Disjunctive Antecedents for Causal Models

Mario Günther

LMU Munich
mario.guenther@campus.lmu.de

Abstract

Sartorio [4] argues convincingly that disjunctive causes exist. To treat disjunctive causes within Halpern and Pearl [2]'s framework of causal models, we extend their causal model semantics by disjunctive antecedents and propose a refinement of their definition of actual causation.

1 Introduction

Halpern and Pearl [2] define actual causation based on a causal model semantics of conditionals. The semantics is restricted to antecedents that do not contain disjunctions. "We might consider generalizing further to allow disjunctive causes", so Halpern and Pearl [2, p. 853], but they discard the idea, because there be "no truly disjunctive causes once all the relevant facts are known".

In contrast, Sartorio [4] argues for the existence of disjunctive causes by putting forward a switching scenario, in which all the relevant facts are known. Sartorio's Switch provides motivation to extend Halpern and Pearl [2]'s causal model semantics and definition of actual causation to be applicable to causes that have a particular disjunctive form. Accordingly, we lift the restriction of causal models to non-disjunctive antecedents such that we can express arbitrary Boolean combinations in a conditional's antecedent.

In Section 2, we translate Sartorio's Switch in a causal model. En passant we introduce Halpern and Pearl [2]'s causal model semantics and definition of actual causation. In Section 3, we extend Halpern and Pearl [2]'s causal model semantics by antecedents having a disjunctive form. This allows us to refine Halpern and Pearl's definition of actual causation such that it captures disjunctive causes of the type found in Sartorio's Switch.

2 Sartorio's Switch and Causal Models

Sartorio [4] argues for the existence of disjunctive causes. She invokes roughly the following scenario to back up her claim.

Example 1. **Sartorio's Switch** (Sartorio [4, p. 523–528])
Suppose a train is running on a track onto which a person is tied. Although there is a switch determining on which of two tracks the train continues, the tracks reconverge before the place, where the person is captivated. Now, Sartorio adds details to this typical switching scenario. A person, called Flipper, flips the switch such that the train continues on the left track. Moreover, there is construction work carried out on the right track. Another person, called Reconnecter, reconnects the right track before the train would have arrived in case Flipper hadn't flipped the switch. The train travels on the left track and kills the trapped person.

Sartorio proposes that the disjunction 'Flipper flips the switch and/or Reconnecter reconnects' is the actual cause of the person's death, while both individually 'Flipper flips the switch' and 'Reconnecter reconnects' are not actual causes of the person's death.

In her judgment, she complies with Lewis [3]'s simple counterfactual analysis of actual causation. 'Flipper flips the switch' (and 'Reconnecter reconnects') is not an actual cause of the person's death. For, if it were not the case that 'Flipper flips the switch' (or 'Reconnecter reconnects' respectively), the person would die nevertheless. Additionally, the conjunction 'Flipper flips the switch and Reconnecter reconnects' is no actual cause of the person's death. For, if it were not the case, the person might die nevertheless, viz. in case one of Flipper and Reconnecter does what they do. However, the disjunction 'Flipper flips the switch or Reconnecter reconnects' is an actual cause of the person's death. For, if it were not the case, the person would not die. Sartorio [4, p. 530] confirms that "the death happened because *at least one of them* did what they did."

Sartorio [4] makes the intuition strong that Flipper's redirection is not a cause given that there was an alternative route, even if that route is never actualized. She thinks that "the mere fact that there was an alternative route is sufficient to rob the event of the redirection of its causal powers." (p. 532) Accordingly, Flipper's redirection to the left track renders Reconnecters reconnection of the right track causally inefficacious, and, conversely, the reconnection renders the redirection causally inefficacious. The core of her reasoning goes as follows: "If either event had happened without the other, then that event would have been causally efficacious [...]. But, when both events happen, they deprive each other of causal efficacy." (p. 531) However, so argues Sartorio, the outcome must still depend on the existence of some viable causally efficacious path. Hence, the disjunctive fact that at least one path was causally efficacious is the cause of the outcome.

We translate now Sartorio's Switch in a causal model and check which formulas qualify as actual causes according to Halpern and Pearl [2]'s definition of actual causation.

2.1 Halpern and Pearl's Causal Model Semantics

Halpern and Pearl [2, pp. 851-852]'s causal model semantics of conditionals is defined with respect to a causal model over a signature.

