



Team Semantics for Interventionist Counterfactuals: Observations vs. Interventions

Fausto Barbero¹ · Gabriel Sandu¹

Received: 31 August 2018 / Accepted: 7 September 2020 / Published online: 19 December 2020
© The Author(s) 2020

Abstract

Team semantics is a highly general framework for logics which describe dependencies and independencies among variables. Typically, the (in)dependencies considered in this context are properties of sets of configurations or data records. We show how team semantics can be further generalized to support languages for the discussion of interventionist counterfactuals and causal dependencies, such as those that arise in manipulationist theories of causation (Pearl, Hitchcock, Woodward, among others). We show that the “causal teams” we introduce in the present paper can be used for modelling some classical counterfactual scenarios which are not captured by the usual causal models. We then analyse the basic properties of our counterfactual languages and discuss extensively the differences with respect to the Lewisian tradition.

Keywords Team semantics · Interventionist counterfactual · Dependence logic · Causation · Manipulationist theories · Structural equation models

1 Introduction

1.1 The Goal of the Paper

We shall introduce a logical framework which can express both functional and (interventionist) counterfactual dependence. The former is an “accidental” correlation formally defined in terms of the existence of a functional correspondence between the relevant variables. The latter is a more robust correlation expressed by a family of special conditionals whose antecedent is made true by an intervention. In order to express both kinds of dependence we shall merge two existing approaches: a) team semantics, a framework developed recently for the purpose of analyzing the interaction

✉ Fausto Barbero
fausto.barbero@helsinki.fi

¹ Department of Philosophy, History and Art studies, University of Helsinki, Unioninkatu 40, 00014, Helsinki, Finland

of functional dependence with logical constructs [4, 12, 14, 23, 30, 40]; and b) causal models, a popular framework which analyses causal dependencies in terms of structural equations relating appropriately chosen variables [15, 17, 20, 32, 33, 38].

The main notion which emerges from the combination of the two approaches is that of a causal team (Section 2). Essentially, a causal team is a team, defined as a set of assignments over a given collection of variables, expanded with a collection of structural equations associated with some of its variables. Due to its team component, a causal team will be able to account for functional correlations, which are the basis of some of our evidential, observational reasoning; and due to its structural equations component, a causal team will be able to support counterfactual claims. We will define two operations on causal teams which will allow us to capture the difference between these two types of reasoning, each corresponding to one kind of conditional. The first operation consists of restricting a causal team to a *subteam* whose assignments satisfy a given formula of the language. It allows us to define a conditional, which we call *selective implication*, which generalizes material implication. The second operation is that of an *intervention* on a causal team, which leads to a new causal team whose assignments satisfy a given formula. This will be used to define (a generalization of) *interventionist counterfactuals*. These two operations, which are given a model-theoretic treatment, will be the main ingredients in the definition of “ φ is true in T ” where T is a causal team and φ is a formula of the object language. The focus of the paper is on the cases in which φ expresses a functional correlation or an interventionist counterfactual, possibly conditional on observations. We think that the presence of both selective and counterfactual conditional makes the underlying language more adequate than alternative formalisms (e.g. Pearl’s) for expressing arguments involving causal claims where both types of reasoning are present, e.g.

On occasions where p , in fact, was not the case, q would have accompanied it, had p been the case (von Wright, [43], p. 26).

The probability Q that a subject who died under treatment would have recovered had he or she not been treated is epsilon (Pearl, [32], p. 20).

The model-theoretic analysis of interventions and interventionist counterfactuals is not new in the literature. Similar treatments, with variations in the details, may be found in e.g. [3, 11, 16, 49]. More recent work made use of interventionist counterfactuals as an ingredient for the study of counterfactuals in natural language [8, 35, 36] and for the study of actual (token) causation [15, 17, 20]. However, the expression of the three kinds of dependencies, functional, counterfactual, and observational in one single framework is, to the best of our knowledge, unique in the literature, and it would not be possible without the additional dimension of team semantics. It allows us to investigate some of the logical rules that connect these forms of dependence and to highlight some central differences among them. One major difference appears in their behaviour with respect to interventions, that is, their “stability” or “invariance”. Causal teams will be seen (Section 5) to provide a nice illustration of this difference.

As we said, one of the advantages of causal team semantics is that it supports a richer class of languages than both causal models and team semantics. A natural question is then whether causal teams bring something new also in the context of

traditional (interventionist) counterfactual languages. In one sense, the answer is no: the logic of such languages coincides (modulo minor details) with that which arises over causal models (see Section 4.1 and Appendix C). On the other hand, we observe that moving from causal models to causal teams seems to be a necessary step in order to model, in an interventionist spirit, some counterfactual scenarios that arise from the literature; we consider two such examples in Section 3.

Finally, let us mention one more goal of the paper. Even though the axiomatics of both the Lewis-Stalnaker and the interventionist counterfactuals are well-enough understood, it seems to us that a thorough comparison of the two counterfactual logics is still missing. We will advance this enterprise by furthering the investigation – initiated in [3] – of which Lewisian laws fail for *nested* interventionist counterfactuals (Section 4.2); and clarify to what extent some non-Lewisian laws (permutation, importation and exportation) are sound for interventionist counterfactuals (Section 4.3).

A summary of some of the ideas presented in the paper appeared in conference proceedings [2], but the present paper largely extends those ideas in both scope and depth. The omitted details of some of the proofs can be found in the Arxiv draft [1].

1.2 Background

In the philosophical literature (Goodman, Stalnaker, D. Lewis, G. H. von Wright, among others) one increasingly finds evidence that robust causal relationships are conceptually related to counterfactual assumptions. However, the exact nature of this connection is still much debated. Some authors claim that this relationship is a form of reduction: for example, D. Lewis in [28] claims that causation can be defined in terms of an appropriate form of counterfactual dependence, which is in turn definable in non-causal terms. In Pearl [32] the direction of the reduction is reversed¹: interventionist counterfactual statements are assigned a meaning in terms of basic causal assumptions, which are encoded by structural equations. This picture is made more complicated by the observation that more sophisticated notions of cause are then definable in terms of the counterfactuals (see [21, 45]); and by Hitchcock's observation that a structural equation can be seen, in turn, as encoding a set of counterfactuals. The approach we follow here is closest to Pearl's treatment of structural equation models, which we now shortly review.

The definition of a structural equation model (SEM) presupposes the choice of two disjoint sets of variables: a set \mathbf{V} of endogenous variables and a set \mathbf{U} of exogenous variables. The latter will represent factors whose causes are not further analysed.² To

¹The idea that causal assumptions are parts of the semantic definition of counterfactuals also underlies some recent developments of *premise semantics*: see [8, 26, 36, 42].

²The word “exogenous” is sometimes used in the literature with different meanings, for example to denote *unobserved* variables. To appreciate the difference, consider the hour displayed on a clock: it is an observable variable, but we may treat it as exogenous if we omit from the model any description of the internal mechanisms of the clock. An unobserved variable, instead, might be used to mark an hypothetical internal component of the clock that is involved in the calculation and display of the hour. We do not discuss unobserved variables in the present paper.

each endogenous variable $X \in \mathbf{V}$, an equation of the form

$$X := f_X(PA_X)$$

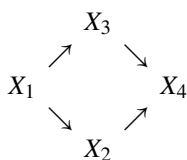
is associated, where the variables in the set $PA_X \subseteq (\mathbf{U} \cup \mathbf{V}) \setminus \{X\}$ are called the *parents of X* . The standard interpretation of a functional equation $X := f_X(PA_X)$ is that of a law which specifies a value of X for any possible combination of the values of PA_X . The explicit specification of the function has the advantage of indicating which changes in PA_X will change X and by how much, something which turned out to be useful in formulating various notions of cause [20, 32, 44].

To each system of equations is associated a *directed graph*, which summarizes the causal dependencies among variables. A (directed) graph is a pair $G = (V, E)$, where V is called the set of *vertices* and $E \subseteq V \times V$ is the set of *arrows*. The directed graph which is associated to the given system of structural equations has as vertices the variables of the system (i.e. $\mathbf{U} \cup \mathbf{V}$); and it contains an arrow from vertex X to vertex Y if and only if $X \in PA_Y$.

We can illustrate these definitions with an example. It is originally due to Daniel Hausman [18] but here we shall present a simplified version borrowed from Woodward [45]. The example is about a salt solution which flows through pipes and mixed chambers. The concentration of salt in each chamber measured by the variables X_1, \dots, X_4 depends on the upstream chambers to which it is connected, according to the following equations

1. $X_2 = aX_1$
2. $X_3 = bX_1$
3. $X_4 = cX_2 + dX_3$.

The system of equations induces the following graph



where, as we see, X_1 is the only exogenous (independent) variable, that is, a variable with no incoming arrows, and X_2, X_3 and X_4 are endogenous (dependent) ones.

Implicit in the specification of the arrows of a graph through a system of structural equations, there is also a very natural means to compute the effect of assuming the antecedent of a counterfactual conditional. This operation, called an *intervention*, takes now the form of a transformation of the relevant set of structural equations. There are several notions of interventions on the market, but the one which is closest in spirit to the definition that will be given in this paper is due to Pearl. According to it, an intervention consists in

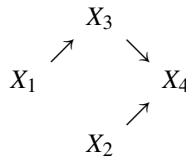
lifting X from the influence of the old functional mechanism $X = f_X(PA_X)$ and placing it under the influence of a new mechanism that sets the value x while keeping all other mechanisms undisturbed. ([32], p. 70; we changed notation).

In other words, an intervention $do(X = x)$ on a variable X is an operation which replaces the equation for X with a new equation $X = x$ while preserving the equations associated with the other variables intact.

In our example, if we intervene on the endogenous variable X_2 and set its value to 1 (i.e., $do(X_2 = 1)$), the result will be the altered system of equations

- 1*. $X_2 = 1$
2. $X_3 = bX_1$
3. $X_4 = cX_2 + dX_3$

which induces the new graph:



Typically a structural equation model is further enriched with a description of the state of the exogenous variables, which may take the form either of a full specification of their actual values (*deterministic SEM*), or of a probability distribution (*semideterministic SEM*). To give an example of the former, we might stipulate that the actual value of the exogenous variable X_1 in our example is some number m . Then, due to the acyclicity of the graph, the system of equations (1)–(3) uniquely determines values for the endogenous variables: $X_2 = am$, $X_3 = bm$ and $Z = cam + dbm$. If we now perform the intervention $do(X_2 = 1)$, then, X_2 depends no more on X_1 and its value becomes 1 (i.e. in Pearl’s terms X_2 has been “lifted from the influence of the old mechanism $X_2 = aX_1$ ”). After the intervention, the values of the other endogenous variables X_3 and X_4 must be recomputed using the value $X_1 = m$ and the new value $X_2 = 1$, thus obtaining $X_3 = bm$ and $X_4 = c + dbm$. This computation may be summarized by stating that the (interventionist) counterfactual

$$X_2 = 1 \Box \rightarrow (X_3 = bm \wedge X_4 = c + dbm)$$

holds in the state where X_1 takes value m .

We take note of the fact that after the intervention $do(X_2 = 1)$, the two other equations of the system, (2) and (3), have remained unaltered, a property referred to in the literature as *modularity*.

2 Causal Team Semantics

2.1 From Teams to Causal Teams

Team semantics was introduced by W. Hodges [23, 24] in order to provide a compositional interpretation of the (game-theoretically defined) semantics of Independence-Friendly logic [19, 30]. It has subsequently been used to extend first-order logic with database dependencies (e.g. Dependence logic [40], Independence logic [14], Inclusion logic [12]), and similar extensions have been proposed for propositional logics

[47, 48] and modal logics [41]. Generalizations of teams have been used as descriptive languages for probabilistic dependencies [10], for the expression of quantum phenomena [25], and for the modelisation of Bayesian networks [9].

The basic idea of team semantics is that notions such as dependence and independence, which express properties of relations (instead of individuals), cannot be captured by Tarskian semantics, which evaluates formulas on single assignments³. The appropriate unit for semantic evaluation is instead the *team*, i.e., a *set* of assignments (all sharing a common variable domain). In our context, a team can be seen as a description of the configurations that are available in a given system of variables; or, under an epistemic interpretation, as a description of our uncertainty over the actual configuration of the system.

In the most common versions of team semantics, all the values of the variables come from a unique domain of individuals associated with an underlying model. However, in order to model causal and counterfactual dependence, we shall need to relax that assumption. Instead we shall take variables to represent properties which may have all kinds of values (as specified by their range). In many applications, we will deal with propositional variables whose range is $\{0, 1\}$. By a *signature* σ we will mean a pair (Dom, Ran) , where Dom is a nonempty finite set (whose elements will be called *variables*) and Ran is a function that associates to each variable $X \in Dom$ a nonempty finite⁴ set $Ran(X)$ of *values* (the *range* of X). An assignment of signature σ will be a mapping $s : Dom \rightarrow \bigcup_{X \in Dom} Ran(X)$ such that $s(X) \in Ran(X)$ for each $X \in Dom$. A **team** T of signature σ will be any set of such assignments.

Here is an example of a team $T = \{s_1, s_2\}$, with $Dom_T = \{X, Y, Z\}$ and range $\{0, 1\}$ for each variable:

	X	Y	Z
s_1	0	0	0
s_2	1	1	0

It is worth noting that such a special “Boolean” team (with binary range for each variable) can as well be identified with a possible-world model, in an obvious way.

The basic semantic relation is $T \models \psi$: the team T satisfies the formula ψ , where ψ is a formula of a formal language to be fully specified later. We are particularly interested in formulas called *dependence atoms* which express the existence of a functional dependency among variables. The semantics of such *dependence atoms* is given by

$$T \models (X_1, \dots, X_n; Y) \text{ if and only if}$$

$$\text{for all } s, s' \in T, \text{ if } s(X_1) = s'(X_1), \dots, s(X_n) = s'(X_n), \text{ then } s(Y) = s'(Y)$$

expressing that Y is functionally determined by $\{X_1, \dots, X_n\}$. Thus, in our example, it holds that $T \models (X; Z)$ but $T \not\models (Z; X)$. In the special case that $n = 0$, we simply write $\models (Y)$ for $\models (; Y)$ and call it a *constancy atom*; notice that $T \models (Y)$ if and only if Y takes the same value in all the assignments of T^- .

As explained earlier, a proper treatment of causal notions requires an account of counterfactual information that is customarily encoded by structural equations. A

³This can be formally proved, see [5].

⁴We will relax the restriction to finite ranges in one later result, Theorem 2.6.

team, by itself, encodes no counterfactual information at all; it does not tell how intervening on a variable will affect the whole system. A “causal team” will have to encode in some way the laws that link together the variables of the system. To exemplify this point, consider a system with two variables, related by the structural equation $Y := X$, which amounts to saying that, whenever X comes to take a specific value x , then also Y comes to take value x , as can be ascertained by applying the identity function to x . Under the assumption that the values 0,1 are in the ranges of X and Y , the following assignments are compatible with the equation:

$X \rightarrow Y$	
0	0
1	1

The arrow from X to Y is drawn to underline the equational dependence of Y on X . Notice that this team satisfies two functional dependencies: $=(X; Y)$, which is specified by the appropriate structural equation, and $=(Y; X)$, which goes in the opposite direction, but is not so specified. Dependencies of this latter kind turn out to be contingent, in the sense that they can be disrupted by some interventions. We will see a few such examples throughout the paper; see in particular Example 2.7 at the end of Section 2.4.

We now extend teams with structural equations by defining “causal teams”; later, in Section 2.4, we will define the notion of *intervention on a causal team*.⁵ We use boldface letters, e.g. \mathbf{X} (resp. \mathbf{x}) to denote sets of variables (resp. sets of values).

Definition 2.1 A **causal team** T of signature (Dom_T, Ran_T) with endogenous variables $\mathbf{V} \subseteq Dom_T$ is a triple $T = (T^-, G_T, \mathcal{F}_T)$, where:

1. T^- is a team of domain Dom_T (the *team component* of T).
2. $G_T = (Dom_T, E)$ is a graph over the set of variables. For any $X \in Dom_T$, we denote as PA_X the set of all variables $Y \in Dom_T$ such that the arrow (Y, X) is in E . We require the graph to be *irreflexive*, i.e. not to have arrows of the form (X, X) .
3. \mathcal{F}_T is a function $\{(V_i, f_{V_i}) \mid V_i \in \mathbf{V}\}$ that assigns to each endogenous variable a $|PA_{V_i}|$ -ary function $f_{V_i} : Ran_T(PA_{V_i}) \rightarrow Ran_T(V_i)$

which satisfies the further consistency constraint:

- (*) For all $s \in T^-$, and all $Y \in \mathbf{V}$, $s(Y) = f_Y(s(PA_Y))$.

The functional component \mathcal{F}_T induces an associated system of structural equations, say one equation

$$Y := \mathcal{F}_T(Y)(PA_Y)$$

for each variable $Y \in \mathbf{V}$. The graph associated to this system of structural equations is G_T . The variables in $Dom_T \setminus \mathbf{V}$ will be called, as usual, *exogenous*.

⁵For simplicity we take our “causal teams” to be *fully specified*, i.e., they provide a full description of the functional equations. Some of the issues that arise when the functions are not fully specified are handled in [2], Section 5.

Clause (*) ensures that the team component satisfies the functional dependencies encoded in \mathcal{F}_T . It further entails a compatibility condition between the team and the graph component:

- a) If $PA_Y = \{X_1, \dots, X_n\}$, then $T^- \models (X_1, \dots, X_n; Y)$.

Notice also that our previous (slightly nonstandard) definition of team entails a compatibility condition between the team and the range components; writing $T^-(X) = \{s(X) \mid s \in T^-\}$ for the set of values that X takes in the assignments of T^- , we have:

- b) $T^-(X) \subseteq \text{Ran}_T(X)$ for each $X \in \text{Dom}_T$.

In this paper we restrict attention to causal teams T whose associated graph G_T is acyclic; such causal teams will be called **recursive**.⁶ A discussion of the developments beyond this restriction can be found in [2].

Example 2.2 We illustrate the previous definition by exhibiting a causal team which captures a “coin tossing” scenario. This is just a toy model, but we will return to the issue of coin tossing in Section 3, as it relates to some famous example discussed by Lewis.

Let us say that you toss a coin, and if it comes up heads, you win; otherwise you lose. First we have to decide what variables are relevant for the description of this scenario. We choose the following:

- U : whether you toss the coin. We treat it as a Boolean variable: it takes value 1 if the coin is tossed and value 0 if the coin is not tossed. So, $\text{Ran}(U) = \{0, 1\}$.
- C : what face of the coin is up: heads (h) or tails (t). (For simplicity: in case the coin is tossed, the variable represents the outcome after tossing. Otherwise, it represents what face of the coin is up in your hand). $\text{Ran}(C) = \{h, t\}$.
- W : whether you win (1) or not (0). $\text{Ran}(W) = \{0, 1\}$.

Notice first that the pair $\sigma = (\text{Dom}, \text{Ran})$, where $\text{Dom} = \{U, C, W\}$ and Ran is the function associating to each of these three variables the range described above, is a signature.

We can now for example describe all possible configurations of this system by the following causal team T of signature σ :

$$T : \begin{array}{c} \begin{array}{|c|c|c|} \hline U & \overset{\curvearrowright}{C \rightarrow} & W \\ \hline 1 & h & 1 \\ \hline 1 & t & 0 \\ \hline 0 & h & 0 \\ \hline 0 & t & 0 \\ \hline \end{array} \end{array}$$

Each row of the table describes, in an obvious way, an assignment of T^- . The arrows between variables are the edges of G_T . They tell us that U and C are exogenous variables, while W is a function of both U and C . We still have to describe this

⁶This slightly eccentric terminology is taken from the literature on structural equation models.

function. From our narrative, it should be clear that the variables are related by one structural equation: $W := 1$ if and only if $U = 1$ and $C = h$; and 0 otherwise. We encode this in the causal team by stipulating that $\mathcal{F}_T(W)$ is the function that returns 1 when U is 1 and C is h ; and returns 0 otherwise.

We might instead be only interested in modelling the case in which you really toss the coin. This scenario is represented by the smaller team S :

$$S : \begin{array}{|c|c|c|} \hline U & C & W \\ \hline 1 & h & 1 \\ \hline 1 & t & 0 \\ \hline \end{array}$$

Notice that, in this smaller team, the state of the coin (C) determines whether you win or not (W), that is, the dependence atom $\models(C; W)$ is true in S . This was not the case in the larger team T .

