr(\varphi, G)$ and $tr_R(\varphi, G)$ are well-formed formulas in the vocabulary of M.

Now, we claim that any given first-order structure M induces a family of functions π_G (one for each graph G compatible with M) that associate to each team T compatible with G a causal team $\pi_G(M, T)$ such that, for all $\varphi \in \mathcal{COD} \cup \mathcal{CO}_\sqcup$,

$$M, T \models tr(\varphi, G) \quad \Rightarrow \quad \pi_G(M, T) \models \varphi.$$

We will write, more briefly, T^G for $\pi_G(M, T)$. We define $T^G = (T^-, G, \mathcal{F})$ as follows:

- $T^- := T$

- G is the parameter of π_G
- for each $V \in End(G)$, $\mathcal{F}(V)$ is the function that associates to each tuple $\mathbf{a} \in M^{ar(f_V)}$ the element $f_V^M(\mathbf{a}) \in M$.

Notice that the domain of T^G is $\mathbf{V}$, the set of vertices of G; for all $V \in \mathbf{V}$, we can take $Ran(V)$ to be M.

Theorem 5.4. *Let M be a first-order structure and T a team on M. For every $G = (\mathbf{V}, E)$ compatible with both M and T, let $T^G = (T, G, \mathcal{F})$ be defined as before. Let $\varphi \in \mathcal{COD}(\sigma) \cup \mathcal{CO}_{\sqcup}(\sigma)$, where σ is the signature given by $Dom = \mathbf{V}$ and $Ran(V) = M$ for each $V \in \mathbf{V}$. Then:*

$$M, T \models tr(\varphi, G) \quad \Rightarrow \quad T^G \models \varphi$$

and, if $\varphi \in \mathcal{CO}$,

$$M, T \models tr^d(\varphi, G) \quad \Rightarrow \quad \forall s \in T : (\{s\}, G, \mathcal{F}) \not\models \varphi.$$

Proof. We prove the claim simultaneously for all teams T (on M) such that G is compatible with T and M. The proof is mostly identical to the left-to-right direction of Theorem 3.18:

- Suppose that $M, T \models tr(X = x, G)$. Then $M, T \models X = x$, and so $s(X) = x$ for all $s \in T$. But since the team component of T^G is precisely T, we have at once that $T^G \models X = x$. A similar argument holds for dependence atoms.
  Suppose now that $M, T \models tr^d(X = x, G)$. Then $M, T \models X \neq x$, and so $s(X) \neq x$ for all assignments $s \in T$, and finally $(\{s\}, G, \mathcal{F}) \not\models X = x$ for all $s \in T$. Instead, there is no dual case for dependence atoms.
- Suppose that $M, T \models tr(\psi, G) \wedge tr(\theta, G)$. Then $M, T \models tr(\psi, G)$ and $M, T \models tr(\theta, G)$, and so by induction hypothesis $T^G \models \psi$ and $T^G \models \theta$ and finally $T^G \models \psi \wedge \theta$.
  If instead $M, T \models tr^d(\psi \wedge \theta, G)$ we have that $M, T \models tr^d(\psi, G) \vee tr^d(\theta, G)$: therefore, $T = T_1 \cup T_2$ where $M, T_1 \models tr^d(\psi, G)$ and $M, T_2 \models tr^d(\theta, G)$. Note that if G is compatible with M and T then it is also compatible with M and T_1 or T_2: thus, by induction hypothesis, $(\{s\}, G, \mathcal{F}) \not\models \psi$ for all $s \in T_1$ and $(\{s\}, G, \mathcal{F}\} \not\models \theta$ for all $s \in T_2$ (note that G and $\mathcal{F}$ are defined in terms of the model M only, and so are the same in the two expressions). But then for every $s \in T = T_1 \cup T_2$ we have that $(\{s\}, G, \mathcal{F}) \not\models \psi \wedge \theta$, as required.
- The case for disjunction is dual to the previous one.
- If $M, T \models tr(\psi \sqcup \theta, G)$ then $M, T \models tr(\psi, G) \sqcup tr(\theta, G)$. But then by induction hypothesis $T^G \models \psi$ or $T^G \models \theta$, and in either case $T^G \models \psi \sqcup \theta$.
  Note that an expression of the form $\psi \sqcup \theta$ is never in $\mathcal{CO}$, so – as in the case of functional dependencies – the second part of the theorem does not need to be proved.
- If $M, T \models tr(\alpha \supset \chi, G)$ then $T = T_1 \cup T_2$, where $M, T_1 \models tr^d(\alpha, G)$ and $M, T_2 \models tr(\chi, G)$.
  Now let $T^\alpha = \{s \in T \mid M, \{s\} \models \alpha\}$: by the induction hypothesis for α we necessarily have that $T^\alpha \subseteq T_2$; so, by downwards closure $M, T^\alpha \models tr(\chi, G)$ and thus – by the induction hypothesis on χ, observing that if G is compatible with M and T it is also compatible with M and T^α – $(T^\alpha, G, \mathcal{F}) \models \chi$, which implies that $(T, G, \mathcal{F}) \models \alpha \supset \chi$.
  Suppose instead that $M, T \models tr^d(\alpha \supset \chi, G)$, where α and χ are both in $\mathcal{CO}$. Then $M, T \models tr(\alpha, G) \wedge tr^d(\chi, G)$, and so – by induction hypothesis and downward closure – for every $s \in T$ we have that $(\{s\}, G, \mathcal{F}) \models \alpha$ and $(\{s\}, G, \mathcal{F}) \not\models \chi$. But then $(\{s\}, G, \mathcal{F}\} \not\models \alpha \supset \chi$ for all such s, as required.

