The HP Definition of Actual Causality

Jingzhi Fang

Department of Philosophy
Peking University

Logic Seminar, May. 29, 2018

Causality

Causality cause-effect relationship among variables or events.

Type causality (general causality) general statements
E.g., “Smoking causes lung cancer” and “printing money causes inflation”

Actual causality (token causality/specific causality) focus on particular events
E.g., “the fact that David smoked like a chimney for 30 years caused him to get cancer last year”

Type causation arises from many instances of actual causation, so that actual causation is more fundamental.

The HP definition of causality

J.Y. Halpern & J. Pearl

The original HP definition was introduced in

Halpern, J. Y. and J. Pearl (2001). Causes and explanations : A structural-model approach.
Part I : Causes. In
Proc. Seventeenth Conference on Uncertainty in Artificial Intelligence (*UAI* 2001), pp. 194–202 ;

it was updated in

Halpern, J. Y. and J. Pearl (2005a). Causes and explanations : A structural-model approach.
Part I : Causes. *British Journal for Philosophy of Science* 56(4), 843–887. ;

the modified definition was introduced in

Halpern, J. Y. (2015a). A modification of the Halpern-Pearl definition of causality.t
In *Proc. 24th International Joint Conference on Artificial Intelligence* (*IJCAI* 2015), pp. 3022–3033.

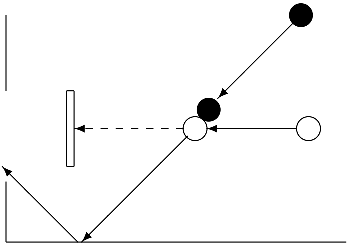
The idea of the definition

> We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed.
>
> —— Hume

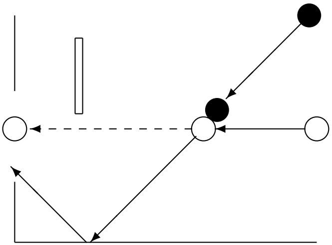
The idea of the definition

- Two quite different notions of causality :
  1. regularity definition :
     consider what actually happens, specifically, which events precede others.
     Roughly speaking, A causes B if As typically precede Bs. (type causality)
  2. counterfactuals : statements counter to fact.

The idea of the definition



(a) Brick would have blocked white ball.



(b) Brick would not have blocked white ball.

Research in psychology has shown that such counterfactual thinking plays a key role in determining causality. People really do consider “what might have been” as well as “what was”.

The idea of the definition

When we say that A caused B, we invariably require (among other things) that A and B both occurred, so when we contemplate A not happening, we are considering a counterfactual.

The idea of the definition

The simple counterfactual definition doesn't always work.

- When it does, we have what's called a *but* $-$ *for* cause
- This is the situation considered most often in the law :
  Andy and Bob both pick up rocks and throw them at a bottle. Andy's rock gets there first, shattering the bottle. Because both throws are perfectly accurate, Bob's would have shattered the bottle had it not been preempted by Andy's throw.

So why is Andy's throw the cause ?

- If Andy hadn't thrown under the contingency that Bob didn't throw, then the bottle would not have shattered.

But then why isn't Billy's throw also a cause ?

- Because it didn't hit the bottle ! (duh . . .)
- Must set appropriate contingencies, which takes into account what actually happened.

Causal models

The model assumes that the world is described in terms of variables, these variables can take on various values.

- If we are trying to determine whether a forest fire was caused by lightning or an arsonist. We can take the world to be described by these variables :
  - *FF* for forest fire, where FF = 1 if there is a forest fire and FF = 0 otherwise ;
  - *L* for lightning, where L = 1 if lightning occurred and L = 0 otherwise ;
  - *MD* for match dropped (by arsonist), where MD = 1 if the arsonist dropped a lit match and MD = 0 otherwise.

Causal models

Some variable may have a causal influence on others. This influence is modeled by **structural equations**.

Split the random variables into

- exogenous variables : values are taken as given, determined by factors outside model
- endogenous variables

Structural equations describe the values of all endogenous variables in terms of exogenous variables and other endogenous variables.

- For the forest-fire example,
  - $U = (i, j)$, where i describes whether the external conditions are such that the lightning strikes and j describes whether the arsonist drops the match, i and j are each either 0 or 1.
  - $L = i$
    $MD = j$
    If we want to model the fact that if the arsonist drops a match or lightning strikes, then a fire starts.
    $FF = max(L, MD)$ $(FF = L \vee MD)$

Causal models

The goal is to give a definition of actual causality in terms of counterfactuals. We need a model that makes it possible to consider the effect of intervening on X and changing its value from x to x'.

- using structural equations makes it easy to define the effect of an intervention
- describing the world in terms of variables and their values makes it easy to describe such interventions

Causal models

- A signature S is a tuple (U, V, R), where U is a set of exogenous variables, V is a set of endogenous variables, and R associates with every variable $Y \in U \cup V$ a nonempty set $R(Y)$ of possible values for Y.
- Context : a setting $\vec{u}$ for the exogenous variables in U

Causal models

Definition 1

- A *causal model* M is a pair (S, F), where S is a *signature*, and F defines a set of structural equations, relating the values of the variables.

- F associates with each endogenous variable $X \in V$ a function denoted F_X such that $F_X : \times_{Z \in (U \cup V - \{X\})} R(Z) \to R(X)$.
- Setting the value of some variable X to x results in a new causal model denoted $M_{X \leftarrow x} = (S, F_{X \leftarrow x})$, where $F_{X \leftarrow x}$ is the result of replacing the equation for X in F by $X = x$ and leaving the remaining equations untouched.

Causal models

- Y **depends on** X if there is some setting of all the variables in $U \cup V$ other than X and Y and the value x and x' of X such that varying x to x' in that setting results in a variation in the value of Y
  (if there is a setting $\vec{z}$ of the variables other than X and Y and values x and x' of X such that $F_Y(x, \vec{z}) \neq F_Y(x', \vec{z})$).
  Is it more reasonable to say that $\forall x \in R(X), \exists x' \in R(X)...$?
- **Recursive models** : if, for each context $\vec{u}$, there is a partial order $\preceq_{\vec{u}}$ of the endogenous variables such that unless $X \preceq_{\vec{u}} Y$, Y is independent of X in $(M, \vec{u})$, where Y is independent of X in $(M, \vec{u})$ if, for all settings $\vec{z}$ of the endogenous variables other than X and Y , and all values x and x' of X, $F_Y(x, \vec{z}, \vec{u}) = F_Y(x', \vec{z}, \vec{u})$.