2.2 A Basic Language and its Semantics

Before discussing interventions and counterfactuals, we need to specify what it means for a causal team to satisfy atomic formulas and their boolean combinations. Let us fix a signature $\sigma = (Dom, Ran)$. Assuming $\mathbf{X}, Y \in Dom$ and $y \in Ran(Y)$, the kind of language we consider, for now, contains dependence atoms of the form $\models(\mathbf{X}; Y)$, with $\mathbf{X}, Y \in Dom$; atomic formulas of the forms $Y = y$ and $Y \neq y$, where $Y \in Dom$ and $y \in Ran(Y)$; connectives \wedge and \vee . Formulas without occurrences of dependence atoms will be called *classical*. We take $Y = y$ to mean: the value of the variable Y is y . Notice that here, in order to reduce the weight of notations, we are identifying the letter Y , used as a symbol of the object language, with the variable $Y \in Dom$; and y , used as a symbol of the object language, with the object $y \in Ran(Y)$ that it is meant to denote. This convention is usually followed in the literature on structural equation models, and is typically not a source of confusion.⁷

We assign to these formulas semantic clauses which are standard in the literature:

- $T \models \models(\mathbf{X}; Y)$ if and only if for all $s, s' \in T$, $s(\mathbf{X}) = s'(\mathbf{X})$ implies $s(Y) = s'(Y)$.
- $T \models Y = y$ if and only if, for all $s \in T^-$, $s(Y) = y$.
- $T \models Y \neq y$ if and only if, for all $s \in T^-$, $s(Y) \neq y$.
- $T \models \psi \wedge \chi$ if and only if $T \models \psi$ and $T \models \chi$.

The disjunction \vee will be treated as what is sometimes called “tensor” or “splitjunction”; its semantic clause requires the notion of causal *subteam*.

⁷If needed, syntax and semantics can be kept strictly separated in a straightforward way. It suffices to specify two disjoint sets of variables (say, the set *Capital* of Y_i s and the set *Small* of y_i s) and add to the definition of causal team an interpretation function I which associates to each variable in *Capital* a variable in *Dom*, and to each variable in *Small* a value in $\bigcup Ran$. Then formulas $Y = y$ and $Y \neq y$ are part of the language in case $Y \in Capital$, $y \in Small$ and $I(y) \in Ran(I(Y))$; the truth condition, say, for formula $Y = y$ over team T will then be “for all $s \in T^-$, $s(I(Y)) = I(y)$ ”.

Definition 2.3 Given a causal team T , a **causal subteam** S of T is a causal team with the same domain and the same set of endogenous variables, which satisfies the following conditions: 1) $S^- \subseteq T^-$, 2) $G_S = G_T$, 3) $\mathcal{F}_S = \mathcal{F}_T$.

Under the epistemic interpretation of teams, the transition from a team to a subteam reflects an increase in our knowledge of the actual state of the system, given that fewer configurations (assignments) are considered possible.

- $T \models \psi \vee \chi$ if and only if there are two causal subteams T_1, T_2 of T such that $T_1^- \cup T_2^- = T^-$, $T_1 \models \psi$ and $T_2 \models \chi$.⁸

The tensor disjunction extends classical disjunction, in the sense that, over a causal team with singleton team component, it respects the usual Tarskian clause ($T \models \psi \vee \chi$ if and only if $T \models \psi$ or $T \models \chi$).

The next two subsections are devoted to extending the basic language with two additional conditional operators. Section 2.5 will summarize the syntactical details of the resulting language.

2.3 Selective Implication

In the literature, counterfactual statements often appear embedded in a context which describes facts or observations. We have seen (Section 1.1) that von Wright ([43]) considers examples of the form “*on occasions where p , in fact, was not the case, q would have accompanied it, had p been the case*”. Pearl ([32]) analyses queries of the form: “the probability Q that a subject *who died under treatment* ($X = 1, Y = 1$) would have recovered ($Y = 0$) had he or she not been treated ($X = 0$) is epsilon”. We will represent this statement as:

$$(X = 1 \wedge Y = 1) \supset (X = 0 \Box\!\!\rightarrow Pr(Y = 0) = \epsilon)$$

where the symbol $\Box\!\!\rightarrow$ stands for counterfactual implication, while the *selective implication* \supset indicates a restriction of the range of application of the statement to the available evidence; that is, it asserts the truth of the consequent in the subteam generated by those assignments which satisfy the antecedent.⁹ The probabilistic atom describes a probability which is relative to the specific team which is produced during the evaluation of the formula, i.e., the initial team restricted to assignments which satisfy $X = 1 \wedge Y = 1$ and further intervened upon by setting X to 0.

Given a causal team T , and a *classical* formula ψ (that is, a formula as in Subsection 2.2, but without dependence atoms), we define the subteam T^ψ of T by the condition:

- $(T^\psi)^- = \{s \in T^- \mid \{s\} \models \psi\}$.

⁸Defining the union of any pair of *causal* teams sharing the same domain may lead to some problems, as the information encoded by the invariant functions of one team might be incompatible with the information encoded by the other.

⁹The semantics of probabilistic atoms can be found in [2], while a proof of the correctness of this decomposition is available in the Arxiv preprint [1], sec. 6.3; it is part of an intended continuation of this paper.

where we abuse notation (as will be done elsewhere in the paper) by using $\{s\}$ to denote the *causal* subteam S such that $S^- = \{s\}$.

Then, we define¹⁰ selective implication by the clause:

- $T \models \psi \supset \chi$ if and only if $T^\psi \models \chi$.

The consequent χ can be any formula of our current logical language, including formulas which do not denote properties of single assignments. The underlying idea behind the restriction for the antecedents to be classical is that selective implication is a reasonable operator only for antecedents that are *flat* formulas, in the sense with which this word is used in the literature on logics of dependence. (Roughly, it means that asserting the truth of the formula over a team amounts to asserting its truth over each assignment of the team). Later on in Section 2.5 we will also allow for antecedents with occurrences of interventionist counterfactuals (which will turn out to be flat). Selective implication generalizes material implication in the sense that, over causal teams with singleton team component, it respects the usual Tarskian clause ($T \models \psi \supset \chi$ if and only if $T \models \psi$ or $T \models \chi$).

Example 2.4 For an illustration, let us look back at the teams T, S considered in Example 2.2. S was obtained by selecting the assignments in which the variable U (“whether the coin is tossed”) takes value 1; therefore, S is none other than $T^{U=1}$. We have pointed out that $S \models (C; W)$, i.e., $T^{U=1} \models (C; W)$. Then, by the clause for \supset , we may conclude that $T \models U = 1 \supset (C; W)$. In other words, the information encoded in T agrees with the statement that “In case the coin is tossed, the outcome of the the tossing completely determines whether I win”.

2.4 Interventions on Causal Teams

The semantic interpretation of (interventionist) *counterfactual implication* requires, as pointed out, the notion of *intervention on a causal team*. Informally, the idea is that a counterfactual $X = x \Box \rightarrow \psi$ is true in the causal team T if ψ is true in the causal team which results from an intervention $do(X = x)$ on the team T .

We start by discussing the motivation behind this notion. Perhaps the best way to proceed is by referring first to the notion of intervention $do(X = x)$ in which a single variable X is set to a particular value $X = x$ in the structural equations framework. In our introductory section, we have seen that such an intervention amounts to replacing the equation $X = f_X(PA_X)$ with the new equation $X = x$. In the graph corresponding to the system of equations, the intervention $X = x$ breaks all the arrows directed into X but leaves intact the arrows going out of X .

In the context of causal teams, there is one additional complication due to the fact that a causal team is a triple $T = (T^-, G_T, \mathcal{F}_T)$. An intervention, say $do(X = x)$, is, roughly, an operation on T which affects all the three ingredients of the team,

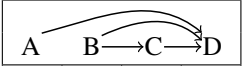
¹⁰As far as we know, this operator has appeared for the first time, with a different notation, in [13]. It is a special case of the so-called *maximal implication* that was introduced in [27]. As long as only downward-closed languages are concerned (see below), it can also be seen as a special case of the *inquisitive implication* of [7].

resulting in a new causal team. The modification of \mathcal{F}_T and G_T goes exactly as in the case of the structural equation framework, described in Section 1.2. The new element is the specification of the way in which the team T^- is affected. It will take the form of an algorithm, which ensures that the resulting team matches with the modified system of equations. Before defining it, we give an example. This is a purely numerical example, whose main purpose is to illustrate the complexities involved in the computation of the effects of an intervention.


Example 2.5 We consider a causal team over variables A, B, C, D with A and C Boolean, while $Ran(B) = Ran(D) = \{0, 1, 2\}$. We assume that the variables respect the equations:

$$\begin{cases} C := 1 \text{ if } B = 2, \text{ and } 0 \text{ otherwise} \\ D := A + C \text{ if } B \text{ is different from } 0, \text{ and } 0 \text{ otherwise.} \end{cases}$$

There are $card(Ran(A)) \times card(Ran(B)) = 2 \times 3 = 6$ assignments compatible with these equations; we include in our team T only the three assignments depicted below. We want to understand what new causal team is produced by the intervention $do(B = 2)$ which fixes B to the value 2. In any case, all assignments in the new team should satisfy $B = 2$:

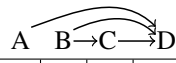
				
T :	0	2	1	1
	1	1	0	1
	1	0	0	0

 \rightsquigarrow

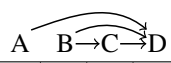
				
	0	2
	1	2
	1	2

The picture also reflects our expectation that, since A is not a descendant of B in the graph, the A -column should be unaffected by the intervention. On the other side, the variables C and D are descendants of B , and must therefore be updated using the functional component \mathcal{F}_T of the causal team.

We cannot first update the D -column, because D is determined as a function of A, B and C , which requires that we first compute the values of $C := f_F(B)$:

				
	0	2	1	...
	1	2	1	...
	1	2	1	...

Finally, we update the D column and we obtain the team $T_{B=1}$:

				
	0	2	1	1
	1	2	1	2

Notice that the correct value for D in the first row might have been read directly from the first row of T ; computing the correct value in the second row, instead, essentially requires using the function component of the causal team.

Now the team $T_{B=2}$ satisfies $C = 1$; we conclude that the initial causal team T satisfies the counterfactual $B = 2 \Box \rightarrow C = 1$.

This example also illustrates the way in which some functional dependencies may be merely contingent: T satisfies $\models (B; D)$, but the intervention $do(B = 1)$ destroys this dependence.

Until now we only considered an intervention over an exogenous variable; such an intervention preserves the graph component of the causal team. Consider now the intervention $do(D = 0)$:

$$T_{D=0}:$$

A	B	$B \rightarrow C$	D
0	2	1	0
1	1	0	0
1	0	0	0

The resulting team has a much simpler graph, given that the intervention removed all the arrows that come into D (and, correspondingly, the function $\mathcal{F}_T(D)$).

The example showed how we can produce algorithmically the causal team generated by the intervention $do(X = x)$ on the causal team T .¹¹ This was possible due to T being recursive, i.e., its graph component being acyclic. We will stick to this assumption throughout the paper, with the exception of Section 3.2.

More formally, an intervention $do(X = x)$ on a recursive causal team $T = (T^-, G_T, \mathcal{F}_T)$ (of endogenous variables \mathbf{V}) generates a new causal team $T_{X=x} = ((T_{X=x})^-, G'(T), \mathcal{F}'_T)$ (of endogenous variables $\mathbf{V} \setminus \{X\}$) such that

1. $\mathcal{F}'_T(Y) = \mathcal{F}_T(Y)$ for every $Y \in \mathbf{V} \setminus \{X\}$ ¹²
2. The graph $G'(T)$ is induced by $\mathcal{F}'_T(X)$ (i.e., the incoming arrows to X are removed)
3. The team $(T_{X=x})^-$ is produced in stages: we first replace each assignment $s \in T^-$ with $s(x/X)$ and then compute how this replacement affects the descendants of X in $G'(T)$ using the functions in \mathcal{F}'_T .¹³

The exact formulation of the algorithm in 3. requires, as the example showed, the introduction of a distance $d(X, Y)$ between X and the other variables $Y \in \text{Dom}_T$. The

¹¹The fact that interventions on a deterministic, recursive structural equation model can be computed in stages is implicit in the literature on causal inference, although we are not aware of it being described in detail. [35] anticipates us in making it explicit (in a different context) how an intervention affects the values of variables in a single assignment. Instead, we have not found in the literature any explicit reference to the role played, in the computation, by the “distance” between variables.

¹²We might have, equivalently, kept X as an endogenous variable, and replaced its function with the function that constantly picks \mathbf{x} . The option we choose here, to deprive X of a function and treat it as exogenous, seems to be more elegant (in the present context).

¹³The function describing X , which has been removed from \mathcal{F}'_T is not needed for recomputing the values of the descendants of X . We only need the new constant value of X as imposed by the intervention.

idea behind this notion of distance is that, if there is a path of length n from X to Y , then Y is *at least* at distance n from X ; the details can be found in Appendix A. We only need the following crucial property of the distance (which holds in finite acyclic graphs):

(R): If $d(X, Y) = n + 1$, then, for each parent Z of Y , $d(X, Z) \leq n$.

Here is the algorithm:

- At stage 0 we produce the team T_0 by replacing each assignment $s \in T^-$ with $s(x/X)$.
- At stage $n+1$ we produce the team T_{n+1} by updating the variables Y such that $d(X, Y) = n + 1$. This happens by replacing each assignment $s \in T_n$ with the assignment s' which is identical to s except that s' assigns to every such Y a unique value $s'(Y) = \mathcal{F}'_T(Y)(s(PA_Y))$.
- The procedure ends at stage

$$\hat{n} = \sup \{d(X, Z) \mid Z \in \text{Dom}_T\}.$$

- We take $(T_{X=x})^- := T^{\hat{n}}$.

Stage $n + 1$ is correctly defined thanks to property (R), which guarantees that all the values of variables in PA_Y have already been recomputed (if needed) in earlier stages of the algorithm.

The same algorithmic procedure can be generalised to the case of simultaneous interventions over multiple variables. We will write $\mathbf{X} = \mathbf{x}$ for a generic conjunction of the form $X_1 = x_1 \wedge \dots \wedge X_n = x_n$, and we will say that such a conjunction is *consistent* if it does not contain two conjuncts $X_i = x_i, X_i = x'_i$ with $x_i \neq x'_i$ (and *inconsistent* otherwise). If $\mathbf{X} = \mathbf{x}$ is a consistent conjunction, we will write $do(\mathbf{X} = \mathbf{x})$ for the intervention that fixes each variable X_i to the corresponding value x_i . In this case the distance $d(\mathbf{X}, Y)$ will have to be reformulated appropriately, as shown in the Appendix.

Given the acyclicity and finiteness of the graph G_T , it is straightforward to show that this algorithm terminates (even in case T^- is infinite).

Theorem 2.6 *Let T be a recursive causal team. If G_T is a finite acyclic graph, then $T_{\mathbf{X}=\mathbf{x}}$ is well-defined.*

See Appendix B for the proof.

Having defined the causal team $T_{\mathbf{X}=\mathbf{x}}$, we are in the position to formulate the semantic clause for counterfactuals of the form $\mathbf{X} = \mathbf{x} \Box \rightarrow \psi$:

$$T \models \mathbf{X} = \mathbf{x} \Box \rightarrow \psi \iff T_{\mathbf{X}=\mathbf{x}} \models \psi.$$

If the antecedent is inconsistent (i.e., it contains two conjuncts $X_i = x_i, X_i = x'_i$ with $x_i \neq x'_i$), the corresponding intervention is not defined; in this case, we postulate the counterfactual to be (trivially) true.

We may now vindicate our earlier claim that functional dependencies which are not backed up by structural equations can be disrupted by interventions.

Example 2.7 Consider again the causal team T with two boolean-valued variables X and Y , related by the structural equation $Y := X$ (i.e. $\mathcal{F}_T(Y)(x) = x$ for each $x \in \text{Ran}(X)$) and with assignments as in the table:

$X \rightarrow Y$	
0	0
1	1

As already observed, this team satisfies both the dependency $=(X; Y)$, which is required for compatibility with the structural equation, and the dependency $=(Y; X)$, which is not supported by the structural equations. When we intervene on the system by fixing Y to value 1, a new table is produced:

X	Y
0	1
1	1

This team, manifestly, does not satisfy the dependency $=(Y; X)$ anymore.

Notice also that, although we have removed the arrow in order to specify that the value of X is not anymore needed to reconstruct the value of Y , the dependency $=(X; Y)$ still holds. This latter dependency is logically implied by the constancy of Y , $=(Y)$, which also holds in the new table.

2.5 Logical Languages

Having defined all the main logical operators that we consider in this paper, we move now to the definition of some languages.

For a fixed signature $\sigma = (Dom, Ran)$, we call *counterfactual-observational* language of signature σ , $\mathcal{CO}(\sigma)$, the language formed by the following rules:

$$Y = y \mid Y \neq y \mid \psi \wedge \chi \mid \psi \vee \chi \mid \psi \supset \chi \mid \mathbf{X} = \mathbf{x} \Box \rightarrow \chi$$

for $Y, \mathbf{X} \in Dom$, $y \in Ran(y)$, $\mathbf{x} \in Ran(\mathbf{X})$, ψ, χ formulas of $\mathcal{CO}(\sigma)$. We will often simply write \mathcal{CO} , omitting reference to the specific signature. The semantics for this language is given by the clauses specified in the previous subsections. Notice that the clause for $\psi \supset \chi$ is still well defined now that occurrences of \supset and $\Box \rightarrow$ are allowed in the antecedent and consequent. We will refer to the sublanguage of $\mathcal{CO}(\sigma)$ without the symbol \supset as the language $\mathcal{C}(\sigma)$.

We will call $\mathcal{COD}(\sigma)$ the language (extending $\mathcal{CO}(\sigma)$) which is formed by the following rules:

$$Y = y \mid Y \neq y \mid =(X; Y) \mid \psi \wedge \chi \mid \psi \vee \chi \mid \theta \supset \chi \mid \mathbf{X} = \mathbf{x} \Box \rightarrow \chi$$

for $Y, \mathbf{X} \in Dom$, $y \in Ran(y)$, $\mathbf{x} \in Ran(\mathbf{X})$, ψ, χ formulas of $\mathcal{COD}(\sigma)$, and, importantly, θ is a formula of $\mathcal{CO}(\sigma)$ (we do not allow dependence atoms in the antecedent of a selective implication).

We will be interested every now and then also in languages that contain an additional connective \sqcup , called Boolean disjunction.¹⁴ In general, given a language \mathcal{L} , we denote by \mathcal{L}_{\sqcup} its extension with \sqcup . The semantics of this operator is given by:

– $T \models \psi \sqcup \chi$ if $T \models \psi$ or $T \models \chi$.

Like the tensor, also the Boolean disjunction behaves classically over causal teams with singleton team components. In general, though, its behaviour is nonclassical.¹⁵

We now list three global properties of the languages introduced so far.¹⁶

Theorem 2.8 *The logic \mathcal{COD}_{\sqcup} is downwards closed, that is: if $\varphi \in \mathcal{COD}_{\sqcup}$, T is a recursive causal team, T' is a causal subteam of T , and $T \models \varphi$, then also $T' \models \varphi$.*

Proof By induction on the syntactical structure of φ . The propositional and atomic cases are routine. For the case of dependence atoms, simply observe that, if $T' \not\models (\mathbf{X}; Y)$, then there are $s, s' \in (T')^-$ such that $s(\mathbf{X}) = s'(\mathbf{X})$ and $s(Y) \neq s'(Y)$. But then these two assignments are also in T^- , from which $T \not\models (\mathbf{X}; Y)$ follows.

Suppose $T \models \psi \supset \chi$. Then by definition $T^\psi \models \chi$. Now $(T')^\psi \subseteq T^\psi$; therefore, by the inductive hypothesis, $(T')^\psi \models \chi$. Thus, by the semantic clause for selective implication, $T' \models \psi \supset \chi$.

Suppose $T \models \mathbf{X} = \mathbf{x} \Box \rightarrow \chi$. Then by definition $T_{\mathbf{X}=\mathbf{x}} \models \chi$. But since the algorithm $do(\mathbf{X} = \mathbf{x})$ acts on each assignment separately, we can conclude that $(T'_{\mathbf{X}=\mathbf{x}})^- \subseteq (T_{\mathbf{X}=\mathbf{x}})^-$. So, by the inductive hypothesis, $T'_{\mathbf{X}=\mathbf{x}} \models \chi$, which yields $T' \models \mathbf{X} = \mathbf{x} \Box \rightarrow \chi$. \square

Under the epistemic interpretation of teams, closure under causal subteams expresses the fact that the properties of the system which are expressible in \mathcal{COD}_{\sqcup} remain stable under our increase of knowledge about the actual state of the system.

Theorem 2.9 *The logic \mathcal{COD}_{\sqcup} has the empty team property, that is: for every $\varphi \in \mathcal{COD}_{\sqcup}$, $\emptyset \models \varphi$. Here \emptyset denotes any recursive causal team with empty team component.*

Proof By induction on the syntax of φ . The new cases are those for \supset and $\Box \rightarrow$ (and the case for \sqcup , which is trivial).

