Isaac Newton Institute, University of Cambridge
Programme on Neural Networks and Machine Learning
Probabilistic Graphical Models Week
September 1–5, 1997

Causal Models: What Does One Need to Know?

Ross D. Shachter
Engineering-Economic Systems and Operations Research
Stanford University

joint work with
David E. Heckerman
Microsoft Research

What does one need to know to build, learn, and use causal models?

What does one need to know to <u>build</u>, <u>learn</u>, and <u>use</u> causal models?

In building a causal model

- **What assumptions are you making?**
- **What are the implications of those assumptions?**

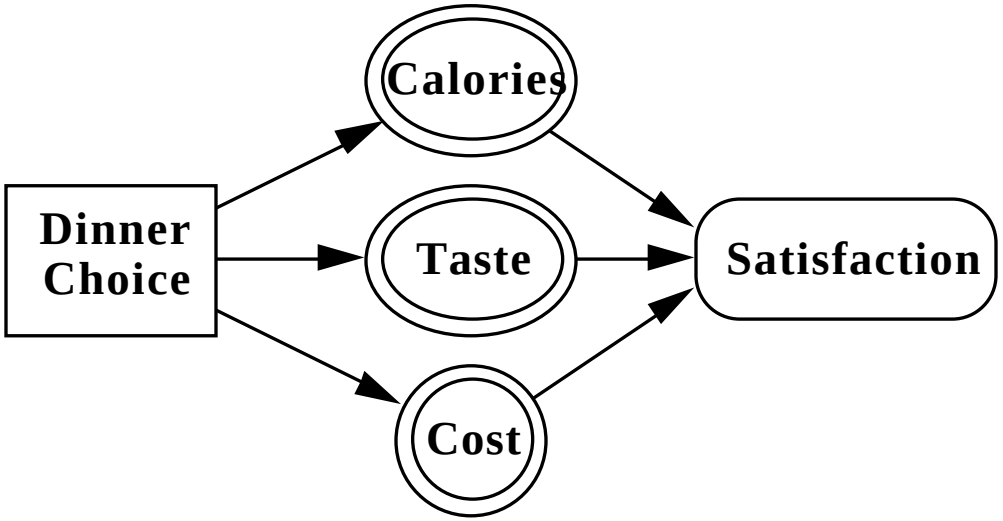
To build and use causal models, what information do you need to get from decision-makers, experts, and data?

Of course, it depends on which tasks will you be performing with a causal model, such as:

- **Inference from passive observation**
- **Predicting the effects of actions/interventions**
- **Inferring the causes of effects**

Related Work:
Pearl; Pearl et al; Spirtes, Glymour, Scheines; Rubin; Simon; Robins; Howard; and others.

Brief Introduction to Influence Diagrams

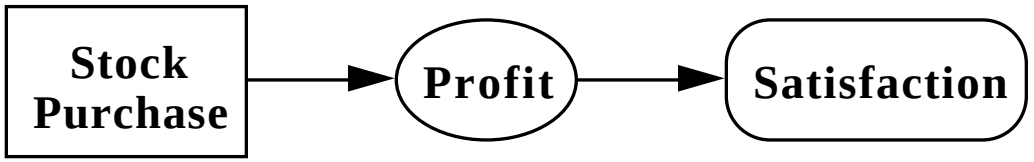


Models the situation faced by a single, rational decision-maker with a one-time decision.

A decision is an irrevocable allocation of resources.