- If $\mathbf{X} = \mathbf{x}$ is inconsistent then $tr(\mathbf{X} = \mathbf{x} \,\Box\!\!\to \psi, G)$ and $\mathbf{X} = \mathbf{x} \,\Box\!\!\to \psi$ are both trivially true and there is nothing to prove.
  Let us suppose then that $\mathbf{X} = \mathbf{x}$ is consistent and $M, T \models tr(\mathbf{X} = \mathbf{x} \,\Box\!\!\to \psi, G)$. Then, as in Theorem 3.18, $T[\mathbf{x}/\mathbf{X}][\mathbf{M}/\mathbf{V}]$ can be split into two subteams T_1 and T_2, the second of which satisfies $tr(\psi, G_{\mathbf{X}=\mathbf{x}})$ and, by flatness and downwards closure, can be assumed to be the set of all assignments in $T[\mathbf{x}/\mathbf{X}][\mathbf{M}/\mathbf{V}]$ that satisfy $Eq(G_{\mathbf{X}=\mathbf{x}})$. Now $G_{\mathbf{X}=\mathbf{x}}$ is compatible with T_2, since T_2 satisfies $Eq(G_{\mathbf{X}=\mathbf{x}})$. Thus we can apply the induction hypothesis to obtain $(T_2)^{G_{\mathbf{X}=\mathbf{x}}} \models \psi$; but, by construction, $(T_2)^{G_{\mathbf{X}=\mathbf{x}}} = (T^G)_{\mathbf{X}=\mathbf{x}}$, and so $T^G \models \mathbf{X} = \mathbf{x} \,\Box\!\!\to \psi$.
  If instead $\psi \in \mathcal{CO}$ and $M, T \models tr^d(\mathbf{X} = \mathbf{x} \,\Box\!\!\to \psi)$, for all $s \in T$ we have that $M, \{s\} \models \exists^1 \mathbf{X} \exists \mathbf{V}(Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr^d(\psi, G_{\mathbf{X}=\mathbf{x}}))$ by downward closure. Then for some choice of functions $\mathbf{F}$ we have that $M, \{s\}[\mathbf{x}/\mathbf{X}][\mathbf{F}/\mathbf{V}] \models Eq(G_{\mathbf{X}=\mathbf{x}}) \wedge \mathbf{X} = \mathbf{x} \wedge tr^d(\psi, G_{\mathbf{X}=\mathbf{x}})$. Thus, there exists some assignment s' obtainable from s via the intervention $\mathbf{X} = \mathbf{x}$ that satisfies $tr^d(\psi, G_{\mathbf{X}=\mathbf{x}})$ with respect to the model M, and we obtain $(\{s'\}, G_{\mathbf{X}=\mathbf{x}}, \mathcal{F}_{\mathbf{X}=\mathbf{x}}) \not\models \psi$ by the inductive hypothesis, where $\mathcal{F}_{\mathbf{X}=\mathbf{x}}$ is the function component of $(T^G)_{\mathbf{X}=\mathbf{x}}$. Write $\{s\}^G$ for the causal team $(\{s\}, G, \mathcal{F})$; since $s' \in ((\{s\}^G)_{\mathbf{X}=\mathbf{x}})^-$, by the previous statement and downwards closure we have that $(\{s\}^G)_{\mathbf{X}=\mathbf{x}} \not\models \psi$; and finally, this implies that $\{s\}^G \not\models \mathbf{X} = \mathbf{x} \,\Box\!\!\to \psi$, for any such $s \in T$, as required. □