Let φ be $\psi \supset \chi$. Notice that $\emptyset^\psi = \emptyset$. By the inductive hypothesis, $\emptyset^\psi = \emptyset \models \chi$. So $\emptyset \models \psi \supset \chi$.

Let φ be of the form $\mathbf{X} = \mathbf{x} \Box \rightarrow \chi$. Any do algorithm, applied to \emptyset , produces a causal team with empty team component (call it \emptyset'); therefore, by the inductive hypothesis, $\emptyset_{\mathbf{X}=\mathbf{x}} = \emptyset' \models \chi$. Thus, by the semantic clause for counterfactuals we obtain $\emptyset \models \mathbf{X} = \mathbf{x} \Box \rightarrow \chi$. \square

¹⁴This connective has also been called “intuitionistic” or “inquisitive” disjunction.

¹⁵For example, the Boolean disjunction of flat formulas may fail to be flat. This does not happen with the tensor. See Theorem 2.10 for the definition of flatness.

¹⁶These kinds of properties are typically used, in the literature on team semantics, for solving issues of definability or for showing that two distinct languages have different expressive power.

We can show that the language \mathcal{CO} (and thus also \mathcal{C}) satisfies a condition that is much more restrictive than downward closure: flatness. Here and in the following, we will often abuse notation and write, say, $\{s\}$ for a causal team T whose support T^- contains the single assignment s ; the context will make clear what are the other components of this causal team. We will then write $\{s\}_{\mathbf{X}=\mathbf{x}}$ for the causal team $T_{\mathbf{X}=\mathbf{x}}$, and so on.

Theorem 2.10 *The logic \mathcal{CO} over causal teams is flat, that is: for every formula φ of \mathcal{CO} and every recursive causal team T , $T \models \varphi$ iff $\{s\} \models \varphi$ for every assignment $s \in T^-$.*

Proof By induction on the complexity of φ . The cases related to propositional connectives are routine. We only have to check the cases of $\varphi = \mathbf{X} = \mathbf{x} \Box \rightarrow \chi$ and $\varphi = \psi \supset \chi$.

Case $\Box \rightarrow$) Suppose first that $T \models \mathbf{X} = \mathbf{x} \Box \rightarrow \chi$, with χ flat. Then, by the semantic clause for counterfactuals, $T_{\mathbf{X}=\mathbf{x}} \models \chi$. By flatness of χ , $\{t\} \models \chi$ for each $t \in (T_{\mathbf{X}=\mathbf{x}})^-$. Now, notice that, for all $s \in T^-$, $\{s\}_{\mathbf{X}=\mathbf{x}}$ is a singleton causal team (see the details of the *do* algorithm) and its unique assignment is an element of $(T_{\mathbf{X}=\mathbf{x}})^-$. Therefore, by the remarks above, $\{s\}_{\mathbf{X}=\mathbf{x}} \models \chi$ for each $s \in T^-$; so, by the semantic clause for counterfactuals, $\{s\} \models \mathbf{X} = \mathbf{x} \Box \rightarrow \chi$ for every $s \in T^-$.

In the other direction, suppose $\{s\} \models \mathbf{X} = \mathbf{x} \Box \rightarrow \chi$ for every $s \in T^-$. Then for every $s \in T^-$, $\{s\}_{\mathbf{X}=\mathbf{x}} \models \chi$. Again, any such causal team is a singleton set, say $\{s\}_{\mathbf{X}=\mathbf{x}} =: \{t_s\}$. But notice that, if t is any assignment in $(T_{\mathbf{X}=\mathbf{x}})^-$, then $t = t_s$ for some $s \in T^-$. So we can apply the inductive hypothesis and conclude that $T_{\mathbf{X}=\mathbf{x}} \models \chi$. Therefore $T \models \mathbf{X} = \mathbf{x} \Box \rightarrow \chi$.

Case \supset) Suppose $T \models \psi \supset \chi$, with ψ, χ flat. Then, by the semantic clause for \supset , $T^\psi \models \chi$. By the induction hypothesis, $\{s\}^\psi = \{s\} \models \chi$ for each $s \in (T^\psi)^-$. Thus $\{s\} \models \psi \supset \chi$. Suppose instead $s' \in T^-$ is such that $\{s'\} \not\models \psi$. Then $\{s'\}^\psi = \emptyset$. Now $\{s'\}^\psi = \emptyset \models \chi$; so, by the clause for \supset , $\{s'\} \models \psi \supset \chi$. In conclusion, for any $s \in T^-$, $\{s\} \models \psi \supset \chi$.

Vice versa, suppose $\{s\} \models \psi \supset \chi$ for all $s \in T^-$. Let $s \in T^-$ be such that $\{s\} \models \psi$. Then $\{s\}^\psi = \{s\}$; our assumption then yields $\{s\} \models \chi$. Since χ is flat, we can conclude that $T^\psi \models \chi$. So $T \models \psi \supset \chi$. \square

We emphasise once more that, unlike in similar results for team semantics (see e.g. [40]), the truth of a flat formula in a causal team based on a singleton team uses, through structural equations, information that goes beyond that which is encoded in the single assignment of the team. A singleton causal team can be identified with a structural equation model (see Section 1.2) enriched with a description of the “actual” state for all variables in the system. From this perspective, a consequence of the flatness result is a kind of conservativity of the causal team semantics over the structural equation modelling of recursive systems. As long as one works within a sufficiently poor (i.e. flat) language, the truth of a formula in a causal team can be reduced to its truth over (a collection of) enriched structural equation models. However, as soon as non-flat language components such as dependence atoms or Boolean disjunction are introduced, structural equation models become insufficient, and causal teams are

needed. Indeed, it is easy to build counterexamples to flatness using these components. The constancy atom $=(X)$ is true over any causal team whose support is a singleton; but it is false e.g. over the team $S = \{(X, 0), \{(X, 1)\}\}$. The disjunction $X = 0 \sqcup X = 1$ is true over both singleton teams $S^{X=0}$ and $S^{X=1}$, yet it is false in S .

2.6 Dual Negation

We can extend our language \mathcal{CO} (resp. \mathcal{C}) with the dual negation \neg interpreted by the following semantic clause:

- $T \models \neg\varphi \iff \text{for all } s \in T^-, \{s\} \not\models \varphi$.¹⁷

It is immediate to extend the proof of Theorem 2.10 to show that the language \mathcal{CO} extended with dual negation is still a flat language (which further entails that it is also downward closed and has the empty team property).

Once we have the dual negation in the language, we can show that the selective implication $\psi \supset \chi$ is equivalent, like its classical relative, to $\neg\psi \vee \chi$. The argument, which requires 1) downward closure and 2) flatness of the antecedent, goes like this: by the semantic clause for the tensor, $T \models \neg\psi \vee \chi$ holds if and only if there are S_1, S_2 causal subteams of T such that $S_2 \models \chi$ and, for all $s \in S_1^-$, $\{s\} \not\models \psi$. Then $(T^\psi)^- \subseteq S_2$. Now if the logic under consideration is downward closed, we can conclude that $T^\psi \models \chi$, that is, $T \models \psi \supset \chi$. In the opposite direction, we have that $T \models \psi \supset \chi$ implies that $T^\psi \models \chi$, while (by flatness of the antecedent) $T \setminus T^\psi \models \neg\psi$. The existence of such a partition of T proves that $T \models \neg\psi \vee \chi$.

It is easy to find examples of languages that are not downward closed, e.g., by adding a contradictory negation, the *independence atoms* $X \perp Y$ proposed in [14], or probabilistic statements. For these kinds of logics the above syntactical definition of selective implication is in general incorrect. However, $\psi \supset \chi$ is usually still definable, in non-downward-closed logics, by the more complex formula $\neg\psi \vee (\psi \wedge \chi)$ (see [13]).

The dual negation $\neg\varphi$ of a \mathcal{CO} formula φ can be expressed in the language \mathcal{CO} itself. That is, for every \mathcal{CO} formula φ , we can produce a \mathcal{CO} formula φ^d , called the *dual* of φ , such that for any causal team T we have:

$$T \models \varphi^d \text{ iff } T \models \neg\varphi.$$

The formula φ^d is obtained by the following syntactical transformations:

- $(X = x)^d$ is $X \neq x$
- $(X \neq x)^d$ is $X = x$
- $(\psi \wedge \chi)^d$ is $(\psi^d \vee \chi^d)$
- $(\psi \vee \chi)^d$ is $(\psi^d \wedge \chi^d)$
- $(\psi \supset \chi)^d$ is $\psi \wedge \chi^d$
- $(\psi \Box \rightarrow \chi)^d$ is $\psi \Box \rightarrow \chi^d$

Then we have:

Theorem 2.11 (Duality) *Let T be a causal team and $\varphi \in \mathcal{CO}$. Then $T \models \varphi^d$ iff $T \models \neg\varphi$.*

¹⁷This definition is reasonable due to the flatness of \mathcal{CO} , theorem 2.10 above.

Proof By induction on the syntactical complexity of $\varphi \in \mathcal{CO}$. We consider the least obvious cases.

Let $\varphi = \psi \supset \chi$. Then we have the following equivalences: $T \models \varphi^d \iff T = T^\psi$ and $T^\psi \models \chi^d \iff$ (by inductive hypothesis) $T = T^\psi$ and $T \models \neg\chi \iff$ (by definition of \neg) $T = T^\psi$ and for every $s \in T^-$, $\{s\} \not\models \chi \iff$ for every $s \in T^-$, $\{s\}^\psi \not\models \chi \iff$ for every $s \in T^-$, $\{s\} \models \psi \supset \chi \iff T \models \neg(\psi \supset \chi)$.

Let $\psi = \psi \sqsupset \chi$. Then $T \models \varphi^d \iff T \models \psi \sqsupset \chi^d \iff T_\psi \models \chi^d \iff$ (by inductive hypothesis) $T_\psi \models \neg\chi \iff$ for all $s \in (T_\psi)^-$, $\{s\} \not\models \chi \iff$ for all $t \in T^-$, $\{t\}_\psi \not\models \chi \iff$ for all $t \in T^-$, $\{t\} \not\models \psi \sqsupset \chi \iff T \models \neg(\psi \sqsupset \chi)$. \square

This Theorem, together with our earlier remarks about the definability of $\psi \supset \chi$ by $\neg\psi \vee \chi$, entails that the selective implication is definable in the \mathcal{CO} language by $\psi^d \vee \chi$. In other words, we could have adopted as our basic language the smaller language \mathcal{C} . Given the significant role that the selective implication plays in our developments¹⁸ we prefer to stick to the \mathcal{CO} language.

We wish to point out two more properties of the dual negation.

Theorem 2.12 (Complementarity) *Let T be a causal team, $\varphi \in \mathcal{CO}$, and $S = \{s \in T^- \mid \{s\} \models \varphi\}$. Then $T^- \setminus S = \{s \in T^- \mid \{s\} \models \varphi^d\}$.*

Proof By induction on the syntactical complexity of $\varphi \in \mathcal{CO}$. We consider the least obvious cases.

Let $\varphi = \psi \supset \chi$. Then $\{s\} \models \varphi^d \iff \{s\} \models \psi \wedge \chi^d \iff$ (by inductive hypothesis) $\{s\} \models \psi$ and $\{s\} \not\models \chi \iff \{s\} \not\models \psi \supset \chi$.

Let $\varphi = \psi \sqsupset \chi$. Then $\{s\} \models \varphi^d \iff \{s\} \models \psi \sqsupset \chi^d \iff \{s\}_\psi \models \chi^d \iff$ (by inductive hypothesis) $\{s\}_\psi \not\models \chi \iff \{s\} \not\models \psi \sqsupset \chi$. \square

Theorem 2.13 (Non-contradiction) *Let T be a recursive causal team, $\varphi \in \mathcal{CO}$. Suppose $T \models \varphi$ and $T \models \varphi^d$. Then $T^- = \emptyset$.*

The proof is a simple induction on the syntax.

The dual negation \neg must be contrasted with the (*team-theoretical*) contradictory negation:

$$- \quad T \models \sim\varphi \iff T \not\models \varphi$$

Languages with contradictory negation are neither flat nor downward closed, and they also lack the empty team property. Contradictory negation usually increases considerably the expressive power of a language, but we shall not be concerned with it in the present paper.

3 Causal Teams as a Modeling Tool

In the present section we will show, by exhibiting two case studies, that causal team semantics allows for richer modelling possibilities than its special case, structural

¹⁸Especially for extensions to probabilistic reasoning, as sketched in [2].

equation modelling. In particular, we point out that some counterfactual scenarios considered in the literature on Lewis-Stalnaker counterfactuals essentially require causal teams (with more than one assignment) if we want to treat the conditionals as interventionist counterfactuals.

The first example is a simple coin-tossing scenario considered by Lewis in connection with a controversy about two Stalnakerian principles:

- The *principle of conditional excluded middle* (CEM)

$$\models (\varphi \Box \rightarrow \psi) \vee (\varphi \Box \rightarrow \neg\psi)$$

- The *principle of consequent distribution* (D)

$$\varphi \Box \rightarrow (\psi \vee \chi) \models (\varphi \Box \rightarrow \psi) \vee (\varphi \Box \rightarrow \chi).$$

In the second subsection we show how to model a counterexample to the Lewis-Stalnaker interpretation of counterfactuals that was proposed by McGee. It turns out that the counterfactuals involved in the example are correctly evaluated in a *non*-recursive causal team, whose structural equations plausibly do not have a causal interpretation. In this example, although the “actual world” seems to be captured by a deterministic SEM (i.e. by a causal team with exactly one assignment), the process of evaluation of counterfactuals inevitably *produces* causal teams with more than one assignment.

3.1 Lewis’s Coin Tossing

In his critique of the *conditional excluded middle* and *consequent distribution* principles, Lewis [29] considered a simple coin-tossing scenario. Let the Boolean variable X stand for the event “whether the coin is tossed” and Y stand for the outcome of the tossing ($h = \text{heads}$, $t = \text{tails}$). Lewis finds problematic the following instance of (D):

$$\begin{aligned} X = 1 \Box \rightarrow (Y = h \vee Y = t) \models \\ (X = 1 \Box \rightarrow Y = h) \vee (X = 1 \Box \rightarrow Y = t). \end{aligned}$$

Stalnaker defended the validity of (D) and ensured it by requiring the antisymmetry of the similarity relation: if there are two possible worlds v and w such that v is as close to r as w is, $v \preceq_r w$, and w is as close to r as v is, $w \preceq_r v$, then v is identical to w . Lewis, on the other hand, thinks it is true that if the coin had been flipped it would have come up either head or tails but he denies both disjuncts. Consequently he rejects the antisymmetry of the model and the principles (D) and (CEM) that come with it.

We want to argue that Lewis’s example is not a counterexample to (D) when interpreted on structural equations models, nor on our framework. In the former case, this happens simply because there seems to be no way to describe Lewis’ example by means of a deterministic structural equation model. Instead, Lewis’s example can be naturally modelled by the following team T :

X	Y
1	h
1	t

where the first assignment corresponds to the case in which the outcome of the flipping of the coin is head, and the second to the case in which it is tails. The team has no structural equations.

Here the fact that the team has more than one assignment makes a difference. Although on singleton causal teams T we have $T \models \varphi \vee \psi$ if and only if $T \models \varphi$ or $T \models \psi$, this does not hold for teams with at least two assignments. In this case we have both

$$T \models X = 1 \Box \rightarrow (Y = h \vee Y = t)$$

and

$$T \models (X = 1 \Box \rightarrow Y = h) \vee (X = 1 \Box \rightarrow Y = t)$$

as T can be partitioned into two singleton causal teams, each making true one of the disjuncts. On the other side, both disjuncts are false, i.e.

$$T \not\models (X = 1 \Box \rightarrow Y = h)$$

and

$$T \not\models (X = 1 \Box \rightarrow Y = t)$$

exactly as Lewis wants. We thus managed to preserve both the validity of (D) and block Lewis's counterexample thanks to the generalization of disjunction as a tensor interpreted on arbitrary causal teams.

Let us mention in passing that, as far as \mathcal{CO} formulas are concerned, also the related principle (CEM) and the (unconditional) law of the excluded middle (LEM) hold for our counterfactuals, for reasons that will be clarified in Section 4. We do not consider these principles in relation to formulas with dependence atoms or occurrences of \sqcup , since we have not defined a dual negation for such formulas.¹⁹ We might instead consider extending the languages with the contradictory negation

$$- \quad T \models \sim \varphi \iff T \not\models \varphi$$

and contemplate principles such as (D \sim), (CEM \sim) and (LEM \sim) where \sim takes the place of the dual negation \neg . However, (LEM \sim) (and consequently (D \sim) and (CEM \sim)) already fails for constancy atoms: $\models (X) \vee \sim \models (X)$ is falsified on any causal team T , since no causal subteam of T – not even the empty one – satisfies $\sim \models (X)$. Similarly, $X = 1 \vee \sim X = 1$ fails on a team where X takes the constant value 1: (LEM \sim) fails even when applied to atomic \mathcal{CO} formulas.

3.2 McGee's Counterfactuals

Our second example involves a system of equations whose graph is *cyclic*, i.e., a *nonrecursive* system. As hinted already in [2], there is a natural way of extending the notion of intervention to cover also the case of non-recursive causal teams. The idea is as follows. Suppose we want to compute the intervention $do(\mathbf{X} = \mathbf{x})$ on a causal

¹⁹In principle, we could discuss dual negations within the language \mathcal{CO}_{\sqcup} . Indeed, analysing the truth clause for \neg we see that the dual negation of a \mathcal{CO}_{\sqcup} formula can be defined by enriching the definition of dualization with the clause $(\psi \sqcup \chi)^d = \psi \wedge \chi$. Similarly, the dual negation of a dependence atom can be defined as \perp (any contradiction, i.e. any formula, such as $X = x \wedge X \neq x$, which is satisfied only by causal teams with empty team component).

team T with exogenous variables \mathbf{U} . Let $\mathbf{W} = \mathbf{U} \cup \mathbf{X}$ and $\mathbf{U}' = \mathbf{U} \setminus \mathbf{X}$ (so that \mathbf{W} is the *disjoint* union of \mathbf{U}' and \mathbf{X}). An intervention $do(\mathbf{X} = \mathbf{x})$ is supposed to produce a new causal team of exogenous variables \mathbf{W} . The most natural way to obtain it seems to be to include in the intervened team all the assignments s that are “solutions” to the associated system of equations and which respect the constraint that $s(\mathbf{U}'\mathbf{X}) = \mathbf{u}'\mathbf{x}$ for some tuple \mathbf{u}' of \mathbf{U}' -values occurring in T . This somewhat abstract definition is illustrated in the example below.

McGee’s paper [31] is mostly well-known for its counterexample to modus ponens for *indicative* conditionals. However, McGee also discusses the counterfactuals

- a) If Reagan hadn’t won the election, and a Republican had won, it would have been Anderson.
- b) If Reagan hadn’t won the election, then, if a Republican had won, it would have been Anderson.

Under the similarity model, (a) comes out true in the actual world, where Reagan was the winner, slightly ahead of the Democrat candidate, Carter, and Anderson was a distant third. McGee finds also (b) true (as predicted by the law of Exportation, which we will discuss in Section 4.3). However, the similarity-based model makes (b) false, as the possible world the most similar to the actual world in which Reagan did not win is a world in which Carter came first and Reagan slightly second, with Anderson again a distant third. Thus the Lewis-Stalnaker theory wrongly predicts that

- c) If Reagan hadn’t won the election, then, if a Republican had won, it would have been Reagan

is true. McGee’s conclusion is that similarity models are not adequate as a semantics for counterfactuals (p. 467).

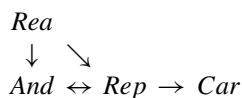
We model this example by a causal team T with a domain consisting of four Boolean variables (using self-explanatory abbreviations; *Rea* is the only exogenous variable)

<i>Rea</i>	<i>And</i>	<i>Car</i>	<i>Rep</i>
1	0	0	1

and three equations (which clearly express non-causal dependencies):

$$\begin{aligned} Rep &:= Rea \vee And \\ And &:= \neg Rea \wedge Rep \\ Car &:= \neg Rep. \end{aligned}$$

The equations induce the graph



We take note that the graph is cyclic, given the path

$$And \longrightarrow Rep \longrightarrow And$$

It can be checked that the only assignment of the team, call it s , is consistent with the equations.

The cyclicity of the graph introduces some complications in defining interventions on T which might not output a unique assignment for each given assignment in T . Indeed, this is the case in our example. An intervention $Rea = 0$ leads to a new causal team $T_{Rea=0}$ with a new system of equations eq' ($Rep = Rea \vee And$, $And = \neg Rea \wedge Rep$, $Car = \neg Rep$, $Rea = 0$) and a team component with two assignments which are all the solutions to eq'

Rea	And	Car	Rep
0	1	0	1
0	0	1	0

The table shows that the new system is compatible with both $Rep = 0$ and $Rep = 1$. Then

$$T_{Rea=0} \models Rep = 1 \Box \rightarrow And = 1$$

given that $(T_{Rea=0})_{Rep=1}$ is the team

Rea	And	Car	Rep
0	1	0	1

with the equations $Rep = 1$, $And = \neg Rea \wedge Rep$, $Car = \neg Rep$ and $Rea = 0$.