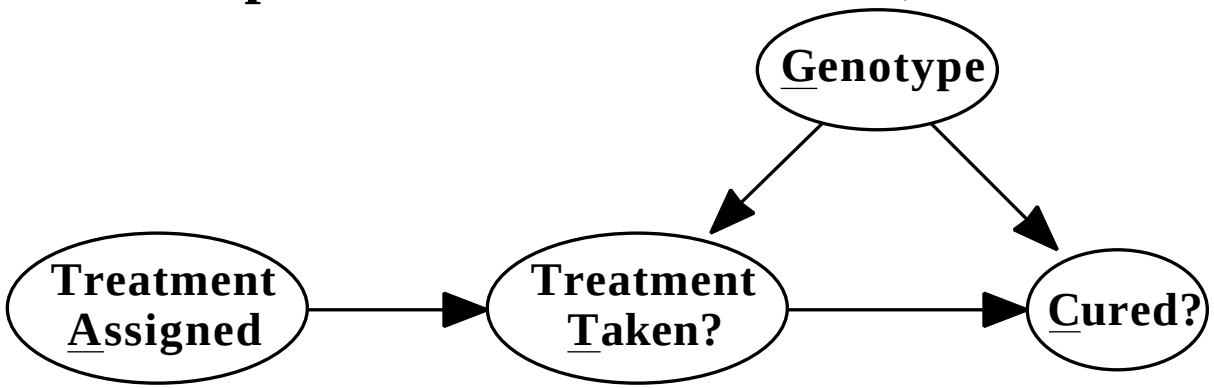
Belief network with added components

- ovals represent uncertain quantities
- rectangles represent decisions, under the decision-maker's control
- double ovals represent deterministic functions
- [rounded rectangles represent the objective value which the decision-maker seeks to maximize]

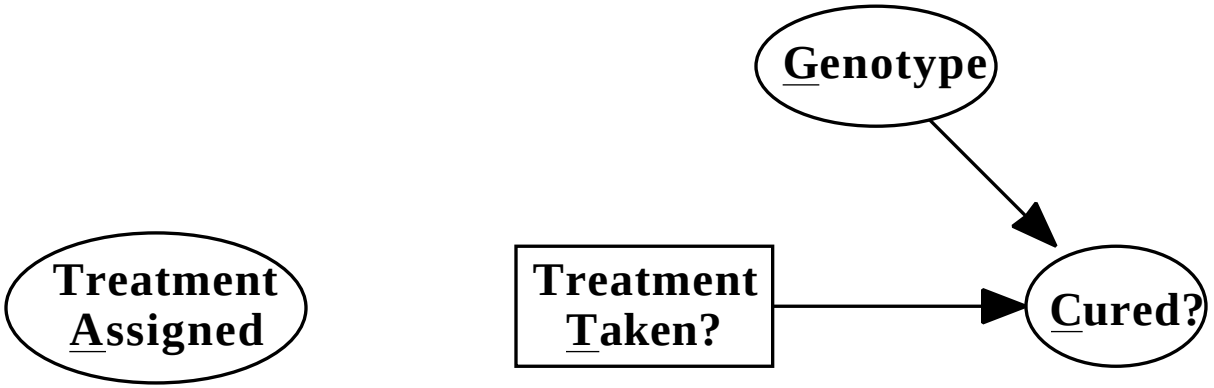


Pearl's Causal Theory

Given a belief network with a causal interpretation, a causal graph, the parents of each node represent its direct causes,



you model interventions or manipulations of variables simply by transforming them into decisions and cutting their incoming arcs.



Other, more complex, operations allow you to reason counterfactually.

Pearl's Causal Theory (Continued)

Given uncertain variables U, DAG G with nodes corresponding to U, and for each $x \in U$, there is a decision x' corresponding to x with an alternative set(x) for each possible value x of x plus an alternative "idle",

if for each $x \in U$, Pa(x) $\cup$ {x'} are "direct causes" for x

then the relationships among the variables are characterized by

$$x = f_x\ (\mathrm{Pa}(x),\ x',\ e_x\ ) \text{ for all } x \in U,$$

where f_x is a deterministic function satisfying x = x if x' = set(x)

and {e_x } are exogenous variables (possibly dependent).

This is called the <u>causal theory</u> underlying the causal graph.