Definition 1. Signature
A signature $\mathcal{S}$ is a triple $\mathcal{S} = \langle \mathcal{U}, \mathcal{V}, \mathcal{R} \rangle$, where $\mathcal{U}$ is a finite set of exogenous variables, $\mathcal{V}$ is a finite set of endogenous variables, and $\mathcal{R}$ maps any variable $Y \in \mathcal{U} \cup \mathcal{V}$ on a non-empty (but finite) set $\mathcal{R}(Y)$ of possible values for Y.

Definition 2. Causal Model
A causal model over signature $\mathcal{S}$ is a tuple $M = \langle \mathcal{S}, \mathcal{F} \rangle$, where $\mathcal{F}$ maps each endogenous variable $X \in \mathcal{V}$ on a function $F_X : (\times_{U \in \mathcal{U}} \mathcal{R}(U)) \times (\times_{Y \in \mathcal{V} \setminus \{X\}} \mathcal{R}(Y)) \mapsto \mathcal{R}(X)$.

The mapping $\mathcal{F}$ defines a set of (modifiable) structural equations modeling the causal influence of exogenous and endogenous variables on other endogenous variables. The function F_X determines the value of $X \in \mathcal{V}$ given the values of all the other variables in $\mathcal{U} \cup \mathcal{V}$. Note that $\mathcal{F}$ defines no structural equation for any exogenous variable $U \in \mathcal{U}$.

Intuitively, a simple conditional $[Y = y]X = x$ is true in a causal model M given context $\vec{u} = u_1, ..., u_n$, if the intervention setting $Y = y$ results in the solution $X = x$ for the structural equations.[1] Such an intervention induces a submodel $M_{Y=y}$ of M.

[1] The solution is unique, because we consider only recursive causal models. We write $\vec{X}$ for a (finite) vector of variables $X_1, .., X_n$, and $\vec{x}$ for a (finite) vector of values $x_1, .., x_n$ of the variables. Hence, we abbreviate $X_1 = x_1, .., X_n = x_n$ by $\vec{X} = \vec{x}$. For simplicity, we do not properly distinguish between the vector and its set $\{\vec{X} = \vec{x}\}$.

Definition 3. Submodel
Let $M = \langle \mathcal{S}, \mathcal{F} \rangle$ be a causal model, $\vec{X}$ a (possibly empty) vector of variables in $\mathcal{V}$ and $\vec{x}, \vec{u}$ vectors of values for the variables in $\vec{X}, \vec{U}$. We call the causal model $M_{\vec{X}=\vec{x}} = \langle \mathcal{S}_{\vec{X}}, \mathcal{F}^{\vec{X}=\vec{x}} \rangle$ over signature $\mathcal{S}_{\vec{X}} = \langle \mathcal{U}, \mathcal{V} \setminus \vec{X}, \mathcal{R}|_{\mathcal{V} \setminus \vec{X}} \rangle$ a submodel of M. $\mathcal{F}^{\vec{X}=\vec{x}}$ maps each variable in $\mathcal{V} \setminus \vec{X}$ on a function $F_Y^{\vec{X}=\vec{x}}$ that corresponds to F_Y for the variables in $\mathcal{V} \setminus \vec{X}$ and sets the variables in $\vec{X}$ to $\vec{x}$.

We can describe the structure of Sartorio's Switch using a causal model including four binary variables:

- an exogenous variable T, where $T = 1$ if the train arrives and $T = 0$ otherwise;
- an endogenous variable F, where $F = 1$ if Flipper flips the switch and $F = 0$ otherwise;
- an endogenous variable R, where $R = 1$ if Reconnecter reconnects and $R = 0$ otherwise;
- an endogenous variable D, where $D = 1$ if the person dies and $D = 0$ otherwise.

Leaving the functions F_F, F_R and F_D implicit, the set of structural equations is given by:

- $F = T$
- $R = T$
- $D = max(F, R)$

In words, Flipper flips the switch ($F = 1$), if the train arrives ($T = 1$). Reconnecter reconnects ($R = 1$), if the train arrives. The person dies ($D = 1$), if at least one of $F = 1$ and $R = 1$ is the case. These recursive dependencies of the structural equations are depicted in Figure 1.

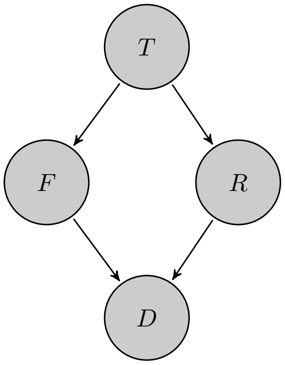

Figure 1: The causal network for Sartorio's Switch. The arrows represent the dependences of the structural equations.