Language

Definition 2

- Given a signature $S = (U, V, R)$, a *primitive event* is a formula of the form $X = x$, for $X \in V$ and $x \in R(X)$.
  A *causal formula* (*over* S) is one of the form $[Y_1 \leftarrow y_1, ..., Y_k \leftarrow y_k]\phi$, where
  -ϕ is a Boolean combination ($\wedge$, $\vee$, and $\neg$) of primitive events .
  -$Y_1, ..., Y_k$ are distinct variables in V, and
  -$y_i \in R(Y_i)$.
  Such a formula is abbreviated as $[\vec{Y} \leftarrow \vec{y}]\phi$. The special case where k = 0 is abbreviated as $[]\phi$ or, more often, just ϕ. Let $L(S)$ consist of all Boolean combinations of causal formulas.

Semantics

Definition 3

- A *causal setting* is a pair $(M, \vec{u})$ consisting of a causal model M and context $\vec{u}$.

Definition 4

- In a recursive model M, the satisfaction relation $\models$ is defined as follows :
  - $(M, \vec{u}) \models X = x$ iff the value of X is x once we set the exogenous variables to $\vec{u}$.
  - $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\phi'$ iff $(M_{\vec{Y} \leftarrow \vec{y}}, \vec{u}) \models \phi'$ (where ϕ' is a Boolean combination of primitive events)
  - $(M, \vec{u}) \models \neg\phi$ iff $(M, \vec{u}) \not\models \phi$
  - $(M, \vec{u}) \models \phi \wedge \psi$ iff $(M, \vec{u}) \models \phi$ and $(M, \vec{u}) \models \psi$
  - $(M, \vec{u}) \models \phi \vee \psi$ iff $(M, \vec{u}) \models \phi$ or $(M, \vec{u}) \models \psi$

  Here ϕ and ψ are Boolean combinations of causal formulas.

The HP definition of actual cause

- Actual cause : $X_1 = x_1 \wedge ... \wedge X_k = x_k$ ($\vec{X} = \vec{x}$)
- The events that can be caused : arbitrary Boolean combination of primitive events.

The HP definition of actual cause

Definition 5

- $\vec{X} = \vec{x}$ is an *actual cause* of ϕ in the causal setting $(M, \vec{u})$ if the following three conditions hold :
  AC1. $(M, \vec{u})| = (\vec{X} = \vec{x})$ and $(M, \vec{u})| = \phi$.
  AC2. See below.
  AC3. $\vec{X}$ is minimal ; there is no strict subset $\vec{X}'$ of $\vec{X}$ such that $\vec{X}' = \vec{x}'$ satisfies conditions AC1 and AC2, where $\vec{x}'$ is the restriction of $\vec{x}$ to the variables in $\vec{X}'$.

- The original HP definition : AC1+AC2(a)+AC2(b^o)+AC3
  The updated HP definition : AC1+AC2(a)+AC2(b^u)+AC3
  The modified HP definition : AC1+AC2(a^m)+AC3

AC2(a)-a necessity condition

- $X = x$ is a *but − for cause* of ϕ in $(M, \vec{u})$ if AC1 holds and there exists some x' such that $(M, \vec{u}) \models [X \leftarrow x']\neg\phi$.

AC2(a)-a necessity condition

Definition 6

- AC2(a). There is a partition of V into two disjoint subsets $\vec{Z}$ and $\vec{W}$ (so that $\vec{Z} \cap \vec{W} = \emptyset$) with $\vec{X} \subseteq \vec{Z}$ and a setting $\vec{x}'$ and $\vec{w}$ of the variables in $\vec{X}$ and $\vec{W}$, respectively, such that $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}]\neg\phi$.

AC2(b^o)-a sufficiency condition

Definition 7

- AC2(b^o). If $\vec{z}^*$ is such that $(M,\vec{u}) \models \vec{Z} = \vec{z}^*$, then for all subsets $\vec{Z}'$ of $\vec{Z} - \vec{X}$, we have $(M,\vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W} \leftarrow \vec{w}, \vec{Z}' \leftarrow \vec{z}^*]\phi$.

AC2(b^u)- a sufficiency condition

Suppose that a prisoner dies either if A loads B's gun and B shoots or if C loads and shoots his gun. Taking D to represent the prisoner's death, we have $D = (A \wedge B) \vee C$. In the actual context u, A loads B's gun, B does not shoot, but C does load and shoot his gun, so the prisoner dies. That is, A = 1, B = 0, and C = 1. Clearly C = 1 is a cause of D = 1.

- With AC2(b^o), A = 1 is a cause of D = 1.
  For we can take $\vec{W} = \{B, C\}$ and consider the contingency where B = 1 and C = 0. $(M, \vec{u}) \models [A \leftarrow 0, B \leftarrow 1, C \leftarrow 0](D = 0)$, whereas $(M, \vec{u}) \models [A \leftarrow 1, B \leftarrow 1, C \leftarrow 0](D = 1)$.

AC2(b^u)-a sufficiency condition

Definition 8

- AC2(b^u). If $\vec{z}^*$ is such that $(M, \vec{u}) \models \vec{Z} = \vec{z}^*$, then for all subsets $\vec{W}'$ of $\vec{W}$ and subsets $\vec{Z}'$ of $\vec{Z} - \vec{X}$, we have $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W}' \leftarrow \vec{w}, \vec{Z}' \leftarrow \vec{z}^*]\phi$.

AC2(b^u)-a sufficiency condition

Suppose that a prisoner dies either if A loads B's gun and B shoots or if C loads and shoots his gun. Taking D to represent the prisoner's death,we have $D = (A \wedge B) \vee C$. In the actual context u, A loads B's gun, B does not shoot, but C does load and shoot his gun, so the prisoner dies. That is, A = 1, B = 0, and C = 1. Clearly C = 1 is a cause of D = 1.