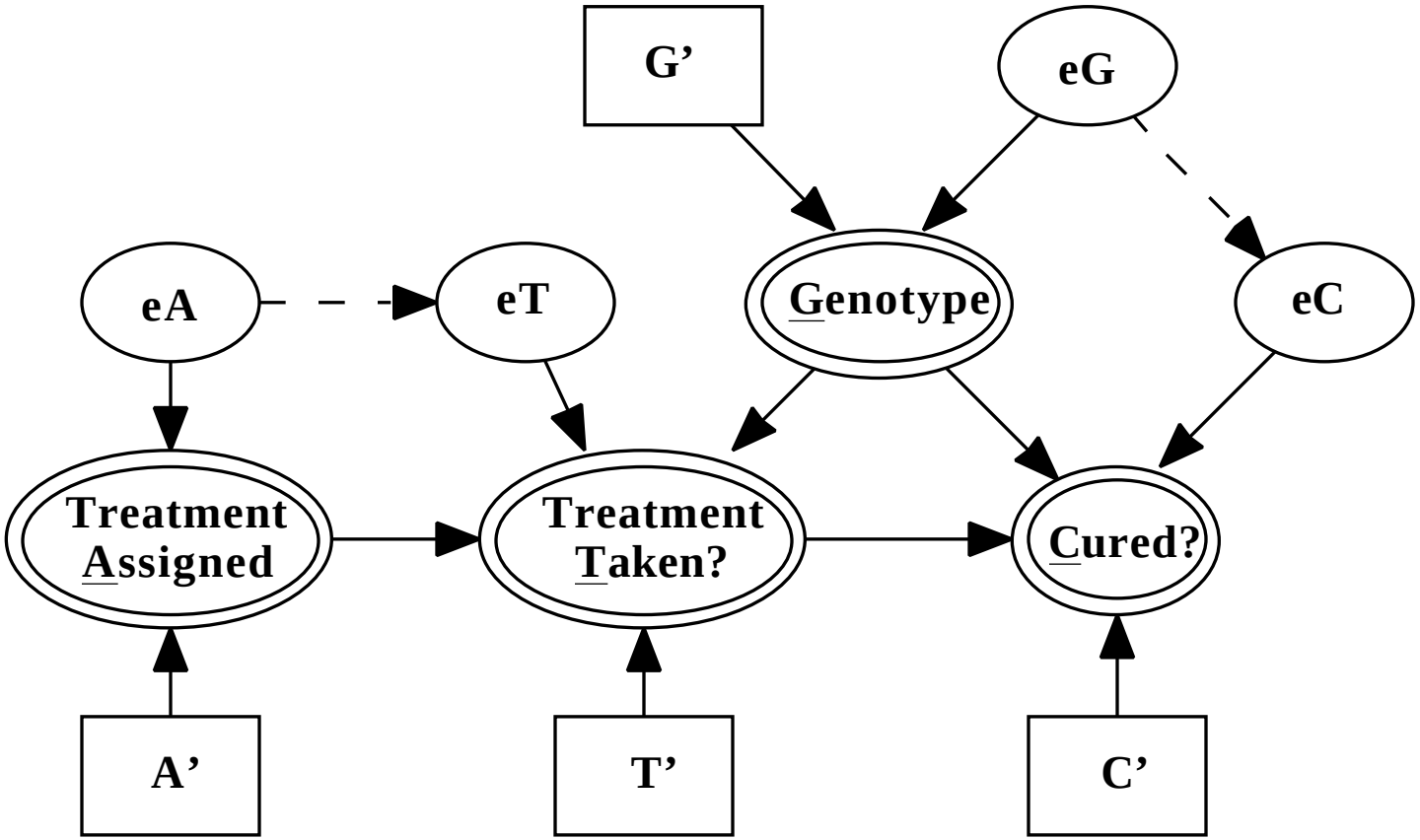
This theory is great, but there are a few questions . . .

- **Where does it come from?**
- **What assumptions are you making?**
- **What are the functions and exogenous variables?**
- **What is the set variable x'?**

. . . and, ultimately,

What is meant by a "cause" for x?

Pearl's Causal Theory (Continued)



What is a set variable?

The model itself depends on <u>how</u> the variables are set.

"Magic Genie" problem: the mechanism isn't specified

- **Does the patient believe he/she is receiving the treatment?**
- **Is medication administered covertly in his/her food?**
- **Is his/her jaw wired shut to prevent taking the treatment?**
- **Is the patient killed to prevent taking the treatment?**

"Fat Hand" problem: side effects possible even given a mechanism

- **Patient infected by dirty needle or other patients when treated**
- **Patient has to cut back on necessities to pay for treatment**
- **(Proximal components--heat in circuits, Piper Alpha)**