To illustrate the causal model semantics, let us check whether or not the conditional $[F = 1]D = 1$ is true in the causal model M of Sartorio's Switch (given the context $t = 1$). Intuitively, the intervention that sets $F = 1$ induces a submodel $M_{F=1}$ of M. If the solution to the structural equations of $M_{F=1}$ satisfies $D = 1$, then $[F = 1]D = 1$ is true in the causal model M under context $T = t$. In this case, we write $\langle M, t\rangle \models [F = 1]D = 1$.

In the scenario of Sartorio's Switch, $\langle M, t\rangle \models [F = 1]D = 1$ iff $\langle M_{F=1}, t\rangle \models D = 1$. The structural equations for the submodel $M_{F=1}$ are:

- $F = 1$
- $R = T$
- $D = max(F, R)$

We see that the solution to the structural equations of $M_{F=1}$ satisfies $D = 1$, and thus M satisfies the conditional $[F = 1]D = 1$ (given t). Notice the difference between the structural equation $F = T$ and $F = 1$: the former depends on T, whereas the latter does not. After the intervention that sets $F = 1$, the variable F is treated similar to an exogenous variable, i. e. it is assigned a value by its structural equation that does not depend on other (exogenous and/or endogenous parent) variables.[2] The structural equations for the variables in $\mathcal{V} \setminus \{F\}$ remain unchanged.

2.2 Halpern and Pearl's Definition of Actual Causation

The basic idea behind Halpern and Pearl [2]'s definition is to extend Lewis's notion of causal dependence to a notion of contingent dependence. Lewis [3, p. 563] defines causal dependence between two occurring events C and E in terms of counterfactual dependence. E causally depends on C iff (i) C and E occurred, and (ii) the simple counterfactual criterion is satisfied: if C had not happened, E would not have happened. Furthermore, he identifies actual causation with the transitive closure of causal dependence. Hence, C is an actual cause of E iff there is a chain of causal dependencies from C to E. Halpern and Pearl extend this definition by (possibly non-actual) contingencies: C is an actual cause of E iff E causally depends on C *under certain contingencies*. Roughly, contingent dependence makes it possible that even if E does not counterfactually depend on C in the actual situation, E counterfactually depends on C under certain contingencies.[3]

Based on their causal model semantics for conditionals, Halpern and Pearl [2, p. 853] propose the following definition of actual causation.

Definition 4. Actual Causation
$\vec{X} = \vec{x}$ is an actual cause of ϕ in $\langle M, \vec{u}\rangle$ iff the following three conditions hold:

AC1. $\langle M, \vec{u}\rangle \models (\vec{X} = \vec{x}) \wedge \phi$.

AC2. There exists a partition $\langle \vec{Z}, \vec{W}\rangle$ of $\mathcal{V}$ with $\vec{X} \subseteq \vec{Z}$ and some setting $\langle \vec{x'}, \vec{w'}\rangle$ of the variables in $\langle \vec{X}, \vec{W}\rangle$ such that if $\langle M, \vec{u}\rangle \models Z = z^*$ for all $Z \in \vec{Z}$, then both of the following conditions hold:

(a) $\langle M, \vec{u}\rangle \models [\vec{X} = \vec{x'}, \vec{W} = \vec{w'}]\neg\phi$.

(b) $\langle M, \vec{u}\rangle \models [\vec{X} = \vec{x}, \vec{W'} = \vec{w'}, \vec{Z'} = \vec{z^*}]\phi$ for all subsets $\vec{W'}$ of $\vec{W}$ and all subsets $\vec{Z'}$ of $\vec{Z}$.

[2] Intuitively, we may think of a value assignment $X = x$ in model M by an intervention as overruling the structural equation in M.

[3] Note that Halpern and Pearl do not take the transitive clossure for their definition of actual causation. In contrast to Lewis's dictum, they think [2, p. 844] that causation is not always transitive.

AC3. $\vec{X}$ is minimal; no subset of $\vec{X}$ satisfies conditions AC1 and AC2.

AC1 requires both that the actual cause $\vec{X} = \vec{X}$ and its effect ϕ are true in the actual (contextualized) model. AC3 ensures that only the conjuncts of $\vec{X} = \vec{x}$ "essential" for changing ϕ in AC2(a) are part of a cause: "inessential elements are pruned." (Halpern and Pearl [2, p. 853]) As proven by Eiter and Lukasiewicz [1], AC3 implies that an actual cause is always a single conjunct of the form $X = x$, if the set of endogenous variables is finite.