- With AC2(b^o), A = 1 is a cause of D = 1.
  For we can take $\vec{W} = \{B, C\}$ and consider the contingency where B = 1 and C = 0. $(M, \vec{u}) \models [A \leftarrow 0, B \leftarrow 1, C \leftarrow 0](D = 0)$, whereas $(M, \vec{u}) \models [A \leftarrow 1, B \leftarrow 1, C \leftarrow 0](D = 1)$.
- AC2(b^u) fails because $(M, \vec{u}) \models [A \leftarrow 1, C \leftarrow 0](D = 0)$.

AC2(a^m)

Definition 8

- AC2(a^m). There is a set $\vec{W}$ of variables in V and a setting $\vec{x}'$ of the variables in $\vec{X}$ such that if $(M, \vec{u}) \models \vec{W} = \vec{w}^*$, then $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}^*]\neg\phi$.

- The modified definition does not need to mention $\vec{Z}$ (although $\vec{Z}$ can be taken to be the complement of $\vec{W}$).
- The need for a sufficiency condition arises only if we are considering contingencies that differ from the actual setting in AC2(a).
  AC2(b^o) holds : $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W} \leftarrow \vec{w}^*]\phi$
  AC2(b^u) holds : $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W}' \leftarrow \vec{w}^*]\phi$

Agreement among the definitions

Theorem

- (a) If X = x is part of a cause of ϕ in $(M, \vec{u})$ according to the modified HP definition, then X = x is a cause of ϕ in $(M, \vec{u})$ according to the original HP definition.
  (b) If X = x is part of a cause of ϕ in $(M, \vec{u})$ according to the modified HP definition, then X = x is a cause of ϕ in $(M, \vec{u})$ according to the updated HP definition.
  (c) If X = x is part of a cause of ϕ in $(M, \vec{u})$ according to the updated HP definition, then X = x is a cause of ϕ in $(M, \vec{u})$ according to the original HP definition.
  (d) If $\vec{X} = \vec{x}$ is a cause of ϕ in $(M, \vec{u})$ according to the original HP definition, then $|\vec{X}| = 1$.

Agreement among the definitions

Proposition

- If X = x is a but-for cause of ϕ in $(M, \vec{u})$, then X = x is a cause of ϕ according to all three variants of the HP definition.

Proof AC1 and AC3 are trivially satisfied.
According to the condition, there must be a possible value x' of X such that $(M, \vec{u}) \models [X \leftarrow x']\neg\phi$. Let $\vec{W} = \emptyset$, AC2(a) and AC2(a^m) holds.
If $(M, \vec{u}) \models \vec{Z} = \vec{z}^*$, $(M, \vec{u}) \models [X \leftarrow x, \vec{Z}' \leftarrow \vec{z}^*]\phi$ for all subsets Z' of $V - \{X\}$.
Because $\vec{W} = \emptyset$, AC2(b^u) follows immediately from AC2(b^o).

Example 1 : Forest fire

In the conjunctive model M^c, $FF = L \wedge MD$.

- Consider the context (1, 1), so the lightning strikes and the arsonist drops the match. Both $L = 1$ and $MD = 1$ are but-for causes for $FF = 1$. By proposition, both $L = 1$ and $MD = 1$ are causes of $FF = 1$ in $(M^c, (1, 1))$ according to all three variants of the definition. By AC3, it follows that $L = 1 \wedge MD = 1$ is not a cause of $FF = 1$ in $(M^c, (1, 1))$.

Example 1 : Forest fire

In the disjunctive model M^d, $FF = L \vee MD$.

- With the original and updated definition, we have that both $L = 1$ and $MD = 1$ are causes of $FF = 1$ in $(M^d, (1, 1))$.
  In the case of $L = 1$ (the argument for $MD = 1$ is identical) :
  Clearly, AC1 and AC3 are satisfied.
  For AC2, let $\vec{Z} = \{L, FF\}$, $\vec{W} = MD$, $x = 0$, and $w = 0$. Clearly, $(M^d, (1, 1)) \models [L \leftarrow 0, MD \leftarrow 0](FF = 0)$, AC2(a) is satisfied. Moreover, $(M^d, (1, 1)) \models [L \leftarrow 1, MD \leftarrow 0](FF = 1)$ and $(M^d, (1, 1)) \models [L \leftarrow 1](FF = 1)$, AC2($b^u$) and AC2($b^o$) are satisfied.
- In the case of the modified definition $L = 1 \wedge MD = 1$ is a cause of $FF = 1$. This shows why theorem is worded in terms of parts of causes.

Example 2 : Voting

Consider the voting scenario where there are 11 voters. If Andy wins 6–5, then all the definitions agree that each of the voters for Andy is a cause of Andy's victory. Suppose that Andy wins 11–0. The original and updated HP definition would still call each of the voters a cause of Andy winning. According to the modified HP definition, any subset of six voters is a cause of Andy winning. If we think of the subset as being represented by a disjunction, it can be thought of as a but-for cause of Andy winning.

- $X = x$ is a *but − for cause* of ϕ in $(M, \vec{u})$ if AC1 holds and there exists some x' such that $(M, \vec{u}) \models [X \leftarrow x']\neg\phi$.
- $\vec{X} = \vec{x}$ $(X_1 = x_1 \vee ... \vee X_n = x_n)$ is a *but − for cause* of ϕ in $(M, \vec{u})$ if AC1 holds and only some $\vec{x}' = (x_1', ..., x_n')$ such that $x_i \neq x_i'$ for all $i = 1, ..., n$ satisfies $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}']\neg\phi$.

Example 3 : Throwing rocks

A coarse causal model M_{RT} :

- one exogenous variable u, which determines whether Bob and Andy throw
  three endogenous variables :
  - AT for "Andy throws", with values 0 (Andy does not throw) and 1 (he does)
  - BT for "Bob throws", with values 0 (he doesn't) and 1 (he does)
  - BS for "bottle shatters", with values 0 (it doesn't shatter) and 1 (it does)

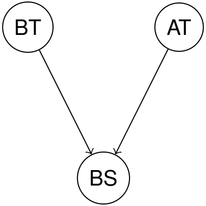
  $BS = AT \vee BT$
- Both BT=1 and AT=1 are classified as causes of BS=1 in $(M_{RT}, \vec{u})$ according to the original and updated HP definition (the conjunction $AT = 1 \wedge BT = 1$ is the cause according to the modified HP definition).
- M_{RT} cannot distinguish the case where both rocks hit the bottle simultaneously from the case where Andy's rock hits first.