For the sake of our applications, we are more interested in a result of this kind for the recursive case. If we start from an acyclic graph G, the same procedure works if we replace tr with tr_R, and in this case obviously we obtain a recursive causal team T^G.

Theorem 5.5. *Let M be a first-order structure and T a team on M. For every acyclic graph G compatible with both M and T,*

$$M, T \models tr_R(\varphi, G) \quad \Rightarrow \quad T^G \models \varphi.$$

Proof. The proof is mostly identical to the left-to-right part of 4.18, with small modifications analogous to those added in the proof of 5.4, and using the following additional fact:

- $((T^G)_{\mathbf{X}=\mathbf{x}})^- = T^-[\mathbf{x}/\mathbf{X}, \mathbf{F}^1/\mathbf{D}^1, \dots, \mathbf{F}^\kappa/\mathbf{D}^\kappa, \dots]$ (see Lemma 4.17 for notations and an analogous proof). □

5.2. The satisfiability problem

We now discuss the decidability of the satisfiability problem over a finite variable domain (in the recursive semantics). We say that a causal team T is a causal team over Dom if the signature of T is of the form (Dom, Ran), for an appropriate function Ran.

Theorem 5.6. *Let Dom be a finite set of variables, and let φ vary over $\mathcal{COD}$ or $\mathcal{CO}_{\sqcup}$ formulas with variables in Dom. Then it is decidable whether φ is $\models^R$-satisfiable in some* recursive *causal team over Dom.*

We point out that this theorem, together with Theorem 4.16, implies that also $\models^H$-satisfiability and $\models^A$-satisfiability are decidable over recursive causal teams.

Proof. Let $\varphi \in \mathcal{COD}$. Now, if φ is satisfied by some nonempty (recursive) causal team $T = (T^-, G, \mathcal{F})$ then, by Theorem 4.18 and downward closure, for every $s \in T^-$ it holds that $M_T, \{s\} \models Eq(G) \wedge tr_R(\varphi, G)$. Conversely, if $M, \{s\} \models Eq(G) \wedge tr_R(\varphi, G)$ then (by $M, \{s\} \models Eq(G)$) G is compatible with both M and $T = \{s\}$; therefore, by Theorem 5.5 the causal team T^G (which is recursive) satisfies φ. Therefore, $\exists \mathbf{Z}(Eq(G) \wedge tr_R(\varphi, G))$ (where $\mathbf{Z}$ is the set of all free variables of $Eq(G) \wedge tr_R(\varphi, G)$) is a sentence of dependence logic that is satisfiable if and only if there is a recursive causal team with graph G that satisfies φ. Therefore, $\psi := \bigvee_G \exists \mathbf{Z}(Eq(G) \wedge tr_R(\varphi, G))$, where G ranges over all acyclic graphs with the variables in Dom as nodes, is a dependence logic sentence that is satisfiable if and only if φ is satisfiable by some recursive causal team of domain Dom. Since all quantifiers occur in ψ in positive positions, we can extract all the existential quantifiers in ψ so as to obtain a prenex formula ψ^* that has an $\exists^*$ prefix. Each such sentence is satisfiable iff the first-order sentence $(\psi^*)^f$ ("flattening") that is obtained by removing all dependence atoms is satisfiable. The reason for this is that, by downward closure, ψ^* is satisfied on a team if and only if it is satisfied by a singleton team; and on singleton teams, all dependence atoms evaluate as true. Since the construction of $(\psi^*)^f$ from φ is clearly a computable procedure, our problem is reduced to satisfiability in the existential fragment of first-order logic, which is decidable (also in case the vocabulary is not purely relational) by a classical result of Gurevich ([16]).