To understand AC2, it is helpful to think of $\vec{X} = \vec{x}$ as the minimal set of conjuncts that qualifies as a cause of the effect ϕ, and to think of $\vec{Z} = \vec{z}$ as the active causal path(s) from $\vec{X}$ to ϕ.

AC2(a) is reminiscent of Lewis [3]'s simple counterfactual criterion: ϕ would be false, if it were not for $\vec{X} = \vec{x}$. The condition says that there is a setting $\vec{X} = \vec{x}$ changing ϕ to $\neg\phi$, if the variables not on the active causal path(s) take on certain values, i. e. $\vec{W} = \vec{w}'$. The difference to the counterfactual criterion is that ϕ's dependence on $\vec{X} = \vec{x}$ may be tested under certain contingencies $\vec{W} = \vec{w}'$, which are non-actual for $\vec{w}' \neq \vec{w}$. Note that those contingent tests allow to identify more causal relationships than the simple counterfactual criterion.

AC2(b) restricts the contingencies allowed to be considered. The idea is that any considered contingency does not affect the active causal path(s) with respect to $\vec{X} = \vec{x}$ and ϕ. In other words, AC2(b) guarantees that $\vec{X}$ alone is sufficient to change ϕ to $\neg\phi$. The setting of a contingency $\vec{W} = \vec{w}'$ only eliminates spurious side effects that may hide $\vec{X}'s$ effect. The idea behind AC2(b) is implemented as follows: (i) setting a contingency $\vec{W} = \vec{w}'$ leaves the causal path(s) unaffected by the condition that changing the values of any subset $\vec{W}'$ of $\vec{W}$ from the actual values $\vec{w}$ to the contingent values $\vec{w}'$ has no effect on ϕ's value. (ii) At the same time, changing the values of $\vec{W}'$ may alter the values of the variables in $\vec{Z}$, but this alteration has no effect on ϕ's value.

We apply now Halpern and Pearl [2]'s definition of actual causation to the causal model of Sartorio's Switch. The result is that each of $F = 1$ and $R = 1$ is an actual cause of $D = 1$. However, the conjunction $F = 1 \wedge R = 1$ and the disjunction $F = 1 \vee R = 1$ do not qualify as actual causes of $D = 1$.

We show that $F = 1$ is an actual cause of $D = 1$. (The argument for $R = 1$ is structurally the same as the causal model of Sartorio's Switch is symmetric with respect to F and R.) Let $\vec{Z} = \{F, D\}$, and so $\vec{W} = \{R\}$. The contingency $R = 0$ satisfies the two conditions of AC2: AC2(a) is satisfied, as setting $F = 0$ results in $D = 0$; AC2(b) is satisfied, as setting F back to 1 results in $D = 1$. The counterfactual contingency $R = 0$ is required to reveal the hidden dependence of D on F, or so argue Halpern and Pearl.

We show that $F = 1 \wedge R = 1$ is not an actual cause of $D = 1$ due to the minimality condition AC3. Let $\vec{Z} = \{F, R, D\}$, and so $\vec{W} = \emptyset$. AC2(a) is satisfied, as setting $F = 0 \wedge R = 0$ results in $D = 0$. AC2(b) is satisfied trivially. However, two subsets of $\vec{X} = \{F, R\}$ satisfy the two conditions of AC2 as well, viz. $\vec{X}' = \{F\}$ and $\vec{X} = \{R\}$. Therefore, $\vec{X} = \{F, R\}$ is not minimal and according to AC3 the conjunction $F = 1 \wedge R = 1$ is thus no actual cause of $D = 1$. Minimality is meant to strip "overspecific details from the cause." (Halpern and Pearl [2, p. 857])

The disjunction $F = 1 \vee R = 1$ does not qualify as actual cause of $D = 1$, simply because Halpern and Pearl [2]'s definition of actual causation does not admit causes in form of proper disjunctions, i. e. disjunctions having more than one disjunct. They do not "have a strong intuition as to the best way to deal with disjunction in the context of causality and believe that disallowing it is reasonably consistent with intuitions." (p. 858)

Sartorio [4, p. 530] observes that "there is no general motivation for believing that, when (if) a disjunctive fact is a cause, at least one of its disjuncts must also be a cause." This observation stands in sharp contrast to Halpern and Pearl [2]'s definition of actual causation, according to which both disjuncts individually qualify as actual causes. In the next section, we first define disjunctive antecedents for Halpern and Pearl's causal model semantics; subsequently, we extend their definition of actual causation to cover disjunctive causes as found in Sartorio's Switch.