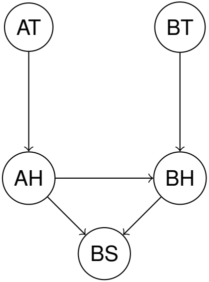
Example 3 : Throwing rocks

M'_{RT} (add two new variables to the model) :

- BH for "Bob's rock hits the (intact) bottle", with values 0 (it doesn't) and 1 (it does) ;
  AH for "Andy's rock hits the bottle", again with values 0 and 1.
  We now modify the equations as follows :
  BS = 1 if AH = 1 or BH = 1 ;
  SH = 1 if AT = 1 ;
  BH = 1 if BT = 1 and AH = 0.

Example 3 : Throwing rocks





According to all three variants of the HP definition, AT = 1 is a cause of BS = 1 in $(M'_{RT}, \vec{u})$, but BT = 1 is not (consider $AT \in \vec{W}$ or $AH \in \vec{W}$, $BH \in \vec{W}$ or $BH \in \vec{Z}$).

Example 4 : Recording a vote

A votes for a candidate. A's vote is recorded in two optical scanners B and C. D collects the output of the scanners ; D' records whether just scanner B records a vote for the candidate. The candidate wins (i.e., WIN = 1) if any of A, D, or D' is 1. The value of A is determined by the exogenous variable. Structural equations from M_V :

- B = A ;
  C = A ;
  D = B $\wedge$ C ;
  D' = B $\wedge\neg$A ;
  WIN = A $\vee$ D $\vee$ D'.
- In the actual context u, A = 1, so B = C = D = WIN = 1 and D' = 0.
  Claim that B = 1 $\wedge$ C = 1 is a cause of WIN = 1 in (M_V, u) according to the updated HP definition.
  - For B = 1, suppose that B = 1 satisfies AC2(a) and AC2(b^u), then A$\in$ $\vec{W}$, A=0.
    $D' \in \vec{W}$ or $D' \in \vec{Z}$: $(M_V, u) \models [B \leftarrow 1, A \leftarrow 0, D' \leftarrow 0]WIN = 0$.
    .
    For C=1, in the same way, consider $(M_V, u) \models [C \leftarrow 1, A \leftarrow 0]WIN = 0$.

  None of B = 1, C = 1, or B = 1$\wedge$C = 1 is a cause of WIN = 1 in (M_V, u) according to the modified HP definition.
  The only cause of WIN = 1 in (M_V, u) according to the modified HP definition is A = 1.

Example 5 : Medical treatment

Billy contracts a serious but nonfatal disease. He is treated on Monday, so is fine Tuesday morning. Had Monday's doctor forgotten to treat Billy, Tuesday's doctor would have treated him, and he would have been fine Wednesday morning. The catch : one dose of medication is harmless, but two doses are lethal. Is the fact that Tuesday's doctor did not treat Billy the cause of him being alive on Wednesday morning ?

- The causal model has three random variables :
  - MT (Monday treatment) : 1-yes ; 0-no
  - TT (Tuesday treatment) : 1-yes ; 0-no
  - BMC (Billy's medical condition) :
    0-OK Tues. and Wed. morning,
    1-sick Tues. morning, OK Wed. morning,
    2-sick both Tues. and Wed. morning,
    3-OK Tues. morning, dead Wed. morning

Example 5 : Medical treatment

Billy contracts a serious but nonfatal disease. He is treated on Monday, so is fine Tuesday morning. Had Monday's doctor forgotten to treat Billy, Tuesday's doctor would have treated him, and he would have been fine Wednesday morning. The catch : one dose of medication is harmless, but two doses are lethal. Is the fact that Tuesday's doctor did not treat Billy the cause of him being alive on Wednesday morning ?

- The causal model has three random variables :
  - MT (Monday treatment) : 1-yes ; 0-no
  - TT (Tuesday treatment) : 1-yes ; 0-no
  - BMC (Billy's medical condition) :
    0-OK Tues. and Wed. morning,
    1-sick Tues. morning, OK Wed. morning,
    2-sick both Tues. and Wed. morning,
    3-OK Tues. morning, dead Wed. morning
- In the actual context u, MT=1, then TT=0, BMC=0.
  According to the modified definition,
  MT = 1 is a cause of BMC = 0 and of TT = 0
  TT = 0 is a cause of Billy's being alive ($BMC = 0 \vee BMC = 1 \vee BMC = 2$).
  MT = 1 is not a cause of Billy's being alive.
  Causality is not transitive !

Theorem

- (a) If X = x is part of a cause of ϕ in $(M, \vec{u})$ according to the modified HP definition, then X = x is a cause of ϕ in $(M, \vec{u})$ according to the original HP definition.
  (b) If X = x is part of a cause of ϕ in $(M, \vec{u})$ according to the modified HP definition, then X = x is a cause of ϕ in $(M, \vec{u})$ according to the updated HP definition.
  (c) If X = x is part of a cause of ϕ in $(M, \vec{u})$ according to the updated HP definition, then X = x is a cause of ϕ in $(M, \vec{u})$ according to the original HP definition.
  (d) If $\vec{X} = \vec{x}$ is a cause of ϕ in $(M, \vec{u})$ according to the original HP definition, then $|\vec{X}| = 1$.

- **Proof** We just need to prove (b), (c) and (d).
  For part (b), suppose that X = x is part of a cause of ϕ in $(M, \vec{u})$ according to the modified HP definition, so that there is a cause $\vec{X} = \vec{x}$ such that X = x is one of its conjuncts. Claim that X = x is a cause of ϕ according to the updated HP definition.
  By definition, there must exist a value $\vec{x}' \in R(\vec{X})$ and a set $\vec{W} \subseteq V - \{\vec{X}\}$ such that if $(M, \vec{u}) \models \vec{W} = \vec{w}^*$, then $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}^*]\neg\phi$.
  If $\vec{X} = \{X\}$, clearly $X = x$ is a cause of ϕ according to the updated definition. By AC3, there is a contradiction.