If $\varphi \in \mathcal{CO}_{\sqcup}$, we can proceed in the same way after replacing $tr_R(\varphi, G)$ with an equivalent dependence logic formula (by removing the $\sqcup$ symbols as explained in the proof of Theorem 3.15). □

We remark that it is not straightforward to extend this result to the general, possibly non-recursive case. Indeed, our main translation tr produces formulas that are equivalent to $\exists^*\forall^*$ formulas *with function symbols*. This fragment of dependence logic (or even first-order logic) is *not* decidable (see [7] or [15]; the result was proved in [16]). On the other hand, the modified translation tr_*, which produces formulas of relational vocabulary, seems not to be suitable for the "converse" embedding proposed in section 5.1.[26]

6. Conclusions

In this paper, we have pursued two lines of thought. The first consisted in systematizing the treatment of causal team semantics in the general (i.e. possibly nonrecursive) case. A definition of interventions in the general case was suggested already in [1,2], but most work in the literature on causal teams has focused on the recursive case. We have shown here that this (call it "Halpern-style") definition is adequate, at least in the sense that it agrees with the uncontroversial definition of intervention on a recursive causal team. On the other hand, we have remarked that the Halpern-style definition of intervention seems to work contrary to the usual intuitions that lie behind causal modelling; the problems with this definition arise already at the level of causal models, and are thus not an artifact of our generalization to causal teams. In view of this objection, we have argued in favour of an alternative definition of intervention; and we have shown that also this alternative clause agrees with the usual definition of intervention in the recursive case. In summary, we have analyzed three alternative semantics $\models^H, \models^A, \models^R$ (which agree in the recursive case) and their mutual relations.

The second line of thought consisted in a systematic effort of embedding the causal-observational languages from [2] (evaluated according to each of the three semantics) into first-order languages. The semantics of classical logic does not come equipped with instruments for encoding causal laws. Nonetheless, we have shown that the simplest of the causal-observational languages, $\mathcal{CO}$, can be embedded into first-order logic; that is, the usual causal reasoning can be encoded into classical logic. The embedding extends straightforwardly to the more complex languages $\mathcal{COD}$ and $\mathcal{CO}_{\sqcup}$, but in this case the target language is first-order dependence logic. With a bit of extra effort, the embeddings for the semantics $\models^H$ and $\models^A$ can be sharpened into embeddings into the Bernays-Schönfinkel-Ramsey fragment of first-order/dependence logic (i.e., prenex formulas with $\exists^*\forall^*$ prefix and relational vocabulary), and similarly, when using $\models^R$, one can obtain embeddings into purely existential fragments. These embeddings come at the price of non-straightforward translations and of a restrictive choice of the models on which to interpret the translations. We believe,

[26] The main problem here is that the constraint $Eq_*(G)$ is insufficient for imposing the correct causal constraints on a generic structure: plausibly, additional quantifiers are needed in order to impose the functionality of the laws.

however, that there is more work to be done in order to generalize the connection between the two semantic frameworks and make it more flexible. This point is illustrated by the variant of the embedding that we presented in section 5. This alternative correspondence ensures that the truth of the translation of φ over an *arbitrary* model (of appropriate signature) entails the truth of φ in an appropriate causal team. The two-way connection established by the two kinds of embeddings allowed us to use a well-known decidability result (for existential dependence logic) to prove the decidability of the satisfiability problem for causal-observational languages over a finite variable domain. This result holds in the recursive case; we leave its extension to the general case as an open problem.