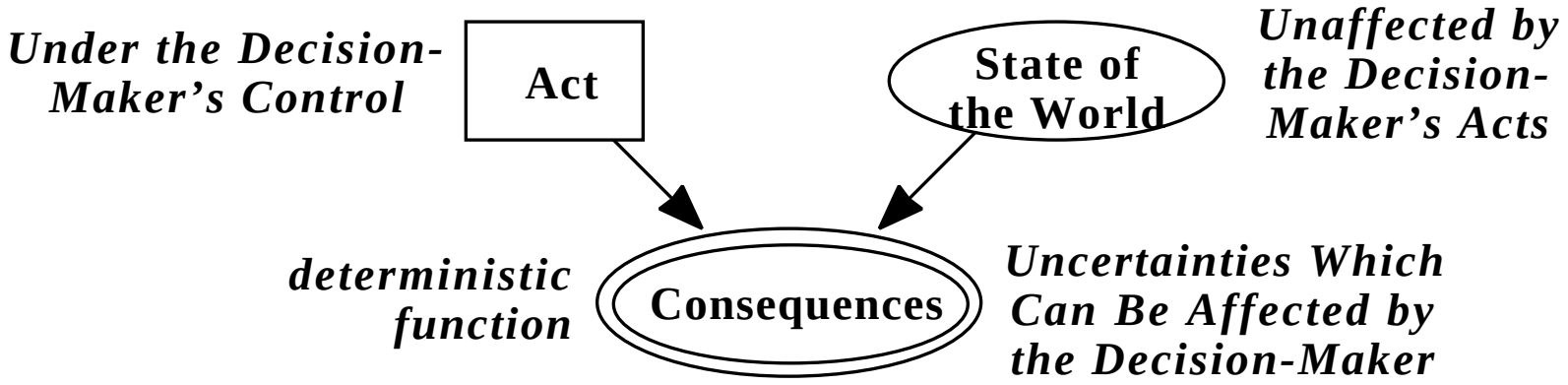
Bottom Line: the cause may depend on how the variables are set!
For example, we have a quite different causal model if

- **A hypochondriac patient is cured by believing he/she has received the treatment, even if it was just a placebo**
- **Treatment Taken? is set to False by killing the patient**

A more primitive framework needs to be established to address this . . .

Savage's Framework

Uncertain quantities can be decomposed into components which are under your control and components which you cannot affect:

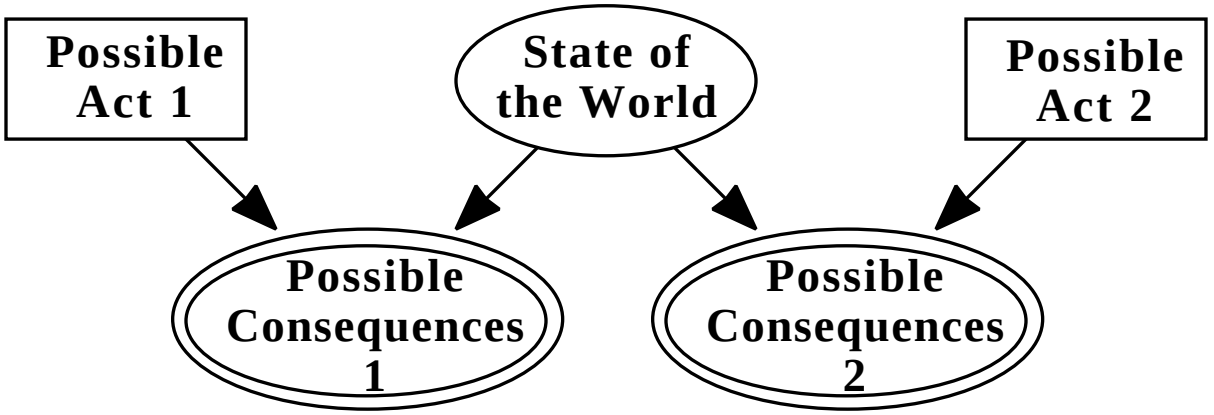


"Your [spouse] has just broken five good eggs into a bowl when you come in and volunteer to finish making the omelet. A sixth egg, which for some reason must either be used for the omelet or wasted altogether, lies unbroken beside the bowl. You must decide what to do with the unbroken egg."