3 An Extension of Causal Model Semantics by Disjunctive Antecedents

Recall Sartorio's Switch of Section 2. Sartorio argues that the person tied to the tracks dies because at least one of Flipper and Reconnecter does what they do. Therefore, the disjunctive fact that at least one track or path was causally efficacious is the cause of the outcome. Moreover, if only one of Flipper's and Reconnecter's events would occur, their *disjunction* would be causally inefficacious, but the single occuring event would be causally efficacious. We identify here two necessary conditions under which there are disjunctive causes: (i) there are more than one potentially efficacious and actually occuring events on different paths ("two tracks"), and (ii) there is an event that switches the paths without being, intuitively, a cause of the outcome ("flipping the switch").

Let us consider Sartorio's Switch using the variables of our causal model. In her switching scenario, Sartorio maintains that $F = 1 \vee R = 1$ is an actual cause of $D = 1$. The disjunction means that $D = 1$ because at least one of $F = 1$ and $R = 1$. On closer inspection, using our identified necessary conditions for disjunctive causes, Sartorio's disjunction means: the actual case $F = 1$ and $R = 1$ results in $D = 1$ *and* the counterfactual case $F = 1$ and $R = 0$ results in $D = 1$ *and* the counterfactual case $F = 0$ and $R = 1$ results in $D = 1$. There are two reasons: (a) if $F = 1$ alone were not sufficient to result in $D = 1$, the disjunction $F = 1 \vee R = 1$ would not be the actual cause. (Mutatis mutandis for $R = 1$.) (b) Both of $F = 1$ and $R = 1$ need actually to be the case. In such a case, if one *or* the other is sufficient for the effect and both occur, then Sartorio judges the disjunction of both to be the cause. In this sense, Sartorio understands the disjunction $F = 1 \vee R = 1$ as a summary of two actually occuring events $F = 1$ and $R = 1$, whose actual co-occurrence robs them of their individual causal efficacy, and which would, individually, be actual causes.

Halpern and Pearl [2]'s causal model semantics does not allow to evaluate the conditional $[F = 1 \vee R = 1]D = 1$. The reason is that they do not allow for disjunctions in the antecedent, and so the submodel $M_{F=1 \vee R=1}$ is undefined. Moreover, the structural equation for D of Sartorio's Switch does apply to values of F and R, but it does not apply to a disjunction such as $F = 1 \vee R = 1$. Hence, the value for D is not determined by the disjunction. Next, we propose a conservative extension of Halpern and Pearl's causal model semantics that allows us to evaluate antecedents that are disjunctive in Sartorio's sense.

3.1 Evaluating Disjunctive Antecedents

As we have just observed, Sartorio's disjunctive causes of the form $A = a \vee B = b$ require that $A = a \wedge B = b$ actually obtain, and if one of $A = a$ or $B = b$ would obtain but not the other, the effect would still follow. We implement now this logic governing Sartorio's disjunctive causes by extending Halpern and Pearl [2]'s framework of causal models.

The idea behind evaluating a conditional with disjunctive antecedent is to check whether the consequent is true in *each* disjunctive situation of the antecedent. We say that a Sartorio disjunction $A = a \vee B = b$ is satisfied if three possible situations are satisfied: (i) $A = a \wedge B = b$, (ii) $A = a \wedge B = \neg b$, and (iii) $A = \neg a \wedge B = b$. We refer to (i)-(iii) as the disjunctive situations or possibilities of the formula $A = a \wedge B = b$. Intuitively, each disjunctive situation corresponds to one intervention that sets the values for a non-disjunctive formula. The result is one submodel per disjunctive situation. The antecedent $[A = a \vee B = b]$, for example, does not correspond to a unique intervention, but rather to three interventions. Each of the interventions results in exactly one submodel. The intervention (i), for instance, results in the submodel $M_{A=a,B=b}$, in which A and B take the same values than in the actual contextualized model $\langle M, \vec{u} \rangle$ given $A = a \vee B = b$ is an actual disjunctive cause in Sartorio's sense.