Thus (b) is true. Notice also that $T \models (Rea = 0 \wedge Rep = 1) \Box \rightarrow And = 1$, given that $T_{(Rea=0 \wedge Rep=1)}$ is the same causal team as $(T_{Rea=0})_{Rep=1}$.²⁰ This confirms the equivalence of (a) and (b), as desired. So, it looks as if McGee's example *can* be modelled by means of interventionist counterfactuals, *provided* we are ready to allow for the causal teams with multiple assignments that may arise after an intervention on a nonrecursive causal model.

It is worth pointing out that after the intervention $do(Rea = 0)$, it is still unsettled, as expected, whether Anderson won the election or not, as witnessed by

$$\begin{aligned} T &\not\models Rea = 0 \Box \rightarrow And = 1 \\ T &\not\models Rea = 0 \Box \rightarrow And = 0. \end{aligned}$$

What is settled, however, after that intervention, is the fact that “whether Anderson is the winner or not” is completely determined by “whether the winner is Republican or not”. This is, indeed, the case, as witnessed by the fact that $T_{Rea=0} \models (Rep; And)$.

4 Counterfactual Logic

In this section we ask two questions. First, what logical principles govern the behaviour of interventionist counterfactuals as interpreted on causal teams? Second, how do these principles differ from those that govern that behaviour according to the Lewis-Stalnaker interpretation?

²⁰We return to this point in Section 4.3.

We shall write $\psi \models \chi$ to mean that χ holds on any recursive causal team on which ψ holds. We write $\psi \equiv \chi$ to express that both $\psi \models \chi$ and $\chi \models \psi$ hold. By $\models \chi$ we mean that χ holds on all recursive causal teams. We will write $\models_\sigma, \equiv_\sigma$ for the analogous notions relativized to a fixed signature σ .

In the following, for better readability, we will often write $\neg\psi$ for the dualization ψ^d (defined in Section 2.6).

4.1 Axiomatization of \mathcal{CO}

We present a complete deductive system for $\mathcal{CO}(\sigma)$ languages (over recursive causal teams of signature σ) which is largely adapted from the system presented by Halpern [16], combined with some ideas from [3]. The axioms of Halpern's system which involve material implication could be rendered in our system using selective implication. We preferred to convert them into inference rules; in this way, the system we obtain is still sound for downward closed languages which extend \mathcal{CO} (such as \mathcal{CO}_\sqcup , \mathcal{COD} , \mathcal{COD}_\sqcup).²¹ Another basic difference is that Halpern's treatment of exogenous and endogenous variables differs from ours in at least two respects: it allows interventions on endogenous variables only; and it encodes the exogenous/endogenous distinction in the signature. We see no philosophical reasons nor technical advantages in keeping this distinction.

Writing down one cumbersome axiom ("Recursivity") will be easier by introducing an abbreviation (as in [16]) for " X affects Y ":²²

$$X \rightsquigarrow Y \text{ iff } \bigvee \{ \mathbf{Z} = \mathbf{z} \sqsupset [(X = x \sqsupset Y = y) \wedge (X = x' \sqsupset Y = y')] \}$$

where the (tensor) disjunction ranges over all sets of variables $\mathbf{Z} \subseteq \text{Dom} \setminus \{X, Y\}$, all tuples $\mathbf{z} \in \text{Ran}(\mathbf{Z})$, all $x \neq x' \in \text{Ran}(X)$, $y \neq y' \in \text{Ran}(Y)$.²³

Our inference system $AX_{\mathcal{CO}}(\sigma)$ is then described by the following list, which includes four axiom schemes and twelve inference rules. We write \vdash_σ for derivability in $AX_{\mathcal{CO}}(\sigma)$. When a rule is marked with an asterisk, we mean that also its inverse should be included in the system.

- All \mathcal{CO} instances of classical tautologies.
- *Modus ponens*: $\frac{\psi \quad \psi \supset \chi}{\chi}$.
- *Effectiveness*: $(\mathbf{X} = \mathbf{x} \wedge Y = y) \sqsupset Y = y$.
- *Definiteness*: $\bigvee_{y \in \text{Ran}(Y)} (\mathbf{X} = \mathbf{x} \sqsupset Y = y)$.
- *A-Triviality*: $(\psi \wedge (X = x \wedge X = x')) \sqsupset \chi$.
- *Composition 1*: $\frac{\mathbf{X} = \mathbf{x} \sqsupset W = w \quad \mathbf{X} = \mathbf{x} \sqsupset Y = y}{(\mathbf{X} = \mathbf{x} \wedge W = w) \sqsupset Y = y}$.

²¹Of course, in these larger languages some rules (*Modus ponens*, *Definition of \supset* and *Extraction of \neg* ; see the axiomatization below) will be applicable only when the involved formula ψ is in \mathcal{CO} - and thus it has a dual negation. For non-downward-closed languages, the *Definition of \supset* rule must be corrected as follows: $\frac{\psi \supset \chi}{\neg\psi \vee (\psi \wedge \chi)}$.*

²²This causal notion is also referred to, in the literature, as " X is causally relevant to Y "; see e.g. [22].

²³Here $\mathbf{z} \in \text{Ran}(\mathbf{Z})$ abbreviates $z_1 \in \text{Ran}(Z_1)$ and... and $z_n \in \text{Ran}(Z_n)$.

- *Uniqueness*: $\frac{\mathbf{X} = \mathbf{x} \Box \rightarrow Y = y}{\mathbf{X} = \mathbf{x} \Box \rightarrow Y \neq y'} \text{ (for } y \neq y') \text{.}$
- *Definition of \supset* : $\frac{\psi \supset \chi}{\neg \psi \vee \chi}^* \text{.}$
- *Extraction of \wedge* : $\frac{\mathbf{X} = \mathbf{x} \Box \rightarrow \psi \wedge \chi}{(\mathbf{X} = \mathbf{x} \Box \rightarrow \psi) \wedge (\mathbf{X} = \mathbf{x} \Box \rightarrow \chi)}^* \text{.}$
- *Extraction of \vee* : $\frac{\mathbf{X} = \mathbf{x} \Box \rightarrow \psi \vee \chi}{(\mathbf{X} = \mathbf{x} \Box \rightarrow \psi) \vee (\mathbf{X} = \mathbf{x} \Box \rightarrow \chi)}^* \text{.}$
- *Extraction of \neg* : $\frac{\mathbf{X} = \mathbf{x} \Box \rightarrow \neg \psi}{\neg(\mathbf{X} = \mathbf{x} \Box \rightarrow \psi)}^* \text{.}$
- *Extraction of $\Box \rightarrow$* : $\frac{\mathbf{X} = \mathbf{x} \Box \rightarrow (\mathbf{Y} = \mathbf{y} \Box \rightarrow \psi)}{(\mathbf{X}' = \mathbf{x}' \wedge \mathbf{Y} = \mathbf{y}) \Box \rightarrow \psi} \text{ (where } \mathbf{X} \text{ consistent, } \mathbf{X}' = \mathbf{X} \setminus \mathbf{Y} \text{ and } \mathbf{x}' = \mathbf{x} \setminus \mathbf{y}) \text{.}$
- *Exportation*: $\frac{(\mathbf{X} = \mathbf{x} \wedge \mathbf{Y} = \mathbf{y}) \Box \rightarrow \psi}{\mathbf{X} = \mathbf{x} \Box \rightarrow (\mathbf{Y} = \mathbf{y} \Box \rightarrow \psi)} \text{ (where } \mathbf{X} \cap \mathbf{Y} = \emptyset \text{).}$
- *Recursivity*^k: $\frac{X_1 \rightsquigarrow X_2 \quad \dots \quad X_{k-1} \rightsquigarrow X_k}{\neg(X_k \rightsquigarrow X_1)} \text{.}$
- *C-Substitution*: If $\psi \vdash_{\sigma} \chi$, then $\mathbf{X} = \mathbf{x} \Box \rightarrow \psi \vdash_{\sigma} \mathbf{X} = \mathbf{x} \Box \rightarrow \chi$.
- *A-Substitution*: If $\mathbf{X} = \mathbf{x} \dashv\vdash_{\sigma} \mathbf{Z} = \mathbf{z}$, then $\mathbf{X} = \mathbf{x} \Box \rightarrow \psi \vdash_{\sigma} \mathbf{Z} = \mathbf{z} \Box \rightarrow \psi$.

These are all schemes, parametrized over the variables and values allowed by the given signature σ , and over k ; ψ and χ range over \mathcal{CO} formulas. We include the limit cases in which some set of variables is empty (thus, if a counterfactual $\mathbf{X} = \mathbf{x} \Box \rightarrow Y = y$ occurs in an axiom or rule, if $\mathbf{X} = \emptyset$ this formula will be replaced by $Y = y$). We show the completeness of this system for \models_{σ} (over recursive causal teams) in Appendix C.

Most of the schemes correspond straightforwardly to those considered by Halpern in [16]. The *Extraction of \neg , \wedge , \vee* rules were also noticed by Halpern in [16], but were not part of his axiom systems; yet they are needed for our completeness proof, since our language is a bit more complex than Halpern's.²⁴ *Extraction of $\Box \rightarrow$* and *Exportation* are needed because, unlike Halpern, we allow for nested counterfactuals; in particular, *Extraction of $\Box \rightarrow$* is a more compact presentation of an axiom discovered by Briggs [3]. *A-triviality* expresses the vacuous truth of counterfactuals with inconsistent antecedents; we have not found earlier mention of this axiom in the literature. *C-Substitution* and *A-Substitution* have been used in some axiomatizations of Lewisian counterfactuals, and they seem to be used implicitly in the literature on interventionist counterfactuals while proving completeness results, but as far as we know, in the latter context they are explicitly isolated here for the first time. We point out that *A-Substitution* could be replaced by three more elementary invertible rules:

- *A-Commutativity*: $\frac{(\psi \wedge \chi) \Box \rightarrow \varphi}{(\chi \wedge \psi) \Box \rightarrow \varphi}^* \text{.}$

²⁴Halpern includes the *Extraction rules* for \neg , \wedge , \vee in later writings such as [15], where he works with more complex languages.

- *A-Associativity*: $\frac{((\psi \wedge \chi) \wedge \theta) \Box \rightarrow \varphi}{(\psi \wedge (\chi \wedge \theta)) \Box \rightarrow \varphi}^*$.
- *A-Removal*: $\frac{(\psi \wedge (X = x \wedge X = x)) \Box \rightarrow \varphi}{(\psi \wedge X = x) \Box \rightarrow \varphi}^*$.

The completeness proof for this modified system is a bit more difficult (in particular, for what regards the proof of lemma C.10 in the Appendix).

The case of languages $\mathcal{COD}(\sigma)$, $\mathcal{CO}_{\sqcup}(\sigma)$ and $\mathcal{COD}_{\sqcup}(\sigma)$ is much more delicate. One problem which arises in this context is the nonclassical behaviour of the tensor disjunction \vee with respect to non-flat disjuncts: for example, the law of idempotency, $\psi \vee \psi \equiv \psi$, may fail. It is to be hoped that complete natural deduction systems for these languages can be produced in the style of those obtained for propositional dependence logic [47], but this remains a task for future work.

4.2 Invalid Principles

We will give counterexamples to the validity of some well known logical principles (cp. [29], sec. 1.8):

- *strengthening the antecedent*

$$\varphi \Box \rightarrow \psi \models (\varphi \wedge \chi) \Box \rightarrow \psi$$

- *contraposition*

$$\varphi \Box \rightarrow \psi \equiv \neg \psi \Box \rightarrow \neg \varphi$$

- *transitivity*

$$\varphi \Box \rightarrow \psi, \psi \Box \rightarrow \chi \models \varphi \Box \rightarrow \chi$$

- *Modus Tollens*

$$\neg \psi, \varphi \Box \rightarrow \psi \models \neg \varphi$$

- *weak transitivity*:

$$\varphi \Box \rightarrow \psi, (\varphi \wedge \psi) \Box \rightarrow \chi \models \varphi \Box \rightarrow \chi$$

- *Modus Ponens*

$$\varphi, \varphi \Box \rightarrow \psi \models \psi$$

- *conjunction conditionalization*

$$\varphi \wedge \psi \models \varphi \Box \rightarrow \psi.$$

The non-validity of the first three principles is a common feature of both similarity based (Lewis-Stalnaker) and interventionist interpretations of counterfactuals. It is generally agreed that these principles, which are valid for strict implication, are not desirable for the counterfactual conditional. By contrast the last four principles hold for similarity-based counterfactuals but fail for interventionist counterfactuals (the counterexamples require nesting of counterfactuals, i.e. occurrences of $\Box \rightarrow$ in the consequent of a counterfactual). This is the case with our languages \mathcal{CO} and \mathcal{COD} (and their extensions) as well as with those considered in [3]. The four principles do hold for unnested interventionist counterfactuals, as long as only *recursive* causal teams are considered.²⁵

²⁵See [16] and [49] for unnested counterexamples in the nonrecursive case.

We now give counterexamples to the first three principles.

A counterexample to the validity of the *strengthening of the antecedent* is quite straightforward. Consider the causal team T

X	Y	Z
0	1	0

with one equation, $Z := X \wedge Y$. It may be checked that $T \models X = 1 \Box \rightarrow Z = 1$ but $T \not\models X = 1 \wedge Y = 0 \Box \rightarrow Z = 1$; in fact we have $T \models X = 1 \wedge Y = 0 \Box \rightarrow Z = 0$ illustrating how additional information can flip the truth-value of the consequent.

For *contraposition*, notice that our languages allow only quite limited instances, as antecedents can only take the form $\mathbf{X} = \mathbf{x}$. This formula is the dual negation of $\bigvee_{x \in \mathbf{X}} X \neq x$. Therefore, the most general instance of the law of *contraposition* in \mathcal{CO} is:

$$\mathbf{X} = \mathbf{x} \Box \rightarrow \bigvee_{y \in \mathbf{Y}} Y \neq y \models \mathbf{Y} = \mathbf{y} \Box \rightarrow \bigvee_{x \in \mathbf{X}} X \neq x.$$

Now consider the causal team

$X \leftarrow Y$
0

with structural equation $X := Y$. Clearly, it satisfies $X = 1 \Box \rightarrow Y \neq 1$ but not its contrapositive $Y = 1 \Box \rightarrow X \neq 1$.

A counterexample to the principle of *transitivity* is more involved. To that effect, we consider the team T

X_1	X_2	X_3	X_4
0	0	0	1

with 3 equations (all variables are Boolean):

$$\begin{aligned} X_2 &= X_1 \\ X_4 &= \neg X_1 \\ X_3 &= (X_4 \wedge X_2) \vee (X_1 \wedge \neg X_2). \end{aligned}$$

It is straightforward to check that $T \models X_2 = 1 \Box \rightarrow X_3 = 1$, $T \models X_1 = 1 \Box \rightarrow X_2 = 1$ but $T \not\models X_1 = 1 \Box \rightarrow X_3 = 1$.

We now give counterexamples to the last four principles in our list. As pointed out, they are valid on the similarity semantics, valid (in the recursive case) also for unnested interventionist counterfactuals, but fail in general. The fact that these principles hold for unnested counterfactuals in the recursive case derives from an analysis of their axiomatizations on causal models (see [11] and [16]). The difference between the nested and the unnested case was first called into attention in [3] with a counterexample to the modus ponens principle for interventionist counterfactuals. Let us give a general argument which identifies the source of the difference between the behaviour of nested counterfactuals on similarity semantics and on structural equation based models. Consider a possible world model whose “worlds” are causal teams of a fixed signature; we also require that, if T is in the model, also $T_{\mathbf{X}=\mathbf{x}}$ is, for each consistent $\mathbf{X} = \mathbf{x}$. Then we may extract from our semantic clause for $\Box \rightarrow$ a world-selection function: $f(\mathbf{X} = \mathbf{x}, T) = T_{\mathbf{X}=\mathbf{x}}$ if $\mathbf{X} = \mathbf{x}$ is a consistent antecedent, and some fixed causal team with empty team component, otherwise; the truth of a counterfactual $\mathbf{X} = \mathbf{x} \Box \rightarrow \psi$ can then be written as a Stalnakerian condition: $T \models \mathbf{X} = \mathbf{x} \Box \rightarrow \psi$

if and only if $f(\mathbf{X} = \mathbf{x}, T) \models \psi$. Interventionist counterfactuals should then satisfy the Stalnakerian axiom system **VCS**, plus perhaps some additional axioms. How can we then explain that some Stalnakerian logical principles are falsified by (nested) interventionist counterfactuals? The answer is that a world-selection function (call it g) is required by Stalnaker to satisfy two constraints

- (Strong centering): if $T \models \mathbf{X} = \mathbf{x}$, then $g(\mathbf{X} = \mathbf{x}, T) = T$
- (Antisymmetry): if $g(\mathbf{X} = \mathbf{x}, T) \models \mathbf{Y} = \mathbf{y}$ and $g(\mathbf{Y} = \mathbf{y}, T) \models \mathbf{X} = \mathbf{x}$, then $g(\mathbf{X} = \mathbf{x}, T) = g(\mathbf{Y} = \mathbf{y}, T)$

both of which are easily falsified for our function f .²⁶

As expected in light of our explanations, also our counterexample to *Modus Tollens* involves nested counterfactuals. Consider the causal teams

$$T : \begin{array}{|c|c|} \hline X \rightarrow Y \\ \hline 1 & 1 \\ \hline \end{array} \quad T_{Y=1} : \begin{array}{|c|c|} \hline X & Y \\ \hline 1 & 1 \\ \hline \end{array} \quad (T_{Y=1})_{X=0} : \begin{array}{|c|c|} \hline X & Y \\ \hline 0 & 1 \\ \hline \end{array} \quad T_{X=0} : \begin{array}{|c|c|} \hline X \rightarrow Y \\ \hline 0 & 0 \\ \hline \end{array}$$

with equation $Y := X$. Let ψ be $Y = 1$ and χ be $X = 0 \Box \rightarrow Y = 1$. Then $\neg\psi$ is $Y \neq 1$ and $\neg\chi$ is $X = 0 \Box \rightarrow Y \neq 1$. Now $T \models \neg\chi$, and $T \models \psi \Box \rightarrow \chi = Y = 1 \Box \rightarrow (X = 0 \Box \rightarrow Y = 1)$, but $T \not\models \neg\psi$. This example also illustrates our earlier remark that the candidate “selection function” f does not satisfy Stalnaker’s requirements. Indeed, even though $Y = 1$ holds in T , T and $f(Y = 1, T) = T_{Y=1}$ are distinct causal teams, as they differ in the function and graph components: strong centering fails. Notice then that $f(Y = 1, T) = T_{Y=1} \models X = 1$ and $f(X = 1, T) = T \models Y = 1$, but $f(Y = 1, T) = T_{Y=1} \neq T_{X=1} = T = f(X = 1, T)$, which falsifies antisymmetry.