- **Proof** If $|\vec{X}| > 1$, suppose $\vec{X} = (X_1, ..., X_n), X = X_1$, let $\vec{X}_{-1}$ denote all components of the vector except the first one, we want to show that $X_1 = x_1$ is a cause of ϕ in $(M, \vec{u})$ according to the updated HP definition. Clearly, AC1 and AC3 hold.
  We have $(M, \vec{u}) \models [X_1 \leftarrow x_1', \vec{X}_{-1} \leftarrow \vec{x}'_{-1}, \vec{W} \leftarrow \vec{w}^*]\neg\phi$, AC2(a) holds.
  Suppose that there exists a set $\vec{Z}' \subseteq V - \vec{X} \cup \vec{W}$, a set $\vec{W}' \subseteq \vec{W}$ and a set $\vec{X}' \subseteq \vec{X}_{-1}$ such that $(M, \vec{u}) \models \vec{Z}' = \vec{z}'$ and $(M, \vec{u}) \models [X_1 \leftarrow x_1, \vec{X}' \leftarrow \vec{x}'_{-1}, \vec{W}' \leftarrow \vec{w}^*, \vec{Z}' \leftarrow \vec{z}']\neg\phi$. Then $\vec{X}' = \vec{x}_{-1}$ satisfies AC2(a^m). $\vec{X} = \vec{x}$ is not a cause of ϕ according to the modified HP definition because AC3 does not hold.
  For part (c), the proof is again similar in spirit.
  For part (d), suppose that $\vec{X} = \vec{x}$ is a cause of ϕ in $(M, \vec{u})$ according to the original HP definition and $|\vec{X}| > 1$. Let X = x be a conjunct of $\vec{X} = \vec{x}$. We can show that X = x is a cause of ϕ in $(M, \vec{u})$ according to the original HP definition.

$AX_{rec}(S)$

- Restrict S here to signatures where V is finite and the range of each endogenous variable is finite.
- For each signature S, consider $M_{rec}(S)$, the set of all recursive causal models with signature S.

$AX_{rec}(S)$

- $L(S)$ is rich enough to express actual causality in models in $M_{rec}(S)$. Given $\vec{X}, \vec{u}, \phi$, and S, for each variant of the HP definition, there is a formula $\psi \in L(S)$ such that $(M, \vec{u}) \models \psi$ for a causal model $M \in M_{rec}(S)$ iff $\vec{X} = \vec{x}$ is a cause of ϕ in $(M, \vec{u})$.
  For the original definition, $\psi : \vec{X} = \vec{x} \wedge \phi \wedge \bigvee_{\vec{W} \subseteq V - \vec{X}} \bigvee_{\vec{w} \in R(\vec{W})} \bigvee_{\vec{x}' \in R(\vec{X})} [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}]\neg\phi \wedge (\vec{Z} = \vec{z}^* \Rightarrow \bigwedge_{\vec{Z}' \subseteq \vec{Z} - \vec{X}} [\vec{X} \leftarrow \vec{x}, \vec{W} \leftarrow \vec{w}, \vec{Z}' \leftarrow \vec{z}^*]\phi)$, where $\vec{Z} = V - \vec{W}, \vec{z}^* \in R(\vec{Z}) (= \times_{Z \in \vec{Z}} R(Z))$
  For the updated definition, $\psi : \vec{X} = \vec{x} \wedge \phi \wedge \bigvee_{\vec{W} \subseteq V - \vec{X}} \bigvee_{\vec{w} \in R(\vec{W})} \bigvee_{\vec{x}' \in R(\vec{X})} [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}]\neg\phi \wedge (\vec{Z} = \vec{z}^* \Rightarrow \bigwedge_{\vec{Z}' \subseteq \vec{Z} - \vec{X}} \bigwedge_{\vec{W}' \subseteq \vec{W}} [\vec{X} \leftarrow \vec{x}, \vec{W}' \leftarrow \vec{w}, \vec{Z}' \leftarrow \vec{z}^*]\phi)$, where $\vec{Z} = V - \vec{W}, \vec{z}^* \in R(\vec{Z})$
  For the modified definition,
  $\psi : \vec{X} = \vec{x} \wedge \phi \wedge \bigvee_{\vec{W} \subseteq V - \vec{X}} \bigvee_{\vec{x}' \in R(\vec{X})} (\vec{W} = \vec{w}^* \Rightarrow [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}^*]\neg\phi$, where $\vec{w}^* \in R(\vec{W})$

$AX_{rec}(S)$

- To help characterize causal reasoning in $M_{rec}(S)$, where $S = (U, V, R)$, define a formula $Y \rightsquigarrow Z$ for $Y, Z \in V$, read "Y affects Z".
  $(M, \vec{u}) \models Y \rightsquigarrow Z$ iff $(M, \vec{u}) \models \bigvee_{\vec{x} \in R(\vec{X})} \bigvee_{y \in R(Y)} \bigvee_{y' \in R(Y), z \neq z'} [\vec{X} \leftarrow \vec{x}, Y \leftarrow y] Z = z \wedge [\vec{X} \leftarrow \vec{x}, Y \leftarrow y'] Z = z'$, where $\vec{X} = V - \{Y, Z\}$

$AX_{rec}(S)$

C0	All substitution instances of propositional tautologies	
C1	$[\vec{Y} \leftarrow \vec{y}](X = x) \Rightarrow [\vec{Y} \leftarrow \vec{y}](X \neq x')$ *if* $x, x' \in R(X), x \neq x'$	(equality)
C2	$\bigvee_{x \in R(X)} [\vec{Y} \leftarrow \vec{y}](X = x)$	(definiteness)
C3	$[\vec{X} \leftarrow \vec{x}](W = w) \wedge [\vec{X} \leftarrow \vec{x}](Y = y) \Rightarrow [\vec{X} \leftarrow \vec{x}, W \leftarrow w](Y = y)$	(composition)
C4	$[X \leftarrow x, \vec{W} \leftarrow \vec{w}](X = x)$	(effectiveness)
C5	$(X_0 \rightsquigarrow X_1 \wedge ... \wedge X_{k-1} \rightsquigarrow X_k) \Rightarrow \neg(X_k \rightsquigarrow X_0)$ *if* $X_k \neq X_0$	(recursiveness)
C6	(a) $[\vec{X} \leftarrow \vec{x}]\neg\phi \Leftrightarrow \neg[\vec{X} \leftarrow \vec{x}]\phi$	
	(b) $[\vec{X} \leftarrow \vec{x}](\phi \wedge \psi) \Leftrightarrow ([\vec{X} \leftarrow \vec{x}]\phi \wedge [\vec{X} \leftarrow \vec{x}]\psi)$	(determinism)
	(c) $[\vec{X} \leftarrow \vec{x}](\phi \vee \psi) \Leftrightarrow ([\vec{X} \leftarrow \vec{x}]\phi \vee [\vec{X} \leftarrow \vec{x}]\psi)$	
MP	From ϕ and $\phi \Rightarrow \psi$, infer ψ	(modus ponens)