To evaluate a conditional with disjunctive antecedent does – according to the outlined idea – not require to modify Halpern and Pearl [2, pp. 849–852]'s notion of a submodel. Rather, the evaluation requires to look at (possibly) more than one submodel, namely at exactly one submodel for each disjunctive situation. In general, we write ϕ_i, where $(1 \leq i \leq n)$, for the formula that expresses the i-th disjunctive situation of the formula ϕ (that contains only finitely many primitive events).[4]

For clarity, we define an extended causal language.

Definition 5. Extended Causal Language $\mathcal{L}$
The extended causal language $\mathcal{L}$ conains

- the two propositional constants $\top$ and $\bot$,
- a finite number of random variables $\vec{X} = X_1, ..., X_n$ associated with finite ranges $\mathcal{R}(X_1), ..., \mathcal{R}(X_n)$,
- the Boolean connectives $\wedge, \vee, \neg$ and the operator $[]$, and
- left and right parentheses.

A formula ϕ of $\mathcal{L}$ is well-formed iff ϕ has the form

- $X = x$ for $x \in \mathcal{R}(X)$ (primitive event);
- if $\phi, \psi \in \mathcal{L}$, then $\neg\phi, \phi \wedge \psi, \phi \vee \psi \in \mathcal{L}$ (Boolean combinations of primitive events);
- if $[]$ does not occur in $\phi, \psi \in \mathcal{L}$, then $[\phi]\psi \in \mathcal{L}$ (causal conditionals).

For the extended causal language, we define a valuation function. Recall that $\langle M, \vec{u} \rangle \models X = x$ is shorthand for $X = x$ is the solution to all of the structural equations in the recursive model M given context $\vec{u}$.

Definition 6. Valuation Function
A valuation function $v_{\langle M, \vec{u} \rangle}$ (abbreviated as v) is associated with any arbitrary model M and any arbitrary vector $\vec{u}$. $v_{\langle M, \vec{u} \rangle} : \mathcal{L} \mapsto \{1, 0\}$ assigns either 1 or 0 to all formulas of the extended causal language $\mathcal{L}$:

$$\text{(a)} \;\; v(X = x) = \begin{cases} 1, & \text{if } \langle M, \vec{u} \rangle \models X = x \\ 0, & \text{otherwise} \end{cases}$$

[4] Note that the number i of ϕ_i depends on the number d of disjunctions occurring in ϕ. In general, $i \leq 2^{d+1} - 1$, i. e. there are at most $2^{d+1} - 1$ disjunctive situations of ϕ. When the disjuncts are mutually exclusive, there are less disjunctive situations, because some are impossible. Take for example $F = 1 \vee F = 0$ for the binary variable F. Here, there are only two disjunctive situations, because $F = 1 \wedge F = 0$ is impossible. For, if $v(F = 1) = 1$ then $v(F = 0) = 0$ and if $v(F = 0) = 1$ then $v(F = 1) = 0$.

(b) $v(\neg\phi) = 1$ iff $v(\phi) = 0$

(c) $v(\phi \wedge \psi) = 1$ iff $v(\phi) = 1$ and $v(\psi) = 1$

(d) $v(\phi \vee \psi) = 1$ iff $v(\phi) = 1$ or $v(\psi) = 1$

(e) $v([\phi]\psi) = \begin{cases} 1, & \text{if } v(\psi) = 1 \text{ in each } \langle M_{\phi_i}, \vec{u}\rangle \\ 0, & \text{otherwise} \end{cases}$

, where M_{ϕ_i} is a submodel of M such that $\langle M, \vec{u}\rangle \models [\phi_i]\psi$, and ϕ_i is a non-disjunctive formula expressing one disjunctive possibility of ϕ.

Clause (c) of the valuation function entails that $X_1 = x_1, .., X_n = x_n$ is the setting of the variables in the contextualized model $\langle M, \vec{u}\rangle$ iff $X_1 = x_1 \wedge ... \wedge X_n = x_n$ is true in $\langle M, \vec{u}\rangle$. Hence, a vector of primitive events $\vec{X} = \vec{x}$ corresponds to a conjunction of those primitive events $\bigwedge X_i = x_i$ for $1 \leq i \leq n$.

Let us now evaluate a conditional with disjunctive antecedent in the causal model of Sartorio's Switch. We check whether or not $\langle M, t = 1\rangle \models [F = 1 \vee R = 1]FB = 1$. Let ϕ_1, ϕ_2, ϕ_3 express the disjunctive situations of $F = 1 \vee R = 1$. According to clause (e), we need to check whether $v(D = 1) = 1$ in each $\langle M_{\phi_i}, t = 1\rangle$ for $i = 3$. Figure 2 depicts the causal network of the submodel M_{ϕ_1} for the disjunctive situation ϕ_1. M_{ϕ_2} and M_{ϕ_3} look the same for $\phi_2 = (F = 1) \wedge (R = 0)$ and $\phi_3 = (F = 0) \wedge (R = 1)$.