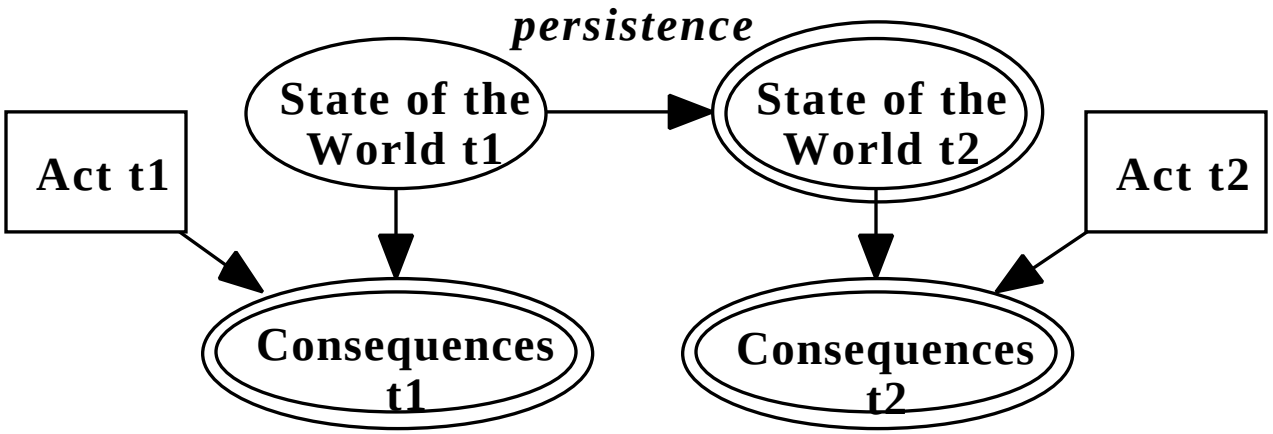
State of the world	Acts: break into bowl	break into saucer	throw away
egg good	6 egg omelet	6 egg omelet and a saucer to wash	5 egg omelet and a good egg destroyed
egg bad	no omelet and 5 good eggs destroyed	5 egg omelet and a saucer to wash	5 egg omelet

Savage's Framework for Temporal Models

Static Model:

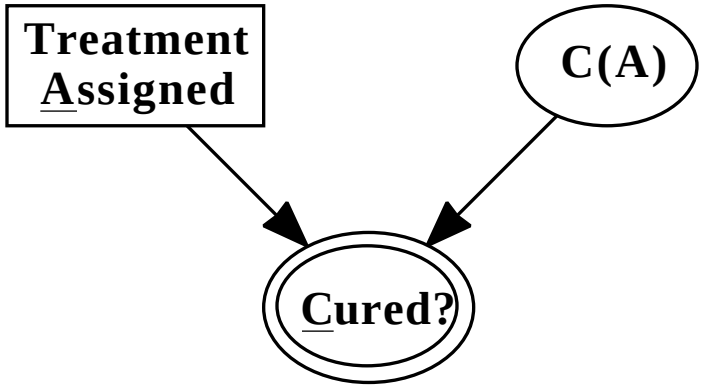


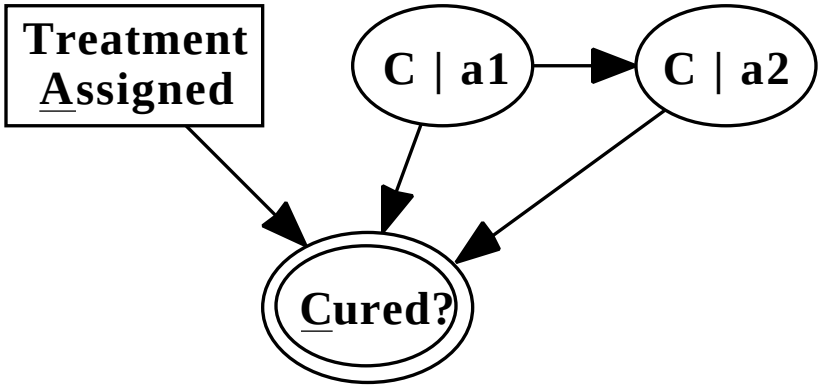
Temporal Model:



Mapping Variable as State of the World

A Mapping Variable describes consequences for all possible actions:



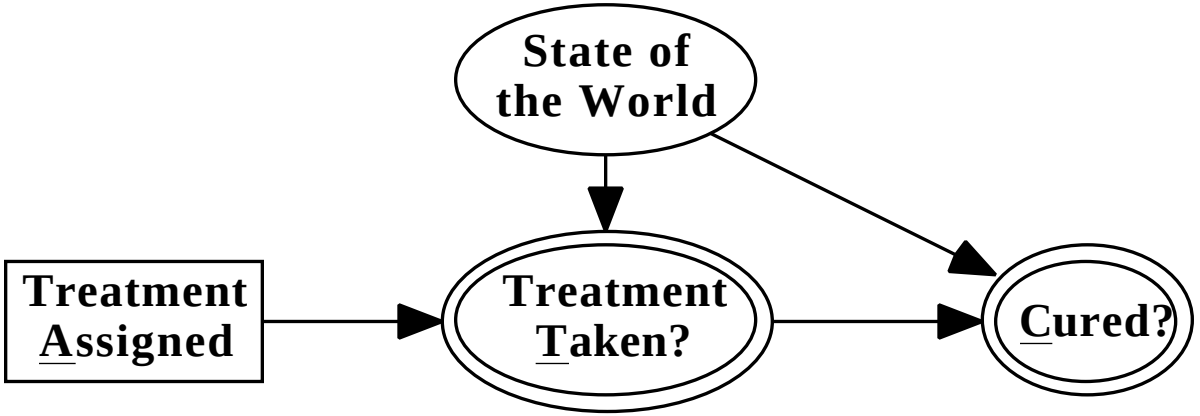


State of the World		Act: Treatment Assigned	Treatment Not Assigned
1:	helped	cured	not cured
2:	hurt	not cured	cured
3:	always cured	cured	cured
4:	never cured	not cured	not cured

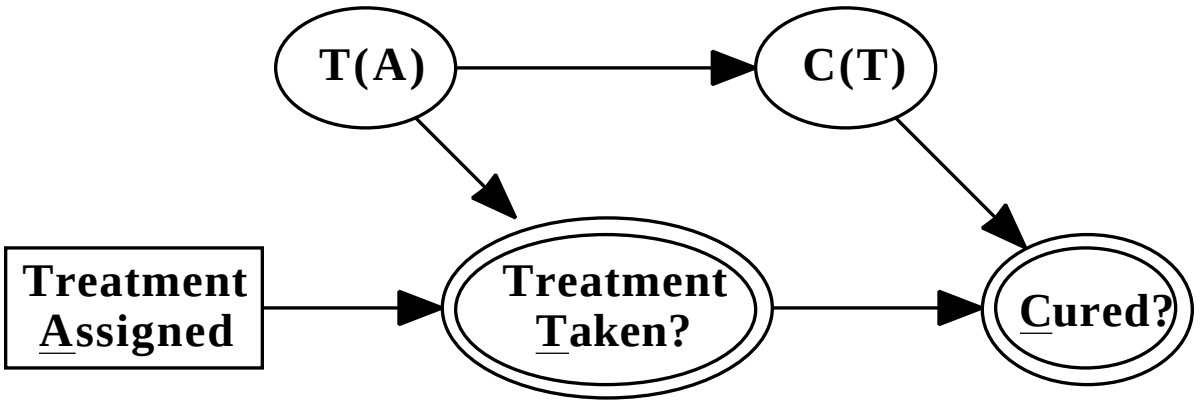
Mapping variables can be thought of as "counterfactuals" because the different cases cannot occur simultaneously. For example, state 1 is {Cured? would be True if Treatment Assigned were True and Cured? would be False if Treatment Assigned were False}

Multiple Mapping Variables

The State of the World can be decomposed into multiple mapping variables.