In [29], Lewis considered some variants of transitivity that *do* hold for his conditionals, such as the *weak transitivity* principle stated above. In the context of causal models or causal teams, weak transitivity has a special instance of particular interest:

$$\mathbf{X} = \mathbf{x} \Box \rightarrow W = w, (\mathbf{X} = \mathbf{x} \wedge W = w) \Box \rightarrow Y = y \models \mathbf{X} = \mathbf{x} \Box \rightarrow Y = y.$$

This inference expresses one half of the axiom of *Composition* that was introduced by Galles&Pearl ([11]); more on this in Appendix C. Its “converse” is also valid

²⁶We will illustrate this point in our counterexample to *Modus Tollens*. These observations are inspired by [49], where J. Zhang uses set-selection functions $g(\mathbf{X} = \mathbf{x}, s) := s_{\mathbf{X}=\mathbf{x}}$ to explain the failure of Lewisian laws for unnested counterfactuals over *nonrecursive* causal models. The functions suggested by him, indeed, fail to satisfy one of the Lewisian constraints for set-selection functions (S4; see [49] for the list). However, we cannot fully agree with Zhang’s analysis: as he himself observes, his set-selection functions do satisfy all of Lewis’s constraints over *recursive* causal models; it should then be impossible to produce counterexamples to Lewisian laws on recursive models, as done in [3] and in the rest of this section. The problem with Zhang’s approach seems to be that it takes worlds to be assignments (without any functional or equational component), while the unit of semantic evaluation is the pair (s, \mathcal{F}) (s assignment, \mathcal{F} functional/equational component). Therefore, the nonvacuous truth of a conditional is expressed by the clause $(s, \mathcal{F}) \models \mathbf{X} = \mathbf{x} \Box \rightarrow \psi$ iff “ $(t, \mathcal{F}) \models \psi$ for all $t \in g(\mathbf{X} = \mathbf{x}, s)$ ”, which is *not* Lewis’s truth condition “ $t \models \psi$ for all $t \in g(\mathbf{X} = \mathbf{x}, s)$ ”. It seems thus to be impossible to draw any conclusion using a function g which returns sets of assignments; one might try to use, instead, a function, similar to our f , which returns set of pairs (s, \mathcal{F}) . Such a function – as our f – can easily be seen to falsify Lewis’s requirements S1 and S4 on recursive causal models, and also S3 on nonrecursive ones.

over recursive causal models and teams. It is none other than the *Composition 1* rule already considered in the axiom system of Section 4.1:

$$\mathbf{X} = \mathbf{x} \Box \rightarrow W = w, \mathbf{X} = \mathbf{x} \Box \rightarrow Y = y \models (\mathbf{X} = \mathbf{x} \wedge W = w) \Box \rightarrow Y = y.$$

The analogy between *weak transitivity* and *composition* suggests that we consider an inverse to the former:

$$- \text{Inverse weak transitivity: } \varphi \Box \rightarrow \psi, \varphi \Box \rightarrow \chi \models (\varphi \wedge \psi) \Box \rightarrow \chi.$$

It turns out that this principle is valid according to the Lewis-Stalnaker semantics; it can also be seen as a way of repairing the “strengthening the antecedent” principle. To see the validity of this rule, observe that the premises state that the closest φ -worlds satisfy ψ and χ . The former assertion (together with Lewis’ *nesting* principle) tells us that the set of the closest φ -worlds and the set of the closest $\varphi \wedge \psi$ -worlds coincide. Therefore, the closest $\varphi \wedge \psi$ -worlds satisfy χ .²⁷ We may now wonder if *weak transitivity* and its converse hold in general for languages evaluated on causal teams. The most general form that *weak transitivity* can take for the kind of languages we have considered in this paper is:

$$\mathbf{X} = \mathbf{x} \Box \rightarrow \mathbf{Y} = \mathbf{y}, (\mathbf{X} = \mathbf{x} \wedge \mathbf{Y} = \mathbf{y}) \Box \rightarrow \psi \models \mathbf{X} = \mathbf{x} \Box \rightarrow \psi.$$

(where for simplicity we may assume that the antecedents of counterfactuals are consistent). If ψ contains no counterfactuals, this principle is valid over causal teams T , for the general reasons mentioned above. We can also give a more informative direct argument: first, from the second premise we get $T_{\mathbf{X}=\mathbf{x}\wedge\mathbf{Y}=\mathbf{y}} \models \psi$. The consistency of $\mathbf{X} = \mathbf{x} \wedge \mathbf{Y} = \mathbf{y}$ allows us to rewrite the latter condition as $(T_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}} \models \psi$.²⁸ The first premise tells us that $T_{\mathbf{X}=\mathbf{x}} \models \mathbf{Y} = \mathbf{y}$; therefore, the intervention $do(\mathbf{Y} = \mathbf{y})$ will not modify the team component of $T_{\mathbf{X}=\mathbf{x}}$; thus $T_{\mathbf{X}=\mathbf{x}}$ and $(T_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}}$ satisfy the same formulas without occurrences of $\Box \rightarrow$, in particular they both satisfy ψ . This also shows the validity of the *Inverse* for ψ without occurrences of $\Box \rightarrow$. We would like to draw attention to the fact that this argument is purely model-theoretical, and so it proves the validity of the rule for unnested counterfactuals of any language that features our version of the interventionist counterfactuals (in particular, for \mathcal{CO} , \mathcal{COD} , \mathcal{CO}_{\sqcup} , \mathcal{COD}_{\sqcup}). Instead, we can easily build counterexamples to *weak transitivity* involving formulas ψ which allow occurrences of $\Box \rightarrow$. Consider the following team T

X	Y	Z
0	1	1

with equation $Y := X \vee Z$. Now $T \models X = 0 \Box \rightarrow Y = 1$ and $T \models (X = 0 \wedge Y = 1) \Box \rightarrow (Z = 0 \Box \rightarrow Y = 1)$, but $T \not\models X = 0 \Box \rightarrow (Z = 0 \Box \rightarrow Y = 1)$. A counterexample to the *Inverse* is obtained by observing that $T \models X = 0 \Box \rightarrow Y = 1$, $T \models X = 0 \Box \rightarrow (Z = 0 \Box \rightarrow Y = 0)$ but $T \not\models (X = 0 \wedge Y = 1) \Box \rightarrow (Z = 0 \Box \rightarrow Y = 0)$.

²⁷Here, for simplicity of exposition, we are implicitly using the *Limit assumption*, i.e. the requirement that for each possible antecedent φ , there exists a set of closest φ -worlds (and not a descending chain of closer and closer φ -worlds). The rule is also valid when the *Limit assumption* is violated.

²⁸This point is detailed in our comments following Theorem 4.1, in Section 4.3.

Our counterexamples to *Modus Ponens* and *conjunction conditionalization* (simpler than Briggs's "execution line" example, [3]) use the following causal teams, with Boolean ranges and functions $f_Y(X) := X$ and $f_Z(X, Y) := X \wedge Y$:

$$\begin{array}{cc}
 T = T^{Y=0}: & \begin{array}{|c|c|c|} \hline \text{X} \rightarrow \text{Y} \rightarrow \text{Z} \\ \hline 0 & 0 & 0 \\ \hline \end{array} & (T^{Y=0})_{X=1}: & \begin{array}{|c|c|c|} \hline \text{X} \rightarrow \text{Y} \rightarrow \text{Z} \\ \hline 1 & 1 & 1 \\ \hline \end{array} \\
 \\
 T_{Y=0}: & \begin{array}{|c|c|c|} \hline \text{X} \quad \text{Y} \rightarrow \text{Z} \\ \hline 0 & 0 & 0 \\ \hline \end{array} & (T_{Y=0})_{X=1}: & \begin{array}{|c|c|c|} \hline \text{X} \quad \text{Y} \rightarrow \text{Z} \\ \hline 1 & 0 & 0 \\ \hline \end{array}
 \end{array}$$

It is then immediate to see that:

1. $T \models Y = 0$
2. $T \models Y = 0 \Box \rightarrow (X = 1 \Box \rightarrow Z \neq 1)$
3. $T \not\models Y = 0 \supset (X = 1 \Box \rightarrow Z \neq 1)$
4. $T \not\models (X = 1 \Box \rightarrow Z \neq 1)$
5. $T \models X = 1 \Box \rightarrow Z = 1$
6. $T \not\models Y = 0 \Box \rightarrow (X = 1 \Box \rightarrow Z = 1)$

Now 1., 2., and 4. provide a counterexample to *Modus Ponens*. Like the other two earlier counterexamples, this one involves nested counterfactuals. We notice in passing that the pair of formulas 2., 3. confirms the unsurprising fact that $\Box \rightarrow$ and \supset have different truth conditions. For a counterexample to conjunction conditionalization, observe that from 1. and 5. we get $T \models Y = 0 \wedge (X = 1 \Box \rightarrow Z = 1)$. On the other hand, the corresponding counterfactual does not hold, as shown by 6.

As in the case of (*inverse*) *weak transitivity* and *Modus Tollens*, it can be shown by direct model-theoretical arguments that *Modus Ponens* and *conjunction conditionalization* hold for non-nested counterfactuals in any of the languages considered in the paper.

4.3 Permutation and Exportation/Importation of Antecedents

The following laws of *permutation*, *exportation* and *importation* of antecedents

$$\begin{aligned}
 (P) : \psi \Box \rightarrow (\psi' \Box \rightarrow \chi) &\equiv \psi' \Box \rightarrow (\psi \Box \rightarrow \chi) \\
 (E) : (\psi \wedge \psi') \Box \rightarrow \chi &\models \psi \Box \rightarrow (\psi' \Box \rightarrow \chi) \\
 (I) : \psi \Box \rightarrow (\psi' \Box \rightarrow \chi) &\models (\psi \wedge \psi') \Box \rightarrow \chi
 \end{aligned}$$

are known to fail quite generally for Stalnaker's [39] and Lewis's [29] counterfactuals (see [37], chap. 8); their validity for natural language counterfactuals is controversial. These principles can be shown to hold for our counterfactuals (under some restrictions). There is a direct link between these three rules and the different ways of producing a complex intervention. Under appropriate assumptions, an intervention can be decomposed into a sequence of two interventions over smaller sets of variables; this justifies exportation and importation. The possibility of swapping the order of the two smaller interventions justifies permutation. The following theorem

specifies the conditions under which these transformations do not modify the effect of an intervention.

Theorem 4.1 *Let T be a causal team with G_T finite acyclic, \mathbf{X}, \mathbf{Y} sets of variables, $\mathbf{x} \in \text{Ran}_T(\mathbf{X})$, and $\mathbf{y} \in \text{Ran}_T(\mathbf{Y})$. Suppose $\mathbf{X} = \mathbf{x} \wedge \mathbf{Y} = \mathbf{y}$ is consistent. Suppose also that, if $X_i = Y_j \in \mathbf{X} \cap \mathbf{Y}$, then $x_i = y_j$. Then $T_{\mathbf{X}=\mathbf{x} \wedge \mathbf{Y}=\mathbf{y}} = (T_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}}$.*

In words, the decomposition of an intervention works correctly for disjoint sets \mathbf{X}, \mathbf{Y} of variables, or more generally when $do(\mathbf{X} = \mathbf{x})$ and $do(\mathbf{Y} = \mathbf{y})$ act in the same way on the common variables $\mathbf{X} \cap \mathbf{Y}$.

The proof is straightforward but tedious and will not be given here (it can be found in [1]). From it we obtain as a corollary the validity of the interventionist versions of importation/exportation of antecedents

$$(IMP) : \frac{\mathbf{X} = \mathbf{x} \Box \rightarrow (\mathbf{Y} = \mathbf{y} \Box \rightarrow \chi)}{(\mathbf{X} = \mathbf{x} \wedge \mathbf{Y} = \mathbf{y}) \Box \rightarrow \chi} \quad (EXP) : \frac{(\mathbf{X} = \mathbf{x} \wedge \mathbf{Y} = \mathbf{y}) \Box \rightarrow \chi}{\mathbf{X} = \mathbf{x} \Box \rightarrow (\mathbf{Y} = \mathbf{y} \Box \rightarrow \chi)}$$

under the above restrictions, over recursive teams. We notice that when the conditions in the statement of the theorem are fulfilled, we also get $(T_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}} = T_{\mathbf{X}=\mathbf{x} \wedge \mathbf{Y}=\mathbf{y}}$ and $(T_{\mathbf{Y}=\mathbf{y}})_{\mathbf{X}=\mathbf{x}} = T_{\mathbf{Y}=\mathbf{y} \wedge \mathbf{X}=\mathbf{x}}$. Since the order of variables is irrelevant in the definition of the *do* algorithm, it follows that $T_{\mathbf{X}=\mathbf{x} \wedge \mathbf{Y}=\mathbf{y}} = T_{\mathbf{Y}=\mathbf{y} \wedge \mathbf{X}=\mathbf{x}}$. Thus $(T_{\mathbf{X}=\mathbf{x}})_{\mathbf{Y}=\mathbf{y}} = (T_{\mathbf{Y}=\mathbf{y}})_{\mathbf{X}=\mathbf{x}}$, which shows that the rule

$$(PERM) : \frac{\mathbf{X} = \mathbf{x} \Box \rightarrow (\mathbf{Y} = \mathbf{y} \Box \rightarrow \chi)}{\mathbf{Y} = \mathbf{y} \Box \rightarrow (\mathbf{X} = \mathbf{x} \Box \rightarrow \chi)}$$

is valid (under the above restrictions) over recursive causal teams.

In order to have a full characterization of the range of validity of the three rules, we still have to see what happens if some of the involved antecedents are inconsistent. When either $\mathbf{X} = \mathbf{x}$ or $\mathbf{Y} = \mathbf{y}$ is inconsistent, (IMP) and (EXP) hold for trivial reasons. But when $\mathbf{X} = \mathbf{x}$ and $\mathbf{Y} = \mathbf{y}$ are both consistent, while $\mathbf{X} = \mathbf{x} \wedge \mathbf{Y} = \mathbf{y}$ is not, then it is easy to build counterexamples to the two principles. In such cases, applying in sequence the interventions $do(\mathbf{X} = \mathbf{x})$ and $do(\mathbf{Y} = \mathbf{y})$ will produce the intervention $do(\mathbf{X}' = \mathbf{x}' \wedge \mathbf{Y} = \mathbf{y})$, where $\mathbf{X}' = \mathbf{X} \setminus \mathbf{Y}$ and $\mathbf{x}' = \mathbf{x} \setminus \mathbf{y}$; i.e., the second intervention overrides the first on the set of variables in common. This fact justifies the more general rule *Extraction of $\Box \rightarrow$* that was included in the axioms for \mathcal{CO} (Section 4.1).

5 Conditionals and Dependencies

In this section we consider the interactions of counterfactuals with other special operators, such as the selective implication, the Boolean disjunction and the dependence atoms. We also investigate what kinds of causal dependencies can be expressed in our languages.

5.1 Selective vs Counterfactual Implication

In this section we focus on the logical properties of selective implication and its interaction with counterfactual implication. Given that selective implication is a generalization of material implication to team semantics, it is natural to ask to what extent properties of the latter generalise to the former. We consider several cases, which show how the generalization is conditioned on some of the properties of causal teams that we investigated earlier. The following inference rules

$$\frac{\theta \supset \chi}{(\theta \wedge \psi) \supset \chi} \quad \frac{\theta \supset \chi}{\theta \supset (\chi \vee \psi)} \quad \frac{\theta \supset (\chi \supset \psi)}{(\theta \wedge \chi) \supset \psi}$$

are sound when the antecedents of the selective implication are \mathcal{CO} formulas; the leftmost rule also requires downward closure of the consequent χ , and the rule in the middle requires the empty team property. The rightmost rule is invertible.

We show how to prove the soundness of the leftmost rule.²⁹ The team $T^{\theta \wedge \psi}$ is a causal subteam of T^θ . By the assumption of the rule, $T^\theta \models \chi$; hence by downward closure, $T^{\theta \wedge \psi} \models \chi$. Therefore $T \models (\theta \wedge \psi) \supset \chi$.

Selective implication also obeys a semantic version of the Deduction Theorem, relative to \mathcal{CO} antecedents.³⁰ To see this, assume that any team that satisfies a \mathcal{CO} formula θ also satisfies χ , and consider, for an arbitrary causal team T , the subteam T^θ . By flatness, $T^\theta \models \theta$, and by our assumption, $T^\theta \models \chi$. Hence $T \models \theta \supset \chi$. Thus all the valid inference rules that are stated in the paper can be internalized in the object language as selective implications, *provided the antecedents are \mathcal{CO} formulas*.

It is standard, after Lewis's seminal book [29], to inquire about the relationships between the counterfactual $\psi \Box \rightarrow \chi$ and the material implication $\psi \rightarrow \chi$. It is known that in the Stalnaker-Lewis approach, [29, 39] the former implies the latter; the converse is also known to hold, provided the antecedent ψ is true. The proof of these facts requires strong centering.

It is natural to ask whether similar principles hold for our interventionist counterfactual and selective implication. Given our analysis of these notions on causal teams, the first question may be rephrased as whether an intervention may be replaced with an observation. Within this perspective, it is also natural to consider other similar questions, e.g., whether the two conditionals respect the same inference rules, whether the order of consecutive applications of the two operations may be reversed (commutativity), etc.

The answer to these questions turns out to be negative, as expected. This has to do, essentially, with the discrepancy between evidential reasoning (observations), encoded by T^ψ , and counterfactual reasoning (interventions), encoded by T_ψ . The former produces a causal subteam of T which, by definition, uses the same structural equations and graph components as T . By contrast, the latter may alter the

²⁹A slightly more complex argument shows that this rule holds, more generally, for the maximal implication of [27], again under the assumption of downward closure of the consequent – yet without restrictions on the antecedent.

³⁰This argument might be extended to the maximal implication of [27], provided the antecedent is a formula closed under union of causal subteams.

equations component and the graph component of T , which may further affect the truth-conditions of some counterfactual statements even when ψ is true in T .

The fact that the uniform replacement of an intervention by an observation may affect the truth-value of the relevant formula was shown while discussing our counterexample to modus ponens in Section 4.2. With the next example, we show that also the logical status of a formula may be affected by such a replacement.³¹ More exactly, there are non-valid formulas which become valid (on teams with appropriate variable domain) once every occurrence of $\Box \rightarrow$ is replaced by \supset . Consider the formula $Y = 1 \supset (X = 1 \Box \rightarrow Y = 1)$. It is false in every causal team whose structural equation component is $Y := X + 1$, such as the causal team T shown in the picture:

$$T = T^{Y=1}: \begin{array}{|c|c|} \hline X \rightarrow Y \\ \hline 0 & 1 \\ \hline \end{array} \quad (T^{Y=1})_{X=1}: \begin{array}{|c|c|} \hline X \rightarrow Y \\ \hline 1 & 2 \\ \hline \end{array}$$

On the other hand, it follows from the semantic clause for selective implication that the formula $Y = 1 \supset (X = 1 \supset Y = 1)$ is valid. For a different kind of counterexample, consider the formula $X = 1 \Box \rightarrow (X = 2 \Box \rightarrow X = 1)$: it is contradictory, i.e. false in every causal team (of nonempty team component) with the appropriate domain of variables (notice that the second intervention rewrites the value assigned to X by the first intervention; given that assignments are functions, the claim follows right away). But $X = 1 \supset (X = 2 \supset X = 1)$ is trivially true in any causal team T , given that $((T^{X=1})^{X=2})^-$ is empty.

Finally, we show the non-commutativity of interventions with observations, that is, we show that the equivalence between $\theta \Box \rightarrow (\chi \supset \psi)$ and $\chi \supset (\theta \Box \rightarrow \psi)$ fails on causal teams in both directions. All the teams in our counterexamples have one structural equation, $Y := X + Z$; it is not really important to specify the ranges of variables. The teams

$$\begin{array}{l} T: \begin{array}{|c|c|c|} \hline X & Z & Y \\ \hline 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 5 \\ \hline \end{array} \quad T_{X=1}: \begin{array}{|c|c|c|} \hline X & Z & Y \\ \hline 1 & 1 & 2 \\ 1 & 2 & 3 \\ \mathbf{1} & 3 & \mathbf{4} \\ \hline \end{array} \quad (T_{X=1})^{X=1}: \begin{array}{|c|c|c|} \hline X & Z & Y \\ \hline 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & 3 & 4 \\ \hline \end{array} \\ \\ T^{X=1}: \begin{array}{|c|c|c|} \hline X & Z & Y \\ \hline 1 & 1 & 2 \\ 1 & 2 & 3 \\ \hline \end{array} \quad (T^{X=1})_{X=1}: \begin{array}{|c|c|c|} \hline X & Z & Y \\ \hline 1 & 1 & 2 \\ 1 & 2 & 3 \\ \hline \end{array} \end{array}$$

show that $T \models X = 1 \supset (X = 1 \Box \rightarrow (Y = 2 \vee Y = 3))$, but $T \not\models X = 1 \Box \rightarrow (X = 1 \supset (Y = 2 \vee Y = 3))$. For the opposite direction consider the teams (with the same graph and equation as before):

$$\begin{array}{l} S: \begin{array}{|c|c|c|} \hline X & Z & Y \\ \hline 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 1 & 3 \\ \hline \end{array} \quad S_{Z=1}: \begin{array}{|c|c|c|} \hline X & Z & Y \\ \hline 1 & 1 & 2 \\ 2 & 1 & 3 \\ \hline \end{array} \quad (S_{Z=1})^{Y=3}: \begin{array}{|c|c|c|} \hline X & Z & Y \\ \hline 2 & 1 & 3 \\ \hline \end{array} \end{array}$$

³¹We thank Tapani Hyttinen for raising this issue.

$S^{Y=3}:$

X	Z	Y
1	2	3
2	1	3

$(S^{Y=3})_{Z=1}:$

X	Z	Y
1	1	2
2	1	3

We have $S \models Z = 1 \Box \rightarrow (Y = 3 \supset Y = 3)$ but $S \not\models Y = 3 \supset (Z = 1 \Box \rightarrow Y = 3)$.