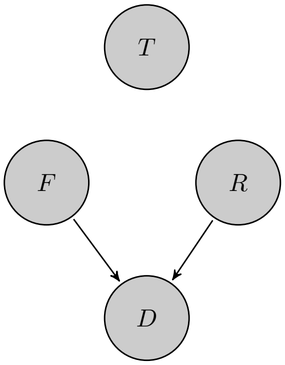

Figure 2: The causal network of M_{ϕ_1} for $\phi_1 = (F = 1) \wedge (R = 1)$.

As $D = max\{F, R\}$ remains unchanged in each $\langle M_{\phi_i}, t = 1\rangle$, we obtain for the three submodels:

(i) $\langle M_{\phi_1}, t = 1\rangle \models D = 1$

(ii) $\langle M_{\phi_2}, t = 1\rangle \models D = 1$

(iii) $\langle M_{\phi_3}, t = 1\rangle \models D = 1$

Hence, $v(D = 1) = 1$ in each $\langle M_{\phi_i}, t = 1\rangle$, and thus the model M satisfies the conditional $[F = 1 \vee R = 1]D = 1$ in context $t = 1$.

3.2 A Refinement of Halpern and Pearl's Definition of Actual Causation

Now that we can evaluate disjunctive antecedents in extended causal models, we propose a refinement or amendment of Halpern and Pearl [2]'s definition of actual causation.

Let $\psi_{\vee_i} = (\vec{X}_i = \vec{x}_i)$ denote the i-th disjunct of the arbitrary Boolean combination ψ of finitely many primitive events.

Definition 7. Actual Causation Refined
ψ is an actual cause of ϕ in $\langle M, \vec{u} \rangle$ iff the following three conditions hold:

AC1R. $\langle M, \vec{u} \rangle \models (\bigwedge \psi_{\vee_i}) \wedge \phi$ for all i.

AC2R. There exists a partition $\langle \vec{Z}, \vec{W} \rangle$ of $\mathcal{V}$ with $\vec{X}_i \subseteq \vec{Z}$ and some setting $\langle \vec{x'}_i, \vec{w'} \rangle$ of the variables in $\langle \vec{X}_i, \vec{W} \rangle$ such that if $\langle M, \vec{u} \rangle \models Z = z^*$ for all $Z \in \vec{Z}$, then both of the following conditions hold:

(a) $\langle M, \vec{u} \rangle \models [\bigwedge \vec{X}_i = \vec{x'}_i, \vec{W} = \vec{w'}]\neg\phi$ for all i.

(b) $\langle M, \vec{u} \rangle \models [\bigvee \vec{X}_i = \vec{x}_i, \vec{W}' = \vec{w'}, \vec{Z}' = \vec{z^*}]\phi$ for all subsets $\vec{W}'$ of $\vec{W}$ and all subsets $\vec{Z}'$ of $\vec{Z}$, and for all i.

AC3R. ψ is minimal; no subsets of the disjuncts $\psi_{\vee_i} = (\vec{X}_i = \vec{x}_i)$ satisfy conditions AC1R and AC2R, and no disjunction of the form $\bigvee(\vec{X}_i = \vec{x}_i) \vee \vec{Y} = \vec{y}$ with $\vec{Y} \subseteq \vec{Z}$, $\vec{Y} \cap \vec{X}_i = \emptyset$ (for all i) and $\vec{Y} \neq \phi$ satisfies AC1R and AC2R.

AC1R requires that each disjunct of the actual cause ψ and its effect ϕ are true in the actual contextualized model. Note that this is equivalent to the big *conjunction* of all disjuncts $\phi_{\vee_i}$ and the effect ϕ being true in the actual contextualized model. (The need for the big conjunction directly follows from Sartorio's first condition necessary for disjunctive causes.)

AC2R requires that each disjunct $\psi_{\vee_i} = (\vec{X}_i = \vec{x}_i)$ of ψ satisfies AC2. That is: (a) setting $\vec{X}_i = \vec{x}_i$ (for any i) changes ϕ to $\neg\phi$, if the variables $\vec{W}$ not on the active causal path(s) take on certain values; (b) guarantees that the disjunction $\bigvee(\vec{X}_i = \vec{x}_i)$ alone is sufficient to change ϕ to $\neg\phi$. Note that AC2R(b) is quite demanding: setting $\bigvee(\vec{X}_i = \vec{x}_i)$ results in a submodel for each disjunctive situation of ψ, and under all of these submodels ϕ is satisfied.