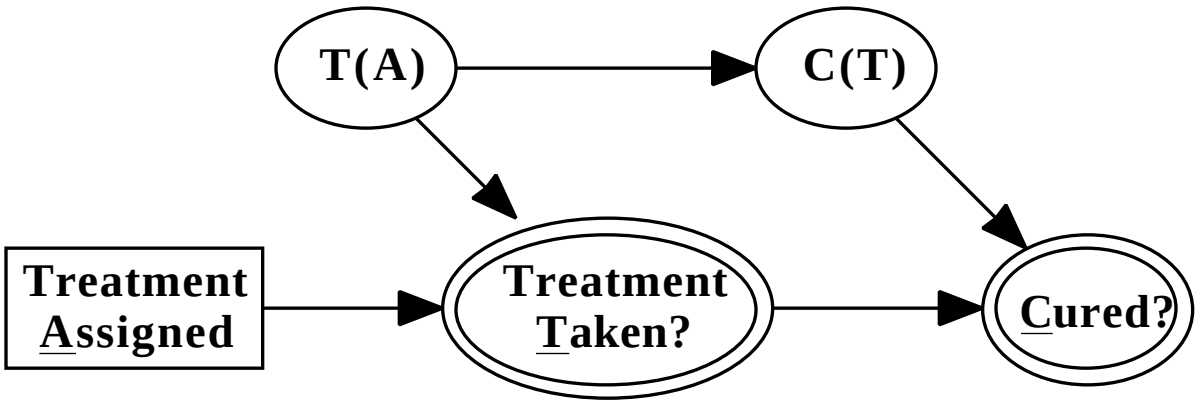


can become

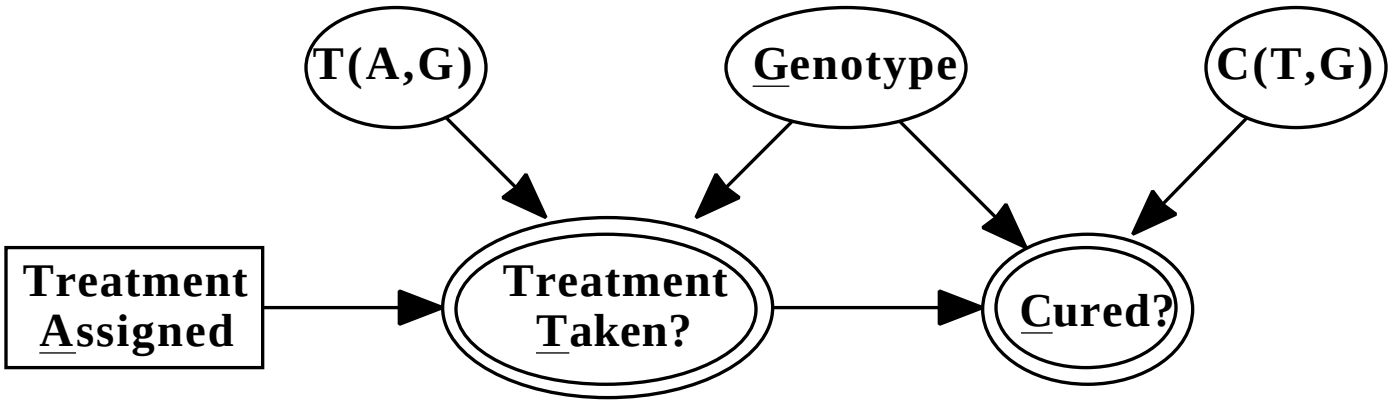


Causal Markov Assumption

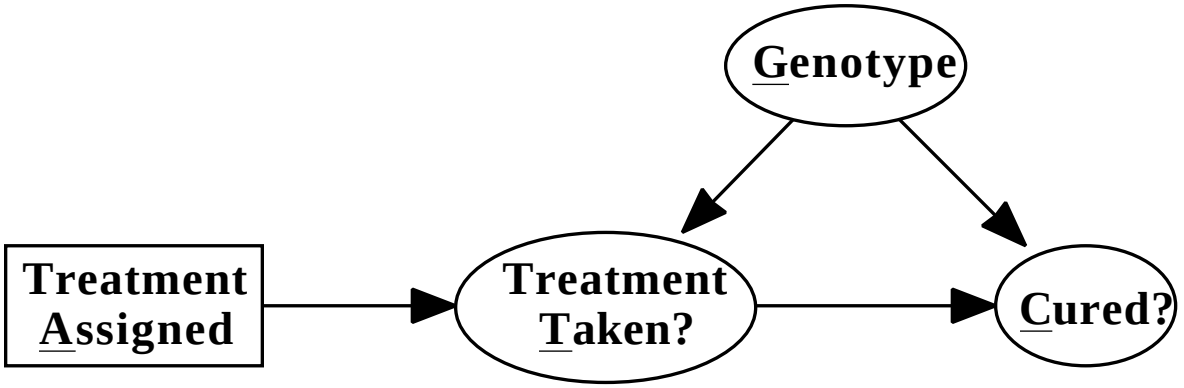
An influence diagram satisfies the Causal Markov Assumption if all mapping variables are independent. In that case, the causal structure corresponds to independence statements in the graph.
It might require the introduction of additional uncertain variables.