We remark that the noncommutativity of $\Box \rightarrow$ with \supset is a phenomenon that lies somehow hidden in the literature on causation, and is dealt with by means of a multiplicity of *ad hoc* notations (“do expressions”, “counterfactual expressions”, etc.) which only cover some specific logico-probabilistic forms.³²

5.2 Dependence Atoms

As we already pointed out, the inferences rules described in Section 4 extend to languages that admit dependence atoms (and/or Boolean disjunctions). In this subsection, we briefly describe some rules that are specific to the extended languages \mathcal{COD} , \mathcal{CO}_{\sqcup} , \mathcal{COD}_{\sqcup} , and that illustrate some of the interactions between dependence atoms (functional dependencies), on one side, and selective implications, counterfactuals and causal dependencies, on the other. These issues will be clearer in the light of the following observations (which derive from [6]): 1) a constancy atom $\models(Y)$ is equivalent (relative to a fixed finite signature $\sigma = (Dom, Ran)$) to the Boolean disjunction $\bigsqcup_{y \in Ran(Y)} Y = y$; more generally, for every causal team T of signature σ , $T \models \models(\mathbf{X}; Y)$ can be shown to be equivalent to the metalanguage assertion that, for all causal subteams S of T ,

$$S \models \bigsqcup_{\mathbf{x} \in Ran(\mathbf{X})} \mathbf{X} = \mathbf{x} \text{ entails } S \models \bigsqcup_{y \in Ran(Y)} Y = y$$

i.e., if a subteam decides the value of \mathbf{X} , then it also decides the value of Y . In the languages considered in this paper, though, we do not have the kind of conditionals that would allow us to internalize the above logical implication.³³ However, a moment of thought shows that we still can define $\models(\mathbf{X}; Y)$ in $\mathcal{CO}_{\sqcup}(\sigma)$ or $\mathcal{COD}_{\sqcup}(\sigma)$ as

$$\bigwedge_{\mathbf{x} \in Ran(\mathbf{X})} \bigsqcup_{y \in Ran(Y)} (\mathbf{X} = \mathbf{x} \supset Y = y).$$

This tells us two things: first, that the expressive power of \mathcal{CO}_{\sqcup} is at least as great as that of \mathcal{COD} ; and secondly, that the following is a sound invertible rule in $\mathcal{COD}_{\sqcup}(\sigma)$:

$$\frac{\bigwedge_{\mathbf{x} \in Ran(\mathbf{X})} \bigsqcup_{y \in Ran(Y)} (\mathbf{X} = \mathbf{x} \supset Y = y)}{\models(\mathbf{X}; Y)} *$$

(as before, the asterisk marks the invertibility of the rule). The rule tells us that if for every tuple of values for \mathbf{X} there is a corresponding value that Y takes (or the tuple \mathbf{X} does not occur in the team), then we can infer that Y is functionally dependent on \mathbf{X} . Its inverse says that, if we assume that Y is functionally dependent on the set \mathbf{X} , then any evidence that \mathbf{X} takes certain specific values implies that Y takes a

³²This point will be substantiated in forthcoming work on causal team semantics for probabilistic languages.

³³The *inquisitive implication* of [7] and the *maximal implication* of [27] allow for this possibility.

constant value. We conjecture that this pair of rules summarizes all the interactions between dependence atoms and \supset in $\mathcal{COD}_{\sqcup}(\sigma)$. The rule (and its inverse) can also be converted into a sound, invertible rule for $\mathcal{COD}(\sigma)$:

$$\frac{\bigwedge_{\mathbf{x} \in \text{Ran}(\mathbf{X})} (\mathbf{X} = \mathbf{x} \supset \Rightarrow(Y))}{\Rightarrow(\mathbf{X}; Y)} *.$$

Notice that the inverse of this rule, unlike the rule for $\mathcal{COD}_{\sqcup}(\sigma)$, does not completely remove dependence atoms from the consequent of \supset ; and in general, we see no reasonable rule for removing constancy atoms from the consequents of \supset without making use of \sqcup . This might suggest that the expressive resources of \mathcal{COD} are weaker than those of \mathcal{CO}_{\sqcup} ; however, the literature on propositional dependence logics warns us against quick conclusions.³⁴

An analogous (but not invertible!) rule connects counterfactual and functional dependence:

$$\frac{\bigwedge_{\mathbf{x} \in \text{Ran}(\mathbf{X})} \bigsqcup_{y \in \text{Ran}(Y)} (\mathbf{X} = \mathbf{x} \Box \rightarrow Y = y)}{\Rightarrow(\mathbf{X}; Y)}$$

or, in the style of $\mathcal{COD}(\sigma)$:

$$\frac{\bigwedge_{\mathbf{x} \in \text{Ran}(\mathbf{X})} (\mathbf{X} = \mathbf{x} \Box \rightarrow \Rightarrow(Y))}{\Rightarrow(\mathbf{X}; Y)}.$$

Instead, the “inverse” inference from $\Rightarrow(\mathbf{X}; Y)$ to $\mathbf{X} = \mathbf{x} \Box \rightarrow \Rightarrow(Y)$ is unsound. The statement $\Rightarrow(\mathbf{X}; Y)$ tells us that

for any value x of X there is associated exactly one value y for Y ;
but we cannot infer that

for every value x there is a value y such that if X had taken the value x then Y would have been y .

This inference pattern can, expectedly, be falsified in many ways. We have already seen one such example in Section 2.4 (ex. 2.7); in that case, a functional dependency $\Rightarrow(Y; X)$ could be broken by an intervention – thereby falsifying the statement $Y = 0 \Box \rightarrow \Rightarrow(X)$ – because it was a contingent dependency, going in the opposite direction as the dependence induced by the structural equations. The following example shows a different possibility. Consider the causal team

T :

$X \rightarrow Y \leftarrow Z$
0 1 1
1 1 0

with associated structural equation $Y := X \vee Z$. For trivial reasons, $T \models \Rightarrow(X; Y)$. However, the intervention $do(X = 0)$ produces the causal team $T_{X=0}$

³⁴ [47] shows that extending a basic propositional language with either \sqcup or with dependence atoms leads to equiexpressive languages (in the sense that the two languages can characterize the same classes of teams). Nonetheless, the Boolean disjunction is not definable in the language with dependence atoms [46].

$$T_{X=0}:$$

$X \rightarrow Y \leftarrow Z$		
0	1	1
0	0	0

and $T_{X=0} \models (X; Y)$. In this example, an intervention on X can make it so that Y is not anymore functionally determined by X , even though here X is an argument of the function which produces Y – a *parent* of Y . Notice that the assignment had the effect of shrinking the set of values available for X in the team, while enlarging the set of values available for Y , and ultimately breaking the relation of functional dependence.³⁵ We also remark that causal teams with at least two assignments are needed to produce these kinds of counterexamples (as long as we only consider *recursive* causal teams).

Although we cannot infer $\mathbf{X} = \mathbf{x} \sqcap \rightarrow (Y)$ from $\models (\mathbf{X}; Y)$, the following much weaker rule is sound:

$$\frac{\models (\mathbf{X}; Y) \quad \mathbf{X} = \mathbf{x}}{\mathbf{X} = \mathbf{x} \sqcap \rightarrow \models (Y)}.$$

5.3 Notions of Cause

In this subsection we show that many of the notions of cause that are considered in the philosophical literature that is inspired by structural equation modeling translate smoothly into the more general framework of causal team semantics. We consider the taxonomy of causal notions that was introduced by Hitchcock in [21], but we stick to the terminology used by Woodward [45]. This classification includes three notions of type causation (direct, total, and contributing cause) and one of token causation (actual cause).

First we consider at length the notion of *direct cause* (and some of its interactions with functional dependencies). Woodward expresses it thus:

A necessary and sufficient condition for X to be a direct cause of Y with respect to some variable set V is that there be a possible intervention on X that will change Y (or the probability distribution of Y) when all other variables in V besides X and Y are held fixed at some value by interventions. ([45], p. 55)

Woodward is ambiguous here as to what “changing Y ” may mean. The remarks that follow the above quote, though, make it clear that what is really meant is that there are *two* possible interventions (of the kind described in the quote) which force Y to take two distinct values.

The assertion $DC_{\sigma}(X; Y)$ that X is a direct cause of Y (relative to the variables in the domain of σ) can then be expressed in the language $\mathcal{C}(\sigma)$ (and its extensions) as follows:

$$\bigvee_{x \neq x', y \neq y', \mathbf{z}} \{[(\mathbf{Z} = \mathbf{z} \wedge X = x) \sqcap \rightarrow Y = y] \wedge [(\mathbf{Z} = \mathbf{z} \wedge X = x') \sqcap \rightarrow Y = y']\}$$

³⁵Other examples can be devised in which just one of either of these two factors suffices to break a functional dependency.

where $\mathbf{Z} = \text{Dom} \setminus \{X, Y\}$ and x, x', y, y', \mathbf{z} vary within the ranges of the corresponding variables.

It would seem more natural to express this condition as a Boolean disjunction; the following results prove the adequacy of the definition in terms of tensor disjunction, and something more: that its truth in a causal team is determined by the functional component alone. The main point which makes this definition work in accordance with the usual behaviour of direct cause is the fact that, if X is really a direct cause of Y , then the specific interventions involved in the definition always produce singleton teams. For brevity, in the proofs we write $A_{x,y,\mathbf{z}}$ for $(\mathbf{Z} = \mathbf{z} \wedge X = x) \sqcap \rightarrow Y = y$. Since some ambiguity may arise when we discuss multiple teams, we write here PA_X^T for the parents of X in the causal team T .

Lemma 5.1 *Let T be a recursive causal team of signature σ , and suppose $T \models DC_\sigma(X, Y)$. Then $X \in PA_Y^T$, and it is not a dummy argument of the function $\mathcal{F}_T(Y)$.*

Proof Suppose $X \notin PA_Y^T$ or X is a dummy argument of $\mathcal{F}_T(Y)$: then, by the definition of $\mathcal{F}_T(Y)$ and of intervention, for every x, x', y, \mathbf{z} we have that $T \models A_{x,y,\mathbf{z}}$ entails $T \models A_{x',y,\mathbf{z}}$; so $T \not\models A_{x',y',\mathbf{z}}$. This contradicts the initial assumption that $T \models DC_\sigma(X, Y)$. \square

Lemma 5.2 *Let T, S be recursive causal teams of signature σ such that $\mathcal{F}_T = \mathcal{F}_S$. Suppose $T \models DC_\sigma(X, Y)$. Then $S \models DC_\sigma(X, Y)$.*

In words, causal teams which have the same signature and the same associated equations also encode the same relationships of direct cause.

Proof Suppose $T \models DC_\sigma(X, Y)$; then, by Lemma 5.1, $X \in PA_Y^T$ and is not a dummy argument of $\mathcal{F}_T(Y) = \mathcal{F}_S(Y)$. Therefore, $X \in PA_Y^S$ by the compatibility condition (*) from the definition of causal team.

$T \models DC_\sigma(X, Y)$ entails, by the clause for \vee , that there is a causal subteam of T , call it T' , such that $T' \models A_{x,y,\mathbf{z}} \wedge A_{x',y',\mathbf{z}}$ for some $x \neq x', y \neq y', \mathbf{z}$. Since $G_T = G_{T'}$, $X \in PA_Y^T$ entails $X \in PA_Y^{T'}$. Since $X \in PA_Y^{T'}$ and $X \in PA_Y^S$, we have that $T'_{\mathbf{Z}=\mathbf{z} \wedge X=x}$ and $S_{\mathbf{Z}=\mathbf{z} \wedge X=x}$ are the same causal team with singleton team component; the same holds for $T'_{\mathbf{Z}=\mathbf{z} \wedge X=x'}$ and $S_{\mathbf{Z}=\mathbf{z} \wedge X=x'}$. Therefore, from $T' \models A_{x,y,\mathbf{z}} \wedge A_{x',y',\mathbf{z}}$ we can infer $S \models A_{x,y,\mathbf{z}} \wedge A_{x',y',\mathbf{z}}$. Since the causal subteam of S with empty team component satisfies each of the other disjuncts of $DC_\sigma(X, Y)$, we can conclude that $S \models DC_\sigma(X, Y)$. \square

Theorem 5.3 *Let T be a recursive causal team of signature $\sigma = (\text{Dom}, \text{Ran})$, and $\mathbf{Z} := \text{Dom} \setminus \{X, Y\}$. The following are equivalent:*

1. $T \models DC_\sigma(X, Y)$.
2. There are $x, x' \in \text{Ran}(X), y, y' \in \text{Ran}(Y), \mathbf{z} \in \text{Ran}(\mathbf{Z})$ such that $x \neq x', y \neq y'$ and $T \models [(\mathbf{Z} = \mathbf{z} \wedge X = x) \sqcap \rightarrow Y = y] \wedge [(\mathbf{Z} = \mathbf{z} \wedge X = x') \sqcap \rightarrow Y = y']$.
3. There are $x, x' \in \text{Ran}(X), y, y' \in \text{Ran}(Y), \mathbf{z} \in \text{Ran}(\mathbf{Z})$ such that $x \neq x', y \neq y'$ and, for every causal team S of signature σ and functional component \mathcal{F}_T , $S \models [(\mathbf{Z} = \mathbf{z} \wedge X = x) \sqcap \rightarrow Y = y] \wedge [(\mathbf{Z} = \mathbf{z} \wedge X = x') \sqcap \rightarrow Y = y']$.

Proof 2. follows from 3. as a special case.

We show that 1. entails 3. Suppose 1. holds, i.e. $T \models DC_\sigma(X, Y)$. Let S be a causal team of signature σ and functional component \mathcal{F}_T . From Lemma 5.2, we get $S \models DC_\sigma(X, Y)$. By the clause for \vee , there is a causal subteam S' of S and $x \neq x'$, $y \neq y'$, \mathbf{z} such that $S' \models A_{x,y,\mathbf{z}} \wedge A_{x',y',\mathbf{z}}$. Furthermore, by downward closure we have $S' \models DC_\sigma(X, Y)$, and thus from Lemma 5.1 we get that $X \in PA_Y^{S'}$, from which it immediately follows that $S'_{\mathbf{Z}=\mathbf{z} \wedge X=x}$ and $S'_{\mathbf{Z}=\mathbf{z} \wedge X=x'}$ have singleton team components. Clearly, if S^* is any causal subteam of S with singleton team component, applying the same two interventions to S^* produces exactly the same pair of causal teams as is obtained by intervening on S' . Therefore, $S^* \models A_{x,y,\mathbf{z}} \wedge A_{x',y',\mathbf{z}}$. By flatness, we get $S \models A_{x,y,\mathbf{z}} \wedge A_{x',y',\mathbf{z}}$. So 3. holds.

The entailment from 2. to 1. immediately follows from the clause for \vee and the fact that causal teams with empty team component satisfy all \mathcal{CO} formulas. \square

We want to show that the notion of direct cause has some kind of logical interaction with dependence atoms. It is to be expected that, if we know that \mathbf{X} is the set of all direct causes of Y , then we can infer that \mathbf{X} functionally determines Y . We write $All-DC_\sigma(\mathbf{X}; Y)$ for the statement that \mathbf{X} is the set of *all* direct causes of Y .³⁶ That is, let $All-DC_\sigma(X; Y)$ abbreviate

$$\bigwedge_{X \in \mathbf{X}} DC_\sigma(X; Y) \wedge \bigwedge_{Z \notin \mathbf{X}} \sim DC_\sigma(Z; Y).$$

Here \sim is contradictory negation (this requires an enrichment of \mathcal{COD}). Now it can be shown that the rule

$$\frac{All-DC_\sigma(\mathbf{X}; Y)}{=\langle \mathbf{X}; Y \rangle}.$$

is a valid inference rule in the language \mathcal{COD} enriched with \sim .

The notion of *total cause* may seem more difficult to express in our languages; part of the difficulty lies in an even greater ambiguity with which it is expressed in the philosophical literature. In Woodward's words:

X is a total cause of Y if and only if there is a possible intervention on X that will change Y or the probability distribution of Y . ([45], p. 51)

One way to interpret this statements is that, for *some* assignment of values to the variables in the system, some intervention on X will change the value of Y . Now, it is straightforward to prove that this statement is equivalent to the assertion that after *fixing* the *nondescendants* of X in some appropriate way, there are two interventions on X which produce two distinct values for Y . In this formulation, " X is a total cause of Y " (relative to a signature σ) can be expressed by a \mathcal{CO} formula which has the same form as $DC_\sigma(X, Y)$, except that \mathbf{Z} is now the set of nondescendants of \mathbf{X} . The appropriateness of this definition over causal teams is due to the fact that, when there is a chain of direct causes from X to Y , the interventions mentioned in the definition

³⁶In part of the literature, \mathbf{X} would then be called the *parent set* of Y . The notion of parent that we have been using throughout the paper is looser, in that we have also been allowing the parent set of a variable Y to contain dummy arguments of the function that generates Y – which are obviously not direct causes.

produce singleton teams. We omit a detailed discussion of this point, but we illustrate it with an example. Consider the causal team

$$T:$$

$X \rightarrow Z \rightarrow Y \leftarrow W$			
0	0	0	0
1	1	0	1

with equations $Z := X$ and $Y := Z \text{ XOR } W$. Here no intervention on X alone can produce a team where Y has a constant value, as the tables show:

$T_{X=0}:$	$X \rightarrow Z \rightarrow Y \leftarrow W$			
	0	0	0	0
	0	0	1	1

$T_{X=1}:$	$X \rightarrow Z \rightarrow Y \leftarrow W$			
	1	1	1	0
	1	1	0	1

However, if we fix W (the only nondescendant of X) e.g. to 0, then further interventions on X permit to settle the value of Y (in two distinct ways):

$T_{W=0 \wedge X=0}:$	$X \rightarrow Z \rightarrow Y \leftarrow W$			
	0	0	0	0
	0	0	0	0

$T_{W=0 \wedge X=1}:$	$X \rightarrow Z \rightarrow Y \leftarrow W$			
	1	1	1	0
	1	1	1	0

Thus X is a total cause of Y .

The third and perhaps most important notion of type causation is called by Woodward *contributing* cause. We say that “ X is a contributing cause of Y ” ([45], p. 59) if there is a chain of direct causes from X to Y , and, while holding fixed in an appropriate way the variables that are *not* on this path, one can make two distinct interventions on X which produce two distinct values of Y . It should be clear, at this point, that this notion can be written in \mathcal{CO} in a similar fashion to direct cause and total cause (adding an external tensor disjunction over paths of direct causes from X to Y , and taking \mathbf{Z} to be the set of variables not occurring on the path).³⁷

Our framework seems to be less adequate for dealing with token causation, i.e., with the causes of a specific event. A typical question of this kind might take the form:

It happened that $Y = y$. Is the fact that $X = x$ a cause of it?

What is usually intended here is that $X = x$ and $Y = y$ are true in the *actual world*. Furthermore, most of the notions of actual cause on the market (e.g., Woodward [45], Halpern and Pearl [17]) rely crucially on the notion of “intervention that fixes a variable to its actual value”. These kinds of devices are lacking in our version of causal teams. And the presence of teams introduces complications of its own, especially in the case of nonrecursive systems: if the system of equations has multiple solutions, it is not always possible to recompute deterministically what the new “actual world” will be after an intervention. For this reason, we think that the modelling of

³⁷In Section 4.1 we have considered a further causal notion from [16], “ X affects Y ”. We believe this simpler notion to be equivalent to contributing cause, but we have not seen any explicit comparison in the literature.

token causation should proceed differently on causal team semantics, replacing talk of actuality by talk of observations, as in the scenario:

We see that $X = x$ and $Y = y$. It is entailed by/compatible with our observations that $X = x$ is a cause $Y = y$.

To model such a scenario we propose a notion of cause that we may call *discernible cause*, derived from the simplest definition of actual causation from [45]. We are aware that such a definition is in need of various refinements in order to treat correctly e.g. cases of overdetermination. We say that $X = x$ is a discernible cause of $Y = y$ if

$$(X = x \wedge Y = y) \supset \left\{ \bigvee_{\mathbf{P}} \bigvee_{\substack{\mathbf{Z} = \mathbf{Dom} \setminus \mathbf{P} \\ x \neq x', y \neq y', \mathbf{z}}} [\mathbf{Z} = \mathbf{z} \supset (A_{x,y,\mathbf{z}} \wedge A_{x',y',\mathbf{z}})] \right\}$$

holds, where \mathbf{P} varies over sets of variables which form a directed path of direct causes from X to Y . The most external implication, $(X = x \wedge Y = y) \supset \dots$ can be read: “in case we observed $X = x$ and $Y = y$ ”. The formula inside square parentheses can be rewritten as a conjunction of formulas of the form $\mathbf{Z} = \mathbf{z} \supset (\mathbf{Z} = \mathbf{z} \Box \rightarrow \psi)$; this is our surrogate for “intervening on the actual value of \mathbf{Z} ”. What such a formula says is that, if the actual value of \mathbf{Z} were \mathbf{z} , then fixing \mathbf{Z} to \mathbf{z} would yield ψ .

The proposed notion of discernible cause explicitly incorporates epistemic aspects. The comments above should illustrate that one possible reading of our selective implication \supset is as an epistemic subjunctive conditional.