Soundness and Completeness

Theorem

- $AX_{rec}(S)$ is a sound and complete axiomatization for the language $L(S)$ in $M_{rec}(S)$.

- **Proof**
  The fact that C1, C2, C4, and C6 are valid is almost immediate.
  C3 : If $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}](W = w)$, and $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}](Y = y)$, in the unique solution to the equations when $\vec{X}$ is set to $\vec{x}$, W = w and Y=y. So the unique solution to the equations when $\vec{X}$ is set to $\vec{x}$ is the same as the unique solution to the equations when $\vec{X}$ is set to $\vec{x}$ and W is set to w. $(\vec{x}, w, y, \vec{z})$
  Thus $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, W \leftarrow w](Y = y)$.
  C5 :
  k=1, if $(M, \vec{u}) \models X_0 \rightsquigarrow X_1$, then $X_0 \preceq_{\vec{u}} X_1$, so we cannot have $X_1 \preceq_{\vec{u}} X_0$ if $X_0 \neq X_1$.
  Thus $(M, \vec{u}) \models \neg X_1 \rightsquigarrow X_0$.
  k=n, $(M, \vec{u}) \models (X_0 \rightsquigarrow X_1 \wedge ... \wedge X_{n-1} \rightsquigarrow X_n) \Rightarrow \neg(X_n \rightsquigarrow X_0)$ *if* $X_n \neq X_0$.
  k=n+1, if $(M, \vec{u}) \models (X_0 \rightsquigarrow X_1 \wedge ... \wedge X_n \rightsquigarrow X_{n+1})$, then $X_0 \preceq_{\vec{u}} X_{n+1}$, so we cannot have $X_{n+1} \preceq_{\vec{u}} X_0$ if $X_0 \neq X_{n+1}$. Thus $(M, \vec{u}) \models \neg X_{n+1} \rightsquigarrow X_0$.
  Therefore we have
  $(M, \vec{u}) \models (X_0 \rightsquigarrow X_1 \wedge ... \wedge X_{k-1} \rightsquigarrow X_k) \Rightarrow \neg(X_k \rightsquigarrow X_0)$ *if* $X_k \neq X_0$.

Completeness

- **Proof** For completeness, it suffices to prove that if a formula ϕ in $L(S)$ is consistent with $AX_{rec}(S)$, then ϕ is satisfiable in $M_{rec}(S)$.
  So suppose that a formula $\phi \in L(S)$, with $S = (U, V, R)$, is consistent with $AX_{rec}(S)$. Consider a maximal consistent set C of formulas that includes ϕ.
  Now construct a causal model $M = (S, F) \in M_{rec}(S)$ and context $\vec{u}$ such that $(M, \vec{u}) \models \psi$ for every formula $\psi \in C$.
  The idea : the formulas in C determine F.
  For each variable $X \in V$, let $\vec{Y}_X = V - \{X\}$. By C1 and C2, for all $\vec{y} \in R(\vec{Y}_X)$, there is a unique $x \in R(X)$ such that $[\vec{Y}_X \leftarrow \vec{y}](X = x) \in C$. For all contexts $\vec{u} \in R(\vec{U})$, define $F_X(\vec{y}, \vec{u}) = x$. This defines F_X for all endogenous variables X, and hence F. Let $M = (S, F)$.

Completeness

- **Proof** Claim that for all formulas $\psi \in L(S)$, $\psi \in C$ iff $(M, \vec{u}) \models \psi$ for all contexts $\vec{u}$.
  Take $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\psi$ to hold if there is a unique solution to the equations in the model $M_{\vec{Y} \leftarrow \vec{y}}$ in context $\vec{u}$, and ψ holds in that solution.
  $\psi : [\vec{Y} \leftarrow \vec{y}]X = x$, by induction on $|V| - |\vec{Y}|$.
  $|V| - |\vec{Y}|$=0, $X \in \vec{Y}$, If $[\vec{Y} \leftarrow \vec{y}](X = x) \in C$, then there is a unique solution $\vec{y}$ to the equations in $M_{\vec{Y} \leftarrow \vec{y}}$ for every context $\vec{u}$ and $X = x$ in that solution.
  Conversely, $[\vec{Y} \leftarrow \vec{y}](X = x)$ is an instance of C4, so must be in C.

Completeness

- **Proof**

  $|V| - |\vec{Y}|$=1, $X \in \vec{Y}$:

  If $[\vec{Y} \leftarrow \vec{y}](X = x) \in C$,

  then there is a unique solution $(\vec{y}, F_W(\vec{y}, \vec{u}))$ for which $W \notin \vec{Y}$ to the equations in $M_{\vec{Y} \leftarrow \vec{y}}$ for every context $\vec{u}$ and $X = x$ in that solution follows from C4, C1 and C6.

  Precisely, if $[\vec{Y} \leftarrow \vec{y}](X = x)$ can be written as

  $[\vec{Y}' \leftarrow \vec{y}, X \leftarrow x']X = x$, $\vec{Y}' = \vec{Y} - \{X\}$, according to C1, we have

  $[\vec{Y} \leftarrow \vec{y}](X \neq x') \in C$, that is $[\vec{Y} \leftarrow \vec{y}](X = x') \notin C$, which contradicts with C4.