The main take is that the most common kinds of causal reasoning can be in principle carried out in (classical) first-order logic, and causal reasoning paired with discussion of contingent functional dependencies can be carried out in dependence logic. A possible direction for future research on this topic is finding out whether the embedding of $\mathcal{CO}$ can be extended also to other kinds of contingent dependencies (such as the inclusion dependencies of [12], the independence atoms of [14] or the multivalued embedded dependencies of [10]). Of course, since these dependencies are not downward closed, the target language will *not* be dependence logic.

On the other hand, having switched the focus of the discourse from the recursive to the general case, we may want to look at *less common* forms of causal reasoning, which may well *not* be embeddable into classical first-order logic. Halpern ([18]) pointed out that, in the general case, it is natural to consider also the dual conditionals, i.e. the interventionist might-counterfactuals. A might-counterfactual $\mathbf{X} = \mathbf{x} \diamond\!\!\rightarrow \psi$ is true if the system of equations obtained after the intervention $do(\mathbf{X} = \mathbf{x})$ has at least one solution (a tuple of values for the variables) that satisfies ψ. In causal team semantics, it is more natural to give a slightly more general semantic clause:

- $T \models \mathbf{X} = \mathbf{x} \diamond\!\!\rightarrow \psi$ if $S \models \psi$ for some causal subteam S of $T_{\mathbf{X}=\mathbf{x}}$ with $S^- \neq \emptyset$.

It is easy to see that this operator does not preserve downward closure nor the empty team property; therefore it is implausible that $\mathcal{CO}$ enriched with this operator can still be embedded into first-order logic (or even dependence logic). The operator does preserve, instead, closure under unions of causal teams (with identical function components); the ideal candidate for an embedding of this extension of $\mathcal{CO}$ seems then to be inclusion logic with nonemptiness atoms, or some more general union-closed logic ([24]). We leave this problem open for future investigations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. Barbero, G. Sandu, Interventionist counterfactuals on causal teams, in: CREST 2018 Proceedings – Electronic Proceedings in Theoretical Computer Science, vol. 286, Open Publishing Association, Jan 2019, pp. 16–30.
[2] F. Barbero, G. Sandu, Team semantics for interventionist counterfactuals: observations vs. interventions, J. Philos. Log. (2020) 1–51.
[3] F. Barbero, K. Schulz, S. Smets, F.R. Velázquez-Quesada, K. Xie, Thinking about causation: a causal language with epistemic operators, in: M.A. Martins, I. Sedlár (Eds.), Dynamic Logic. New Trends and Applications, Springer International Publishing, Cham, 2020, pp. 17–32.
[4] F. Barbero, K. Schulz, F.R. Velázquez-Quesada, K. Xie, Observing interventions: a logic for thinking about experiments, J. Log. Comput. (April 2022).
[5] F. Barbero, F. Yang, Counterfactuals and dependencies on causal teams: expressive power and deduction systems, in: N. Olivetti, R. Verbrugge, S. Negri, G. Sandu (Eds.), Advances in Modal Logic, United Kingdom, vol. 13, College Publications, July 2020, pp. 73–93.