6 Conclusions and Future Developments

We have seen that, even though team semantics, by itself, is insufficient to support interventionist counterfactuals and causal dependencies, it can be extended to overcome this limitation. The resulting causal team semantics enriches each team with components that encode structural equations, and induces a notion of intervention on a team (a procedure that rewrites parts of the team when one of the variables in its domain is acted upon).

This notion of intervention was the main ingredient in our semantic clause for interventionist counterfactuals. We studied the logical principles that govern these counterfactuals by comparison with, on one side, Lewis-Stalnaker counterfactuals, and interventionist counterfactuals over causal models, on the other. In particular, we made some advances in the comparison of *nested* interventionist counterfactuals vs. the Lewisian ones, as it was initiated in [3]; we found more Lewisian principles that can be falsified in the interventionist frameworks, for example *weak transitivity* and *Modus tollens*.

We also used our interventionist counterfactuals to define various notions of cause in Woodward’s style. The team component of causal teams led to interesting generalizations of some of the basic propositional connectives (e.g. tensor, selective implication). It also made it possible to express functional dependencies in the language. Many other operators and notions of dependence have been considered in the

literature on team semantics, and it might be worthwhile to consider their interactions with counterfactuals in the context of future work.

We hope that the paper has shown that causal teams already bring something new when they are applied to “classical” languages, such as the language \mathcal{CO} . The novelty does not lie in the resulting logic (which, save for minor details, is the same as for causal models); it consists, instead, in the possibility of modelling more scenarios, as illustrated by the examples in Section 3. The examples also show that, in nonrecursive systems, it is necessary to consider causal teams with multiple assignments in order to compute interventions over causal models.

In a continuation of this paper we will show how these ingredients can be made to work together with probabilities in the framework of causal multiteams, supporting languages that are compatible with the framework of semi-deterministic SEMs. The operators \supset and \sqcup , which have played a mild role so far, will be seen to be crucial. In particular, we will see that the selective implication can be used to decompose statements of conditional probabilities into more basic statements, and allows expressing forms of conditioning (mixed with interventions) that are more general, as far as we know, than those considered in the literature on structural equation models.

Concerning the deterministic case considered in this paper, future work will have to settle the issue of axiomatizability of the languages containing nonclassical operators, such as functional dependence and the “Boolean disjunction” \sqcup . The literature on propositional dependence logic (in particular [34, 47]) suggests that complete deduction system should be possible to achieve. Towards an adaptation of the results of [47], it might be necessary to find exact characterizations of the expressive power of our formal languages. Finally, the case of nonrecursive causal teams, only sketched here, will need to be more systematically studied, both from a semantic and a proof-theoretic point of view.

Acknowledgements Open access funding provided by University of Helsinki including Helsinki University Central Hospital. The research was conducted under the aid of the Academy of Finland grants n. 286991 and n. 316460. The authors wish to thank Ivano Ciardelli for carefully reading and commenting upon an earlier version of the paper; and the anonymous reviewers and the editor for many significant suggestions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix A: Distance Between Variables in a Graph

We specify here the notion of distance that we have used in the definition of the *do* algorithm. Suppose we are given a graph $G = (\mathbf{V}, E)$; in the context of this paper, the

vertices \mathbf{V} are thought of as variables. The general intuition is that, after an intervention on X , other variables should be updated; and variables at distance $n + 1$ can only be updated after the variables at distance $\leq n$ have been updated.

We define the distance between two variables X and Y to be the maximum length of a directed path from X to Y :

$$d_G(X, Y) = \sup\{ \text{length}(P) \mid P \text{ dir. path of } G \text{ going from } X \text{ to } Y \}$$

if any such paths exists; or -1 otherwise.

Clearly, if the graph is finite and acyclic, $d_G(X, Y) \in \mathbb{N} \cup \{-1\}$ for any pair X, Y . We omit the subscript G when the graph is clear from the context. Notice that in an acyclic graph $d_G(X, Y) \neq d_G(Y, X)$ whenever X and Y are distinct variables.

The algorithm described in Section 2 uses this notion of distance in order to define interventions of the form $do(X = x)$. However, as we saw, we needed to define the more general notion of intervention of the form $do(\mathbf{X} = \mathbf{x})$. This can be achieved by defining a notion of distance between a *set* of variables \mathbf{X} and a variable Y ; we denote it by $d_G(\mathbf{X}, Y)$. There are two ideas involved in this generalization. The first is that we should not count the directed paths that are disrupted by the intervention itself (an intervention $do(\mathbf{X} = \mathbf{x})$ removes all arrows coming into variables of \mathbf{X}); for this purpose, we need to consider the reduced graph $G^{-\mathbf{X}} = (\mathbf{V}, E_{-\mathbf{X}})$, the graph obtained by removing all arrows going into some vertex of \mathbf{X} (i.e., an edge (V_1, V_2) is in $E_{-\mathbf{X}}$ iff it is in E and $V_2 \notin \mathbf{X}$). The second idea is that we should take the maximum of the individual distances between each variable X of \mathbf{X} and Y (relative to the reduced graph). Thus we obtain

$$d_G(\mathbf{X}, Y) = \sup\{ \text{length}(P) \mid P \text{ dir. path of } G^{-\mathbf{X}} \text{ going from some } X \in \mathbf{X} \text{ to } Y \}.$$

Let $\mathbf{X} = \mathbf{x}$ be a *consistent* conjunction (i.e. it does not contain pairs of formulas $X = x$, $X = x'$ with $x \neq x'$). An algorithm for $do(\mathbf{X} = \mathbf{x})$ is then obtained by modifying the algorithm for $do(X = x)$ as follows:

- Stage 0 is replaced with: produce the team T_0 by replacing each assignment $s \in T^-$ with $s(\mathbf{x}/\mathbf{X})$.
- Replace $d_G(X, Y)$ with $d_G(\mathbf{X}, Y)$.

We do not define the algorithm for inconsistent conjunctions.

Using Theorem 4.1, it can be proved (see [1]) that the intervention $do(\mathbf{X} = \mathbf{x})$ (for consistent $\mathbf{X} = \mathbf{x}$) could be equivalently defined as the sequential application of individual interventions $do(X = x)$, for each conjunct $X = x$ in $\mathbf{X} = \mathbf{x}$, applied in any order.

Appendix B: Proof of the Termination of the *do* Algorithm

Theorem B.1 *Let T be a recursive causal team. If G_T is a finite acyclic graph, then $T_{\mathbf{X}=\mathbf{x}}$ is well-defined.*

Proof Assume, at first, that T^- is finite.

Suppose that, for $\mathbf{X} \subseteq \text{Dom}_T, Y \in \text{Dom}_T, d(\mathbf{X}, Y) > \text{card}(G_T)$. Then there is a path P in $G_T^{-\mathbf{X}}$ from some $X \in \mathbf{X}$ to Y such that $\text{length}(P) > \text{card}(G_T)$. Then there is a node which is crossed at least twice by P ; so $G_T^{-\mathbf{X}}$ contains a cycle. Therefore, the larger graph G_T contains a cycle, and we have a contradiction. Thus, $\sup\{d(\mathbf{X}, Z) \mid Z \in \text{Dom}_T\} \leq \text{card}(G_T)$; this means that the “for” cycle in the algorithm goes through a finite number of iterations over the variable n .

Finally, notice that, for each n , the number of variables Z such that $d(\mathbf{X}, Z) = n$ is finite (due to finiteness of G_T), as is the number of assignments t in the team that is undergoing modification (due to the finiteness of T^-). Therefore, the algorithm terminates after a finite number of steps.

If instead T^- is infinite, $(T_{\mathbf{X}=\mathbf{x}})^-$ is produced by an infinitary algorithm which, in each iteration of the “for” cycle, performs simultaneously the substitution $t \mapsto t(f_{Z_1}(s(PA_{Z_1}))/Z_1, \dots, f_{Z_{k_{n+1}}}(s(PA_{Z_{k_{n+1}}})))/Z_{k_{n+1}})$ for all the assignments t in the current team. By the arguments above, such an “algorithm” terminates and yields a well-defined causal team. \square

Appendix C: Complete Axiomatization for Language \mathcal{CO}

As pointed out in the main text, for any given signature $\sigma = (\text{Dom}, \text{Ran})$, a complete inference system for the language $\mathcal{CO}(\sigma)$ (interpreted over causal teams of signature σ) can be obtained by adapting and slightly extending the axiomatization proposed by Halpern [16] for a language interpreted over causal models. See Section 4.1 for the list of axioms and rules.

Under many respects, our framework is closer to that of Briggs [3], in that 1) our notion of signature does not encode a distinction between exogenous and endogenous variables, 2) our language treats exogenous and endogenous variables on a par, e.g. by allowing interventions over both. Because of these differences from Halpern’s approach, our strategy for obtaining the completeness proof will be closer to that used in [3].

The notion of consequence that we want to axiomatize is $\Gamma \models_{\sigma} \psi$: every recursive causal team of signature σ which satisfies all formulas in Γ also satisfies ψ . We write $\Gamma \vdash_{\sigma} \psi$ to say that ψ is the end formula of some proof in $\text{Ax}_{\mathcal{CO}}(\sigma)$ which possibly uses extra premises from Γ .

The soundness of the axioms can be proved by observing their soundness over causal teams with singleton team components; and then using the flatness of $\mathcal{CO}(\sigma)$ (Theorem 2.10). The soundness of many of the rules has been addressed in the main text; the following lemmas yield a general method for proving the soundness of the inference rules of $\text{Ax}_{\mathcal{CO}}(\sigma)$.

Lemma C.1 *The inference rules of $\text{Ax}_{\mathcal{CO}}(\sigma)$ are sound for recursive causal teams with singleton team component.*

Given the obvious identification of deterministic structural equation models with causal teams with singleton team component, this soundness result follows from the

literature [3, 11, 16]. Next we show how to extend this result to all recursive causal teams.

Lemma C.2 *Let $\Gamma \cup \{\psi\} \subseteq \mathcal{CO}(\sigma)$, and suppose that every causal team which has singleton team component and satisfies Γ also satisfies ψ . Then $\Gamma \models_{\sigma} \psi$.*

Proof Let T be a recursive causal team, and suppose that $T \models \Gamma$ and $T \not\models \psi$. From the latter, plus flatness of ψ (Theorem 2.10), it follows that there is an $s \in T^{-}$ such that $(*) \{s\} \not\models \psi$. By $T \models \Gamma$ and the flatness of Γ , it follows that $(**) \{s\} \models \Gamma$. $(*)$ and $(**)$, together, contradict the assumption that Γ entails ψ over recursive causal teams with singleton component. \square

We point out that the following principles are also sound for recursive causal teams (as can be proved by the two previous lemmas):

- *Composition 2:*
$$\frac{\mathbf{X}=\mathbf{x} \sqsupset W=w \quad (\mathbf{X}=\mathbf{x} \wedge W=w) \sqsupset Y=y}{\mathbf{X}=\mathbf{x} \sqsupset Y=y}.$$
- *Reversibility:*
$$\frac{(\mathbf{X}=\mathbf{x} \wedge Y=y) \sqsupset W=w \quad (\mathbf{X}=\mathbf{x} \wedge W=w) \sqsupset Y=y}{\mathbf{X}=\mathbf{x} \sqsupset Y=y} \text{ (for } Y \neq W\text{)}.$$

Composition 2 is a “converse” to *Composition 1*. It was implicitly part of the axiom system of Galles and Pearl [11], but [16] showed that it is unnecessary for a completeness proof.

What we now show is that (for each signature σ) the system $AX_{\mathcal{CO}}(\sigma)$ is complete for \models_{σ} . The strategy will be to show that every consistent set of formulas has a model; we mostly follow the method used in [3]. There is some vagueness in the literature as to what notion of (proof-theoretic) consistency is involved here; the notion we will use is given in the definition below. But first a notational convention. Given a set Γ of $\mathcal{CO}(\sigma)$ formulas, write Γ_{σ}^{+} for its closure under deduction in $AX_{\mathcal{CO}}(\sigma)$.

Definition C.3 A set Γ of $\mathcal{CO}(\sigma)$ formulas is σ -consistent if Γ_{σ}^{+} does not contain any pair of formulas of the form $\psi, \neg\psi$.

Γ is *maximally σ -consistent* if it is σ -consistent and every $\Gamma' \supset \Gamma$ is not σ -consistent.

In the following we shall simply write “consistent” for σ -consistent, since the relevant signature will be clear from the context.

We note that, since $AX_{\mathcal{CO}}(\sigma)$ contains all instances of classical tautologies and modus ponens, one can prove the syntactical deduction theorem as usual.

Lemma C.4 *Let ψ, χ be $\mathcal{CO}(\sigma)$ formulas. Then*

$$\Gamma \vdash_{\sigma} \psi \supset \chi \iff \Gamma, \psi \vdash_{\sigma} \chi.$$

Lemma C.5 *Let Γ be a consistent set of $\mathcal{CO}(\sigma)$ formulas, and ψ a $\mathcal{CO}(\sigma)$ formula. Suppose $\Gamma \not\models_{\sigma} \psi$; then $\Gamma \cup \{\neg\psi\}$ is consistent.*

Proof Suppose $\Gamma \cup \{\neg\psi\}$ is not consistent; then $\Gamma \cup \{\neg\psi\} \vdash_{\sigma} \chi$ and $\Gamma \cup \{\neg\psi\} \vdash_{\sigma} \neg\chi$ for some $\mathcal{CO}(\sigma)$ formula χ . Then, using Lemma C.4, we get that $\Gamma \vdash_{\sigma} \neg\psi \supset \chi$ and $\Gamma \vdash_{\sigma} \neg\psi \supset \neg\chi$. Using classical logic we get $\Gamma \vdash_{\sigma} \psi$, contradicting the assumption. \square

Using this result, one can prove as usual a version of Lindenbaum's lemma.

Lemma C.6 *Let Γ be a consistent set of $\mathcal{CO}(\sigma)$ formulas. Then there is a maximally consistent set Δ of $\mathcal{CO}(\sigma)$ formulas such that $\Gamma \subseteq \Delta$.*

Lemma C.5 also entails that maximally consistent sets of formulas are syntactically complete:

Corollary C.7 *Let Δ be a maximally consistent set of $\mathcal{CO}(\sigma)$ formulas. Then, for every $\psi \in \mathcal{CO}(\sigma)$, either $\psi \in \Delta$ or $\neg\psi \in \Delta$.*

In the next lemma we prove that a maximally consistent set of $\mathcal{CO}(\sigma)$ formulas “uniquely determines” the outcome of interventions.

Lemma C.8 *Let $\sigma = (Dom, Ran)$ be a signature, $\mathbf{X} \cup \{Y\} \subseteq Dom$, $\mathbf{X} \neq \emptyset$. Let Δ be a maximally consistent set of $\mathcal{CO}(\sigma)$ formulas.*

1. *There is exactly one $y \in Ran(Y)$ such that “ $Y = y$ ” $\in \Delta$.*
2. *For every $\mathbf{x} \in Ran(\mathbf{X})$ there is exactly one $y \in Ran(Y)$ such that “ $\mathbf{X} = \mathbf{x} \sqsupset Y = y$ ” $\in \Delta$.*

Proof Let us prove 2. Let $\mathbf{x} \in Ran(\mathbf{X})$. First we show that Δ can contain a formula of the form $\mathbf{X} = \mathbf{x} \sqsupset Y = y$ for at most one value $y \in Ran(Y)$. Indeed, let y' be a distinct value in $Ran(Y)$; observe that Δ is closed under \vdash_{σ} , and thereby under applications of the *Uniqueness* rule; so, since “ $\mathbf{X} = \mathbf{x} \sqsupset Y = y$ ” $\in \Delta$, we obtain that “ $\mathbf{X} = \mathbf{x} \sqsupset Y \neq y'$ ” $\in \Delta$. But notice that this formula is $\neg(\mathbf{X} = \mathbf{x} \sqsupset Y = y')$ (by definition of the dualization). So, by the consistency of Δ , “ $\mathbf{X} = \mathbf{x} \sqsupset Y = y'$ ” $\notin \Delta$.

We then have to show that there is at least one $y \in Ran(Y)$ such that “ $\mathbf{X} = \mathbf{x} \sqsupset Y = y$ ” $\in \Delta$. For suppose this is not the case; then by maximality of Δ , corollary C.7 and the definition of dual negation, “ $\mathbf{X} = \mathbf{x} \sqsupset Y \neq y$ ” $\in \Delta$ for each $y \in Ran(Y)$; thus “ $\bigwedge_{y \in Ran(Y)} (\mathbf{X} = \mathbf{x} \sqsupset Y \neq y)$ ” $\in \Delta$. But this formula is just $\neg(\bigvee_{y \in Ran(Y)} (\mathbf{X} = \mathbf{x} \sqsupset Y = y))$; and “ $\bigvee_{y \in Ran(Y)} (\mathbf{X} = \mathbf{x} \sqsupset Y = y)$ ” $\in \Delta$ by the *Definiteness* axiom. Since Δ is consistent, we have reached a contradiction.

The proof of 1. is analogous, using instances of *Uniqueness* and *Definiteness* with empty antecedents. \square

We will need two substitution lemmas to guarantee that rules need not be applied to the outermost connectives of a formula. The first is for formulas that occur in positive contexts. We write $\varphi[\theta]$ for a formula if we want to highlight that it has an (occurrence of) a subformula θ . If θ' is yet another formula, $\varphi[\theta']$ will denote the formula which results by substituting the occurrence θ with θ' in φ .

Lemma C.9 Let $\varphi[\theta], \theta' \in \mathcal{CO}(\sigma)$ and let θ denote a specific occurrence of a subformula of $\varphi[\theta]$. Assume further that θ is not occurring in any antecedent of \supset or $\Box \rightarrow$. If $\theta \vdash_{\sigma} \theta'$, then $\varphi[\theta] \vdash_{\sigma} \varphi[\theta']$.

Proof We prove it by induction on $\varphi[\theta]$, including more generally the possibility that θ does not occur at all in $\varphi[\theta]$. First base case: $\varphi[\theta]$ is θ . Then $\theta \vdash_{\sigma} \theta'$ by assumption, and θ' is $\varphi[\theta']$. Second base case: $\varphi[\theta]$ does not contain the intended occurrence of θ . Then $\varphi[\theta'] = \varphi[\theta]$, from which obviously $\varphi[\theta] \vdash_{\sigma} \varphi[\theta']$.

Inductive step: $\varphi[\theta]$ is $\psi[\theta] \circ \chi[\theta]$, where \circ is one of the binary connectives of $\mathcal{CO}(\sigma)$. By the inductive hypothesis, $\psi[\theta] \vdash_{\sigma} \psi[\theta']$ and $\chi[\theta] \vdash_{\sigma} \chi[\theta']$. The cases for \wedge and \vee follow easily using classical tautologies; we prove the remaining cases.

- \circ is \supset . Since the occurrence θ is not in $\psi[\theta]$, we have $\psi[\theta] = \psi[\theta']$; but then $\neg\psi[\theta] \vdash_{\sigma} \neg\psi[\theta']$. Furthermore, by the inductive hypothesis, $\chi[\theta] \vdash_{\sigma} \chi[\theta']$. By lemma C.4 we get $\vdash_{\sigma} \neg\psi[\theta'] \supset \neg\psi[\theta]$ and $\vdash_{\sigma} \chi[\theta] \supset \chi[\theta']$.

Since $\vdash_{\sigma} (\neg\psi[\theta'] \supset \neg\psi[\theta]) \supset ((\chi[\theta] \supset \chi[\theta']) \supset ((\psi[\theta] \supset \chi[\theta]) \supset (\psi[\theta'] \supset \chi[\theta'])))$, as this is an instance of a classical tautology, by two applications of *Modus ponens* we get $\vdash_{\sigma} (\psi[\theta] \supset \chi[\theta]) \supset (\psi[\theta'] \supset \chi[\theta'])$. By lemma C.4, then, $\psi[\theta] \supset \chi[\theta] \vdash_{\sigma} \psi[\theta'] \supset \chi[\theta']$.

- \circ is $\Box \rightarrow$. Since by inductive hypothesis we have $\chi[\theta] \vdash_{\sigma} \chi[\theta']$, and furthermore $\psi[\theta] = \psi[\theta']$, by applying the *C-substitution* rule we obtain:

$$\psi[\theta] \Box \rightarrow \chi[\theta] = \psi[\theta'] \Box \rightarrow \chi[\theta] \vdash_{\sigma} \psi[\theta'] \Box \rightarrow \chi[\theta']. \quad \square$$

The second substitution lemma guarantees substitution of equivalents in the antecedents of counterfactuals occurring in the context of a complex formula.