  If $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]X = x$, then there is a unique solution $(\vec{y}, F_W(\vec{y}, \vec{u}))$ to the equations in $M_{\vec{Y} \leftarrow \vec{y}}$ for every context $\vec{u}$ and $X = x$ in that solution. Thus

  $[\vec{Y} \leftarrow \vec{y}]X = x$ must be an instance of C4, so must be in C.

  $X \notin \vec{Y}$:

  If $[\vec{Y} \leftarrow \vec{y}](X = x) \in C$,

  then $F_X(\vec{y}, \vec{u}) = x$, there is a unique solution $(\vec{y}, x)$ to the equations in $M_{\vec{Y} \leftarrow \vec{y}}$ for every context $\vec{u}$ and $X = x$ in that solution.

  If $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]X = x$, then we must have $[\vec{Y} \leftarrow \vec{y}](X = x) \in C$, given how F_X is defined.

Completeness

- **Proof**
  $|V| - |\vec{Y}|$=k>1, and $[\vec{Y} \leftarrow \vec{y}](X = x) \in C$
  To show that there is a solution, define a vector $\vec{v}$ and show that it is in fact a solution.
  If $W \in \vec{Y}$ and $W \leftarrow w$ is a conjunct of $\vec{Y} \leftarrow \vec{y}$, then set the W component of $\vec{v}$ to w. If W is not in Y , then set the W component of $\vec{v}$ to the unique value w such that $[\vec{Y} \leftarrow \vec{y}](W = w) \in C$.
  Claim that $\vec{v}$ is a solution to the equations in $M_{\vec{Y} \leftarrow \vec{y}}$ for all contexts $\vec{u}$.
  Let $V_1 \in V - \vec{Y}$, $V_2 \in V$, v_1 and v_2 be the values of these variables in $\vec{v}$. Then $[\vec{Y} \leftarrow \vec{y}]V_1 = v_1 \in C$, $[\vec{Y} \leftarrow \vec{y}]V_2 = v_2 \in C$, and C contains every instance of C3, it follows that $[\vec{Y} \leftarrow \vec{y}, V_1 \leftarrow v_1]V_2 = v_2 \in C$.

Completeness

- **Proof** Since V_2 was arbitrary, we have proved that for $|V| - |\vec{Y}| = 0$, $[\vec{Y} \leftarrow \vec{y}](X = x) \in C$ iff there is a unique solution $\vec{y}$ to the equations in the model $M_{\vec{Y} \leftarrow \vec{y}}$ for every context $\vec{u}$ and $X = x$ in that solution, for $|V| - |\vec{Y}| = 1$,
  $[\vec{Y} \leftarrow \vec{y}](X = x) \in C$ iff there is a unique solution $\vec{v}'$ to the equations in the model $M_{\vec{Y} \leftarrow \vec{y}}$ for every context $\vec{u}$ ($\vec{v}'$=($\vec{y}$, $F_W(\vec{y}, \vec{u})$)) and $X = x$ in that solution. $\vec{y}$ and $\vec{v}'$ both satisfy the principle of construction of $\vec{v}$.
  By induction hypothesis that $\vec{v}$ is the unique solution to the equations in $M_{\vec{Y} \leftarrow \vec{y}, V_1 \leftarrow v_1}$ for all contexts $\vec{u}$.
  For every endogenous variable Z other than V_1, the equation F_Z for Z is the same in $M_{\vec{Y} \leftarrow \vec{y}}$ and $M_{\vec{Y} \leftarrow \vec{y}, V_1 \leftarrow v_1}$. Every equation except possibly that for V_1 is satisfied by $\vec{v}$ in $M_{\vec{Y} \leftarrow \vec{y}}$ for all contexts $\vec{u}$.
  Since $|V - \vec{Y}| \geqslant 2$, we can repeat this argument starting with a variable in $V - \vec{Y}$ other than V_1 to conclude that, every equation in $M_{\vec{Y} \leftarrow \vec{y}}$ is satisfied by $\vec{v}$ for all contexts $\vec{u}$. That is, $\vec{v}$ is a solution to the equations in $M_{\vec{Y} \leftarrow \vec{y}}$ for all contexts $\vec{u}$.

Completeness

- **Proof** It remains to show that $\vec{v}$ is the unique solution to the equations in $M_{\vec{Y}\leftarrow\vec{y}}$ for all contexts $\vec{u}$. Suppose there were another solution, say $\vec{v}'$, to the equations in $M_{\vec{Y}\leftarrow\vec{y}}$ for all contexts $\vec{u}$. There must be some variable V_1 whose value in $\vec{v}$ is different from its value in $\vec{v}'$. Suppose that the value of V_1 in $\vec{v}$ is v_1 and its value in $\vec{v}'$ is v_1', with $v_1 \neq v_1'$. By construction, $[\vec{Y} \leftarrow \vec{y}](V_1 = v_1) \in C$. By C1, $[\vec{Y} \leftarrow \vec{y}](V_1 \neq v_1') \in C$. Since $\vec{v}'$ is a solution to the equations in $M_{\vec{Y}\leftarrow\vec{y}}$ in context $\vec{u}$, it is easy to check that $\vec{v}'$ is also a solution to the equations in $M_{\vec{Y}\leftarrow\vec{y},V_1\leftarrow v_1'}$ in context $\vec{u}$.
  By the induction hypothesis, $\vec{v}'$ is the unique solution to the equations in $M_{\vec{Y}\leftarrow\vec{y},V_1\leftarrow v_1'}$ in every context. (For $|V| - |\vec{Y}| = 0$, if $\vec{a}$ is a solution to the equations in $M_{\vec{Y}\leftarrow\vec{y}}$ in context $\vec{u}$, then $\vec{a} = \vec{y}$. For $|V| - |\vec{Y}| = 1$, if $\vec{a}$ is a solution to the equations in $M_{\vec{Y}\leftarrow\vec{y}}$ in context $\vec{u}$, then $\vec{a} = (\vec{y}, F_W(\vec{y}, \vec{u}))$ for all context $\vec{u}'$. If not, there is a context $\vec{u}'$ such that $\vec{a} = (\vec{y}, w'), F_W(\vec{y}, \vec{u}) \neq w', F_W(\vec{y}, \vec{u}) = w, F_W(\vec{y}, \vec{u}') = w'$. From the definition of F_W, we have $[\vec{Y} \leftarrow \vec{y}]W = w' \in C$ and $[\vec{Y} \leftarrow \vec{y}]W = w \in C$, contradiction.)