AC3R extends the motivation behind AC3, which is to "prune inessential elements" from the actual causes. The extension demands that if we have another actually occurring disjunct that would alone be sufficient to result in the effect, we need to add it to the disjunctive cause. Correspondingly, we obtain that a formula of the form $(\vec{X} = \vec{x}) \wedge (\vec{Y} = \vec{y})$ for $\vec{X} \cap \vec{Y} = \emptyset$ is more specific and less minimal than $\vec{X} = \vec{x}$, which is in turn more specific and less minimal than $(\vec{X} = \vec{x}) \vee (\vec{Y} = \vec{y})$. Assume this disjunction is an actual cause of some effect. Then the disjunction strips the "overspecific detail" which specific disjunct is causally efficacious (both are!) from the actual cause.

We show now that in $\langle M, t = 1 \rangle$ the disjunction $F = 1 \vee R = 1$ is an actual cause of $D = 1$ according to our refined definition. AC1R is satisfied, as $\langle M, t = 1 \rangle \models (F = 1 \wedge R = 1) \wedge \phi$. AC2R is satisfied as well. To see this, let $\vec{Z} = \{F, R, D\}$, and thus $\vec{W} = \emptyset$. Clearly, $F, R \subseteq \vec{Z}$. But then (a) $\langle M, t = 1 \rangle \models [F = 0 \wedge R = 0]D = 0$. Furthermore, (b) $\langle M, t = 1 \rangle \models [F = 1 \vee R = 1]D = 1$, as we have seen in the previous section. Finally, AC3R is satisfied: no subsets of the disjuncts $F = 1$ and $R = 1$ satisfy AC1R and AC2R; there exists no further disjunct satisfying AC1R and AC2R, as $\vec{Z} \setminus \{F, R\} = \{D\}$ and D is the effect.

According to our refined definition, $F = 1$ does not qualify any more as an actual cause of $D = 1$. (The same holds mutatis mutandis for $R = 1$.) The reason is AC3R: $F = 1$ is not minimal. Why? Because there is a disjunction $F = 1 \vee R = 1$ with $R \subseteq \vec{Z}$, $R \cap F = \emptyset$ and $(R = 1) \neq (D = 1)$ satisfying AC1R and AC2R. Hence, $F = 1$ is "inessential" for $D = 1$ in the sense that it is not required for $D = 1$ to obtain, as the actual event $R = 1$ alone would also be sufficient for $D = 1$ to obtain.

4 Conclusion

We generalized Halpern and Pearl [2]'s causal model semantics to allow disjunctive causes of the type found in Sartorio [4]'s Switch. These disjunctive causes have an actual part, i. e. both disjuncts actually occur, and a counterfactual part, i. e. each disjunct would be sufficient for the effect to occur. Based on the causal model semantics extended by disjunctive antecedents à la Sartorio, we refined Halpern and Pearl's definition of actual causation. Halpern and Pearl's original definition qualifies Flipper's flipping the switch as an actual cause of the captivated person's death and does not allow for disjunctive causes. In contrast, our refined definition disqualifies the individual disjuncts as actual causes but makes Sartorio's disjunction "at least one of Flipper flips the switch and Reconnecter reconnects" an actual cause of the person's death. Our refined definition, therefore, implements Sartorio [4, p. 530]'s observation that "there is no general motivation for believing that, when (if) a disjunctive fact is a cause, at least one of its disjuncts must also be a cause."

References

[1] Eiter, T. and Lukasiewicz, T. (2002). Complexity results for structure-based causality. *Artificial Intelligence* **142**(1): 53 – 89. doi:http://dx.doi.org/10.1016/S0004-3702(02)00271-0. URL `http://www.sciencedirect.com/science/article/pii/S0004370202002710`.

[2] Halpern, J. Y. and Pearl, J. (2005). Causes and Explanations: A Structural-Model Approach. Part I: Causes. *British Journal for the Philosophy of Science* **56**(4): 843–887.

[3] Lewis, D. (1973). Causation. *Journal of Philosophy* **70**(17): 556–567.

[4] Sartorio, C. (2006). Disjunctive Causes. *Journal of Philosophy* **103**(10): 521–538.