Lemma C.10 Let $\psi[\mathbf{X} = \mathbf{x}]$ be a $\mathcal{CO}(\sigma)$ formula without occurrences of \supset . Suppose the occurrence of $\mathbf{X} = \mathbf{x}$ is in the antecedent of a counterfactual occurring in $\psi[\mathbf{X} = \mathbf{x}]$. If $\mathbf{X} = \mathbf{x} \dashv\vdash_{\sigma} \mathbf{Z} = \mathbf{z}$, then $\psi[\mathbf{X} = \mathbf{x}] \dashv\vdash_{\sigma} \psi[\mathbf{Z} = \mathbf{z}]$.

Proof The proof is analogous to that of lemma C.9, using *A-substitution* instead of *C-substitution*. \square

The next lemma shows how to build an appropriate “model” for any maximally consistent set of $\mathcal{CO}(\sigma)$ formulas.

Lemma C.11 Let $\sigma = (\text{Dom}, \text{Ran})$ be a signature, and Δ be a maximally consistent set of $\mathcal{CO}(\sigma)$ formulas. Then there is a recursive causal team T of signature σ such that T^- is a singleton, and $T \models \Delta$.

Proof 1) Defining T . Suppose first that $\text{card}(\text{Dom}) \geq 2$. For each variable $V \in \text{Dom}$, let \mathbf{W}_V be the list of variables in $\text{Dom} \setminus \{V\}$, arranged in some order. To each $V \in \text{Dom}$ we associate a function $g_V : \text{Ran}(\mathbf{W}_V) \rightarrow \text{Ran}(V)$ as follows:

$$g_V(\mathbf{w}) = v \text{ iff } \mathbf{W}_V = \mathbf{w} \Box \rightarrow V = v \in \Delta.$$

We must verify that this function is correctly defined; but this immediately follows from lemma C.8.

Now define PA_V as the set of nonredundant arguments of g_V ; that is, the smallest subset $\mathbf{Z} \subseteq \mathbf{W}_V$ such that, for every fixed tuple $\mathbf{z} \in \text{Ran}(\mathbf{Z})$, $g_V(\mathbf{z}, \cdot)$ is a constant function. Define the graph $G_T := (\text{Dom}, E)$, where $(X, Y) \in E$ iff $X \in PA_Y$. By the definition of g_Y , we have that $Y \notin PA_Y$; so, G_T is irreflexive.

Let $\text{End}(T)$ be the set of variables $V \in \text{Dom}$ such that $PA_V \neq \emptyset$. For each $V \in \text{End}(T)$, define $f_V : \text{Ran}(PA_V) \rightarrow \text{Ran}(V)$ as g_V deprived of the redundant parameters (i.e., for all $pa \in \text{Ran}(PA_V)$, $f_V(pa) := g_V(pa, \mathbf{u})$ for any $\mathbf{u} \in \text{Ran}(\mathbf{W}_V \setminus PA_V)$). Let $\mathcal{F}_T := \{(V, f_V) \mid V \in \text{End}(T)\}$.

By lemma C.8 again, we have that, for each $V \in \text{Dom}$, Δ contains exactly one formula of the form “ $V = v$ ”, for some $v \in \text{Ran}(V)$. Let s be the unique assignment of signature σ such that, for each $V \in \text{Dom}$, $s(V) = v$ iff “ $V = v$ ” $\in \Delta$. Define $T^- := \{s\}$.

Now T is defined as the triple $(T^-, G_T, \mathcal{F}_T)$. In order to show that T is a causal team (of endogenous variables $\text{End}(T)$), we need to prove that T^- and \mathcal{F}_T are compatible, i.e. they respect the clause (*) of definition 2.1. Let $V \in \text{End}(T)$; let $v = s(V)$ and $\mathbf{w} = s(\mathbf{W}_V)$. Now by definition of s we have that “ $V = v$ ” $\in \Delta$ and “ $\mathbf{W}_V = \mathbf{w}$ ” $\in \Delta$. Let pa be the restriction of \mathbf{w} to PA_V . Then by *Composition 1* (for assumptions with empty antecedents) and the fact that Δ is closed under derivations, we have “ $\mathbf{W}_V = \mathbf{w} \sqsupset V = v$ ” $\in \Delta$. But then, by definition of g_V , we have $g_V(\mathbf{w}) = v$, from which $f_V(pa) = v$, as desired.³⁸

In case $\text{card}(\text{Dom}) = 1$, we let G_T and \mathcal{F}_T be empty, and define s and T as above.

- 2) T is recursive. First we show that, if $X \in PA_Y$, then “ $X \rightsquigarrow Y$ ” $\in \Delta$. Write \mathbf{Z} for $\text{Dom} \setminus \{X, Y\}$. Assume $X \in PA_Y$. By the construction of PA_Y , this means that there is a $\mathbf{z} \in \text{Ran}(\mathbf{Z})$ and there are distinct $x, x' \in \text{Ran}(X)$ and distinct $y, y' \in \text{Ran}(Y)$ such that $y = g_Y(\mathbf{z}, x) \neq g_Y(\mathbf{z}, x') = y'$. By the construction of g_Y , we have

$$“(\mathbf{Z} = \mathbf{z} \wedge X = x) \sqsupset Y = y” \in \Delta$$

and

$$“(\mathbf{Z} = \mathbf{z} \wedge X = x') \sqsupset Y = y'” \in \Delta$$

from which, applying classical logic, *Exportation*, the inverse of *Extraction of \wedge* and the closure of Δ under \vdash_σ , we obtain

$$“\mathbf{Z} = \mathbf{z} \sqsupset [(X = x \sqsupset Y = y) \wedge (X = x' \sqsupset Y = y')]” \in \Delta$$

and applying classical logic and closure under \vdash_σ again, we obtain “ $X \rightsquigarrow Y$ ” $\in \Delta$.

Now, for the sake of contradiction, suppose that for some $X_1, \dots, X_n \in \text{Dom}$ we have $X_i \in PA_{X_{i+1}}$ ($i = 1, \dots, n-1$) and $X_n \in PA_{X_1}$. As shown above, then, we have “ $X_i \rightsquigarrow X_{i+1}$ ” $\in \Delta$ and “ $X_n \rightsquigarrow X_1$ ” $\in \Delta$. Since Δ is closed under

³⁸We thank Kaibo Xie (personal communication) for this argument.

the *Recursivity*ⁿ rule, “ $\neg(X_n \rightsquigarrow X_1)$ ” $\in \Delta$. Since Δ is consistent, we have a contradiction.

- 3) $T \models \Delta$. We have to show that, for each $\psi \in \Delta$, $T \models \psi$. First we show that, for every $\psi \in \Delta$, there is a $\psi' \in \Delta$ such that $\psi \dashv\vdash_{\sigma} \psi'$ and furthermore the following conditions hold:
- All counterfactuals occurring in ψ' have atoms of the form $V = v$ as consequents.
 - ψ' has no occurrences of \supset .
 - Antecedents of counterfactuals in ψ contain at most one occurrence of each variable.
 - The variables in the antecedents of counterfactuals in ψ occur in a fixed alphabetical order.

We show how to derive the desired ψ' from ψ ; the inverse derivation proceeds analogously, using the inverses of the rules (the inverse of $\Box \rightarrow$ -*Extraction* is *Exportation*). First of all we remove all occurrences of \supset by using the rule *Definition of \supset* together with lemma C.9. This has to be done inductively, starting from the most external occurrences of \supset (so that the occurrence of \supset satisfies the assumptions of lemma C.9, i.e. it is not in an antecedent of \supset). In this way we obtain a formula satisfying b). Next we apply the extraction rules (plus lemma C.9) until a) is ensured. Then we use lemma C.10, together with *A-Triviality*, classical tautologies and lemma C.4, to ensure c) and d). $\psi' \in \Delta$ since ψ is, and Δ is closed under \vdash_{σ} .

Since $\psi \dashv\vdash_{\sigma} \psi'$, by the soundness of $AX_{\mathcal{CO}}(\sigma)$ we have that $T \models \psi$ iff $T \models \psi'$. Let us show that, for every $\psi \in \Delta$, $T \models \psi'$; we proceed by induction on the syntax of ψ' . The inductive step has cases for \vee and \wedge (remember that here \neg is just an abbreviation):

- Case $\psi' = \theta \wedge \eta$. Since Δ is closed under \vdash_{σ} , from $\psi' \in \Delta$ it follows that $\theta \in \Delta$ and $\eta \in \Delta$. And notice that θ, η satisfy conditions a)-b)-c)-d), so $\theta' = \theta, \eta' = \eta$. Thus, by inductive hypothesis, $T \models \theta$ and $T \models \eta$. Thus $T \models \theta \wedge \eta$.
- Case $\psi' = \theta \vee \eta$. Since Δ is maximally consistent, it must then contain either θ or η (otherwise, Corollary C.7, it would contain both $\neg\theta$ and $\neg\eta$, thus $\neg\theta \wedge \neg\eta$, thus $\neg(\theta \vee \eta)$, contradicting the consistency of Δ). Say it contains θ . Notice that θ satisfies conditions a)-b)-c)-d), so $\theta' = \theta$. Then, by inductive hypothesis, $T \models \theta$. Since $T^- = T^- \cup \emptyset$, and causal teams with empty team component satisfy η , the semantic clause for \vee yields $T \models \theta \vee \eta$.

The base cases are those for formulas of the form $Y = y$, $Y \neq y$ and $\mathbf{X} = \mathbf{x} \Box \rightarrow Y = y$:

- Case $\psi' = Y \neq y$. If $\psi' \in \Delta$, then, by consistency, “ $Y = y$ ” $\notin \Delta$. Then, by definition of s , $s(Y) \neq y$. Since s is the only assignment in T^- , $T \models Y \neq y$.
- Case $\psi' = \mathbf{X} = \mathbf{x} \Box \rightarrow Y = y$ or $\psi' = Y = y$ or $\psi' = Y = y$ (in this latter case, we define $\mathbf{X} := \emptyset$).

If $Y \in \mathbf{X}$, then ψ' is obviously valid on causal teams; so $T \models \psi'$. If instead $Y \notin \mathbf{X}$, we proceed with a further induction on the number $n = \text{card}(\mathbf{W}_Y \setminus \mathbf{X})$.

- If $n = 0$, notice that $\mathbf{X} = \mathbf{W}_V$ as sequences, thanks to c) and d). Then $T \models \psi'$ by the definition of g_Y together with the semantic clause for $\Box \rightarrow$ (or by the definition of s , in case $\mathbf{X} = \emptyset$).
- Case $n > 0$. In case $\mathbf{X} = \emptyset$, from the assumption that “ $Y = y$ ” $\in \Delta$ we get $s(Y) = y$ by definition of s . Since $T = \{s\}$, then, $T \models Y = y$.

Suppose instead $\mathbf{X} \neq \emptyset$. Now, there is at least one variable $Z \in \text{Dom} \setminus (\mathbf{X} \cup \{Y\})$. By lemma C.8 there is a unique $z \in \text{Ran}(Z)$ such that “ $\mathbf{X} = \mathbf{x} \Box \rightarrow Z = z$ ” $\in \Delta$. Since furthermore $\psi' \in \Delta$ and Δ is closed under the *Composition 1* rule, we conclude that “ $(\mathbf{X} = \mathbf{x} \wedge Z = z) \Box \rightarrow Y = y$ ” $\in \Delta$. Since the antecedent of this counterfactual operates on a larger number of variables than the antecedent of ψ' , we can apply the inductive hypothesis and conclude $T \models (\mathbf{X} = \mathbf{x} \wedge Z = z) \Box \rightarrow Y = y$. Applying *Composition 1* with the premises in opposite order, and again the inductive hypothesis, we obtain $T \models (\mathbf{X} = \mathbf{x} \wedge Y = y) \Box \rightarrow Z = z$. Since recursive causal teams satisfy the *Reversibility* rule we conclude that $T \models \mathbf{X} = \mathbf{x} \Box \rightarrow Y = y$. \square

Theorem C.12 *Let $\Gamma \cup \{\psi\}$ be a set of $\mathcal{CO}(\sigma)$ formulas. If $\Gamma \models_{\sigma} \psi$, then $\Gamma \vdash_{\sigma} \psi$.*

Proof First suppose that Γ is not consistent. This means that $\Gamma \vdash_{\sigma} \chi$ and $\Gamma \vdash_{\sigma} \neg\chi$ for some $\mathcal{CO}(\sigma)$ formula χ . Then, using the classical tautology $\chi \supset (\neg\chi \supset \psi)$ and two applications of *Modus ponens*, one obtains $\Gamma \vdash_{\sigma} \psi$.

Suppose instead Γ is consistent. For the sake of contradiction, assume that $\Gamma \models_{\sigma} \psi$ but $\Gamma \not\vdash_{\sigma} \psi$. Then also $\Gamma \cup \{\neg\psi\}$ is consistent (by lemma C.5). So, by Lindenbaum’s lemma there is a maximally consistent Δ such that $\Gamma \cup \{\neg\psi\} \subseteq \Delta$. Then, by lemma C.11 there is a causal team T of signature σ and nonempty team component T^- such that $T \models \Delta$. Therefore, $T \models \Gamma$ (and therefore $T \models \psi$, since $\Gamma \models_{\sigma} \psi$) and $T \models \neg\psi$. By the non-contradiction property (2.13), we conclude that $T^- = \emptyset$: a contradiction. \square

References

1. Barbero, F., & Sandu, G. (2017). Team semantics for interventionist counterfactuals and causal dependence. arXiv:1610.03406.
2. Barbero, F., & Sandu, G. (2019). Interventionist counterfactuals on causal teams. In Finkbeiner, B., & Kleinberg, S. (Eds.) *Proceedings 3rd Workshop on Formal Reasoning about Causation, Responsibility, and Explanations in Science and Technology, Thessaloniki, Greece, 21st April 2018* (pp. 16–30). vol. 286 of Electronic Proceedings in Theoretical Computer Science, Open Publishing Association.
3. Briggs, R. (2012). Interventionist counterfactuals. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 160, 1, 139–166.
4. Caicedo, X., Dechesne, F., Janssen, T.M.V. (2009). Equivalence and quantifier rules for logic with imperfect information. *Logic Journal of the IGPL*, 17, 91–129.
5. Cameron, P., & Hodges, W. (2001). Some combinatorics of imperfect information. *Journal of Symbolic Logic*, 66, 673–684.
6. Ciardelli, I. (2016). Lifting conditionals to inquisitive semantics. *Semantics and Linguistic Theory*, 26(10), 732.
7. Ciardelli, I., & Roelofsen, F. (2011). Inquisitive logic. *Journal of Philosophical Logic*, 40(1), 55–94.
8. Ciardelli, I., Zhang, L., Champollion, L. (2018). Two switches in the theory of counterfactuals. *Linguistics and Philosophy*, 41(6), 577–621.

9. Corander, J., Hyttinen, A., Kontinen, J., Pensar, J., Väänänen, J. (2016). A logical approach to context-specific independence. In *Logic, Language, Information, and Computation - 23rd International Workshop, WoLLIC 2016, Puebla, Mexico, August 16-19th, 2016. Proceedings* (pp. 165–182).
10. Durand, A., Hannula, M., Kontinen, J., Meier, A., Virtema, J. (2016). Approximation and dependence via multiteam semantics. In *Proceedings of the 9th International Symposium on Foundations of Information and Knowledge Systems*, (Vol. LNCS 9616 pp. 271–291): Springer.
11. Galles, D., & Pearl, J. (1998). An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3(1), 151–182.
12. Galliani, P. (2012). Inclusion and exclusion dependencies in team semantics - on some logics of imperfect information. *Annals of Pure and Applied Logic*, 163(1), 68–84.
13. Galliani, P. (2013). Upwards closed dependencies in team semantics. In *Proceedings Fourth International Symposium on Games, Automata, Logics and Formal Verification, GandALF 2013, Borca di Cadore, Dolomites, Italy, 29-31th August 2013* (pp. 93–106).
14. Grädel, E., & Väänänen, J. (2013). Dependence and independence. *Studia Logica*, 101, 399–410.
15. Halpern, J. (2016). *Actual Causality*. MIT Press.
16. Halpern, J.Y. (2000). Axiomatizing causal reasoning. *J. Artif. Int. Res.*, 12(1), 317–337.
17. Halpern, J.Y., & Pearl, J. (2005). Causes and explanations: a structural-model approach. part i: Causes. *The British Journal for the Philosophy of Science*, 56(4), 843–887.
18. Hausman, D.M. (1998). *Causal asymmetries*. Cambridge University Press.
19. Hintikka, J., & Sandu, G. (1989). Informational independence as a semantical phenomenon. In Fensstad, J.E., et al. (Eds.) *Logic, Methodology and Philosophy of Science VIII* (pp. 571–589): Elsevier Science Publishers B.V.
20. Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *The Journal of Philosophy*, 98(06), 273–299.
21. Hitchcock, C. (2001). A tale of two effects. *Philosophical Review*, 110(3), 361–396.
22. Hitchcock, C. Causal modelling. In Beebe, H., Hitchcock, C., Menzies, P. (Eds.) *The Oxford Handbook of Causation* (p. 2009): Oxford University Press.
23. Hodges, W. (1997). Compositional semantics for a language of imperfect information. *Logic Journal of the IGPL*, 5, 539–563.
24. Hodges, W. (1997). Some strange quantifiers. In Mycielski, J., Rozenberg, G., Salomaa, A. (Eds.) *Structures in Logic and Computer Science. Lecture Notes in Computer sci.*, (Vol. 1261 pp. 51–65). London: Springer.
25. Hyttinen, T., Paolini, G., Väänänen, J. (2015). Quantum team logic and Bell's inequalities. *Review of Symbolic Logic*, 08(04), 722–742.
26. Kaufmann, S. (2013). Causal premise semantics. *Cognitive Science*, 37(6), 1136–1170.
27. Kontinen, J., & Nurmi, V. (2009). Team logic and second-order logic. In Ono, H., Kanazawa, M., de Queiroz, R. (Eds.) *Logic, Language, Information and Computation: Lecture Notes in Computer Science 5514* (pp. 230–241). Berlin: Springer.
28. Lewis, D. (1973). Causation. *Journal of Philosophy*, 70(17), 556–567.
29. Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell publishers.
30. Mann, A.L., Sandu, G., Sevenster, M. (2011). Independence-Friendly logic - a game-theoretic approach, vol. 386 of London Mathematical Society lecture note series. Cambridge University Press.
31. McGee, V. (1985). A counterexample to modus ponens. *Journal of Philosophy*, 82(9), 462–471.
32. Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.
33. Pearl, J., Glymour, M., Jewell, N.P. (2016). *Causal inference in statistics: a primer*. John Wiley & Sons.
34. Sano, K., & Virtema, J. (2015). Axiomatizing propositional dependence logics. In Kreutzer, S. (Ed.) *24th EACSL Annual Conference on Computer Science Logic. Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik*, pp. 292–307. Volume: 41 Proceeding volume.
35. Schulz, K.K. (2011). “If you’d wiggled A, then B would’ve changed”. Causality and counterfactual conditionals *Synthese* 179 (03).
36. Schulz, K. (2018). The similarity approach strikes back: negation in counterfactuals. In Sauerland, U., & Solt, S. (Eds.) *Proceedings of Sinn und Bedeutung 22* (pp. 343–360).
37. Sider, T. (2010). *Logic for Philosophy* Vol. 304. Oxford: Oxford University Press.

38. Spirtes, P., Glymour, C., Scheines, R.N. (1993). *Causation Prediction, and Search*, vol. 81 of *Lecture Notes in Statistics*. New York: Springer.
39. Stalnaker, R.C. (1968). A theory of conditionals. *American Philosophical Quarterly*, pp 98–112.
40. Väänänen, J. (2007). Dependence logic: a new approach to Independence Friendly logic, vol. 70 of *London Mathematical Society Student Texts*. Cambridge University Press.
41. Väänänen, J. (2008). Modal dependence logic, new perspectives on games and interaction. In Apt, K., & van Rooij, R. (Eds.) *Vol. 5 of texts in logic and games* (pp. 237–254): Amsterdam University Press.
42. Veltman, F. (2005). Making counterfactual assumptions. *Journal of Semantics* 22, 2(05), 159–180.
43. von Wright, G.H. (1971). *Explanation and Understanding*. Ithaca: Cornell University Press.
44. Woodward, J. (2001). Probabilistic causality, direct causes and counterfactual dependence. *Stochastic Causality*, pp 39–63.
45. Woodward, J. (2003). Making things happen, vol. 114 of *Oxford Studies in the Philosophy of Science*. Oxford University Press.
46. Yang, F. (2017). Uniform definability in propositional dependence logic. *The Review of Symbolic Logic*, 10(1), 65–79.
47. Yang, F., & Väänänen, J. (2016). Propositional logics of dependence. *Annals of Pure and Applied Logic*, 167(7), 557–589.
48. Yang, F., & Väänänen, J. (2017). Propositional team logics. *Annals of Pure and Applied Logic*, 168(7), 1406–1441.
49. Zhang, J. (2013). A Lewisian logic of causal counterfactuals. *Minds and Machines*, 23(1), 77–93.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.