Completeness

- **Proof** Let V_2 be a variable other than V_1 in $V - \vec{Y}$, and let v_2' be the value of V_2 in $\vec{v}'$. In the same way, $\vec{v}'$ is the unique solution to the equations in $M_{\vec{Y}\leftarrow\vec{y}, V_2\leftarrow v_2'}$ in all contexts. It follows from the induction hypothesis that $[\vec{Y} \leftarrow \vec{y}, V_1 \leftarrow v_1']V_2 = v_2'$ and $[\vec{Y} \leftarrow \vec{y}, V_2 \leftarrow v_2']V_1 = v_1'$ are both in C.
  Claim that $[\vec{Y} \leftarrow \vec{y}]V_1 = v_1' \in C$, this together with the fact that $[\vec{Y} \leftarrow \vec{y}]V_1 \neq v_1' \in C$, contradicts the consistency of C.
  To prove that $[\vec{Y} \leftarrow \vec{y}]V_1 = v_1' \in C$, by C5, at most one of $V_1 \rightsquigarrow V_2$ and $V_2 \rightsquigarrow V_1$ is in C. If $V_2 \rightsquigarrow V_1 \notin C$, $\neg V_2 \rightsquigarrow V_1 \in C$, so V_2 does not affect V_1. Since $[\vec{Y} \leftarrow \vec{y}, V_2 \leftarrow v_2']V_1 = v_1' \in C$, it follows that $[\vec{Y} \leftarrow \vec{y}]V_1 = v_1' \in C$.
  Suppose that $V_1 \rightsquigarrow V_2 \notin C$, then an argument analogous to that above shows that $[\vec{Y} \leftarrow \vec{y}]V_2 = v_2' \in C$. If $[\vec{Y} \leftarrow \vec{y}]V_1 = v_1' \notin C$, by C2, $[\vec{Y} \leftarrow \vec{y}]V_1 = v_1'' \in C$ for some $v_1'' \neq v_1$. Applying C3, it follows that $[\vec{Y} \leftarrow \vec{y}, V_2 \leftarrow v_2']V_1 = v_1'' \in C$, which contradicts with $[\vec{Y} \leftarrow \vec{y}, V_2 \leftarrow v_2']V_1 = v_1' \in C$. This completes the uniqueness proof. Besides, it is clear that $X = x$ is in $\vec{v}$.

Completeness

- **Proof** For the converse, suppose that
  $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}](X = x), [\vec{Y} \leftarrow \vec{y}](X = x) \notin C$. Then, by C2,
  $[\vec{Y} \leftarrow \vec{y}](X = x') \in C$ for some $x' \neq x$. By the argument above,
  $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}](X = x')$, a contradiction.
  $\psi : [\vec{Y} \leftarrow \vec{y}]\neg\psi'$
  $[\vec{Y} \leftarrow \vec{y}]\neg\psi' \in C$
  iff $[\vec{Y} \leftarrow \vec{y}]\psi' \notin C$
  iff $(M, \vec{u}) \not\models [\vec{Y} \leftarrow \vec{y}]\psi'$ for all context $\vec{u}$
  iff $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}]\neg\psi'$ for all context $\vec{u}$
  $\psi : [\vec{Y} \leftarrow \vec{y}]\psi_1 \wedge \psi_2$ and $\psi : [\vec{Y} \leftarrow \vec{y}]\psi_1 \vee \psi_2$ are similar to the above proof, using axiom C6.
  Thus $\psi : [\vec{Y} \leftarrow \vec{y}]\psi''$ has been proved.
  The case that ψ is a Boolean combination of formulas of the form $[\vec{Y} \leftarrow \vec{y}]\psi''$ is easy to be checked.

Completeness

- **Proof** One more thing needs to be checked : that $M \in M_{rec}(S)$.
  For each context $\vec{u}$, define a relation $\preceq'$ on the endogenous variables by taking $X \preceq' Y$ if $(M, \vec{u}) \models X \rightsquigarrow Y$. It is easy to see that $(M, \vec{u}) \models X \rightsquigarrow X$, $\preceq'$ is reflexive.
  Let $\preceq$ be the transitive closure of $\preceq'$.
  For suppose, that $X \preceq Y$ and $Y \preceq X$, $X \neq Y$, then there exist variables $X_0, ..., X_n$ and $Y_0, ..., Y_m$ such that $X = X_0 = Y_m$, $Y = X_n = Y_0$, $X_0 \preceq' X_1, ..., X_{n-1} \preceq' X_n$, $Y_0 \preceq' Y_1, ..., Y_{m-1} \preceq' Y_m$. That is
  $(M, \vec{u}) \models X_0 \rightsquigarrow X_1 \wedge ... X_{n-1} \rightsquigarrow X_n \wedge Y_0 \rightsquigarrow Y_1 \wedge ... \wedge Y_{m-1} \rightsquigarrow Y_m$
  Since every instance of C5 is in C, every instance of C5 is true in $(M, \vec{u})$ for all contexts $\vec{u}$.
  If $Y_{m-1} \neq X$, then $(M, \vec{u}) \models \neg Y_{m-1} \rightsquigarrow Y_m$.
  If $Y_{m-1} = X$, $(M, \vec{u}) \models X \rightsquigarrow X_1 \wedge ... X_{n-1} \rightsquigarrow Y \wedge Y \rightsquigarrow Y_1 \wedge ... \wedge Y_{m-2} \rightsquigarrow Y_{m-1}$, repeat the argument, there must exist a Y_i such that $X \neq Y_i$,
  $(M, \vec{u}) \models X \rightsquigarrow X_1 \wedge ... X_{n-1} \rightsquigarrow Y \wedge Y \rightsquigarrow Y_1 \wedge ... \wedge Y_i \rightsquigarrow Y_{i+1}$, then contradiction.
  Thus we have proved that $\preceq$ is a partial order on the endogenous variables. If X affects Y in context $\vec{u}$, then $(M, \vec{u}) \models X \rightsquigarrow Y$, that is $X \preceq Y$. Therefore, $M \in M_{rec}(S)$.