[6] F. Barbero, F. Yang, Characterizing counterfactuals and dependencies over causal and generalized causal teams, Notre Dame J. Form. Log. (2022), in press.
[7] E. Börger, E. Grädel, Y. Gurevich, The Classical Decision Problem, Springer Science & Business Media, 2001.
[8] R. Briggs, Interventionist counterfactuals, Philos. Stud. Int. J. Philos. Anal. Trad. 160 (1) (2012) 139–166.
[9] A. Durand, M. Hannula, J. Kontinen, A. Meier, J. Virtema, Approximation and dependence via multiteam semantics, in: Proceedings of the 9th International Symposium on Foundations of Information and Knowledge Systems, in: LNCS, vol. 9616, Springer, 2016, pp. 271–291.
[10] F. Engström, Generalized quantifiers in dependence logic, J. Log. Lang. Inf. 21 (2012) 299–324.
[11] D. Galles, J. Pearl, An axiomatic characterization of causal counterfactuals, Found. Sci. 3 (1) (Jan 1998) 151–182.
[12] P. Galliani, Inclusion and exclusion dependencies in team semantics - on some logics of imperfect information, Ann. Pure Appl. Log. 163 (1) (2012) 68–84.
[13] P. Galliani, Dependence logic, in: E.N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, Summer 2021 ed., Metaphysics Research Lab, Stanford University, 2021.
[14] E. Grädel, J. Väänänen, Dependence and independence, Stud. Log. 101 (2013) 399–410.
[15] E. Grädel, Decidable fragments of first-order and fixed-point logic - from prefix vocabulary classes to guarded logics, in: Proceedings of Kalmár Workshop on Logic and Computer Science, Szeged, 2003.
[16] Y. Gurevich, The decision problem for standard classes, J. Symb. Log. 41 (2) (1976) 460–464.
[17] J. Halpern, Actual Causality, MIT Press, 2016.
[18] J.Y. Halpern, Axiomatizing causal reasoning, J. Artif. Intell. Res. 12 (1) (May 2000) 317–337.
[19] J.Y. Halpern, From causal models to counterfactual structures, Rev. Symb. Log. 6 (2) (2013) 305–322.
[20] M. Hannula, Å. Hirvonen, J. Kontinen, V. Kulikov, J. Virtema, Facets of distribution identities in probabilistic team semantics, in: European Conference on Logics in Artificial Intelligence, Springer, 2019, pp. 304–320.
[21] M. Hannula, J. Kontinen, M. Lück, J. Virtema, On quantified propositional logics and the exponential time hierarchy, in: Electronic Proceedings in Theoretical Computer Science, vol. 226, Sep 2016, pp. 198–212.
[22] C. Hitchcock, The intransitivity of causation revealed in equations and graphs, J. Philos. 98 (June 2001) 273–299.
[23] W. Hodges, Compositional semantics for a language of imperfect information, Log. J. IGPL 5 (1997) 539–563.
[24] M. Hoelzel, R. Wilke, On the union closed fragment of existential second-order logic and logics with team semantics, in: M. Fernández, A. Muscholl (Eds.), 28th EACSL Annual Conference on Computer Science Logic (CSL 2020), in: Leibniz International Proceedings in Informatics (LIPIcs), vol. 152, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2020, 25.
[25] R. Iemhoff, F. Yang, Structural completeness in propositional logics of dependence, Arch. Math. Log. 55 (7–8) (2016) 955–975.
[26] J. Kontinen, Coherence and computational complexity of quantifier-free dependence logic formulas, Stud. Log. 101 (2) (2013) 267–291.
[27] J. Kontinen, A. Kuusisto, J. Virtema, Decidability of predicate logics with team semantics, in: P. Faliszewski, A. Muscholl, R. Niedermeier (Eds.), 41st International Symposium on Mathematical Foundations of Computer Science (MFCS 2016), in: Leibniz International Proceedings in Informatics (LIPIcs), vol. 58, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2016, 60.
[28] J. Pearl, Causality: Models, Reasoning, and Inference, Cambridge University Press, New York, NY, USA, 2000.
[29] P. Spirtes, C. Glymour, R.N. Scheines, Causation, Prediction, and Search, Lecture Notes in Statistics, vol. 81, Springer, New York, 1993.
[30] R.H. Strotz, H.O. Wold, Recursive vs. nonrecursive systems: an attempt at synthesis (part I of a triptych on causal chain systems), Econometrica J. Economet. Soc. (1960) 417–427.
[31] J. Väänänen, Dependence Logic: A New Approach to Independence Friendly Logic, London Mathematical Society Student Texts, vol. 70, Cambridge University Press, 2007.
[32] H. van Ditmarsch, W. van der Hoek, B. Kooi, Dynamic Epistemic Logic, Synthese Library Series, vol. 337, Springer, Dordrecht, The Netherlands, 2008.
[33] J. Woodward, Making Things Happen, Oxford Studies in the Philosophy of Science., vol. 114, Oxford University Press, 2003.
[34] F. Yang, J. Väänänen, Propositional logics of dependence, Ann. Pure Appl. Log. 167 (7) (2016) 557–589.
[35] J. Zhang, A Lewisian logic of causal counterfactuals, Minds Mach. 23 (1) (2013) 77–93.