becomes



Responsiveness

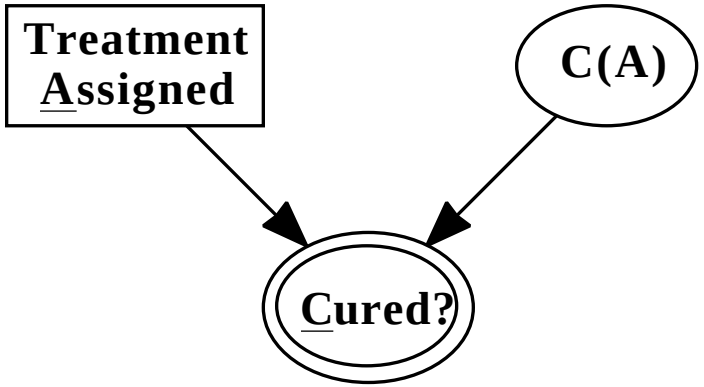


A variable <u>responsive</u> to decisions might change if they are changed.
Usually, variables downstream from a decision are responsive to it.
Treatment Taken? and Cured? are responsive to Treatment Assigned.
Genotype is unreponsive to Treatment Assigned
=> Genotype and Treatment Assigned are independent

A variable <u>responsive to</u> decisions <u>in worlds limited by</u> other variables can change when the decisions change even if the other variables do not.
Usually, limited unresponsiveness corresponds to a blocking of the paths emanating from decisions.
Cured? is unresponsive to Treatment Assigned in worlds limited by Treatment Taken?. Note that Treatment Assigned and Cured? are <u>not independent</u> given Treatment Taken?

Responsiveness (Continued)

A Mapping Variable describes consequences for all possible actions:



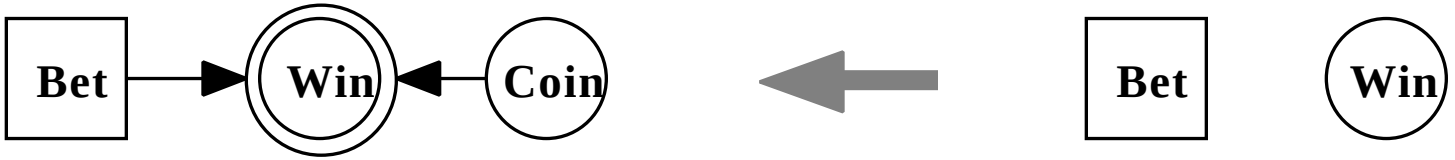
State of the World		Act: Treatment Assigned	Treatment Not Assigned
1:	helped	cured	not cured
2:	hurt	not cured	cured
3:	always cured	cured	cured
4:	never cured	not cured	not cured

Different choices for Treatment Assigned lead to different values of Cured?, so Cured? is responsive to Treatment Assigned.

Responsiveness (Continued)

Given uncertain variables *U*, decisions *D*, state of the world *S*
U[s,d]* is instance of *U* given act *d* and state of the world *s

An uncertain variable *x* is <u>unresponsive</u> to *D* if
$x[s,d_i] = x[s,d_j]$ **for all** d_i **,** d_j **, and states of the worlds *s*.**
That is, *x* is unaffected by *D*.
Example: Today's weather is unresponsive to your clothing choice.
If *x* is unresponsive to *D* (and no variables are observed before *D*) then *x* and *D* are independent, but not necessarily vice versa.



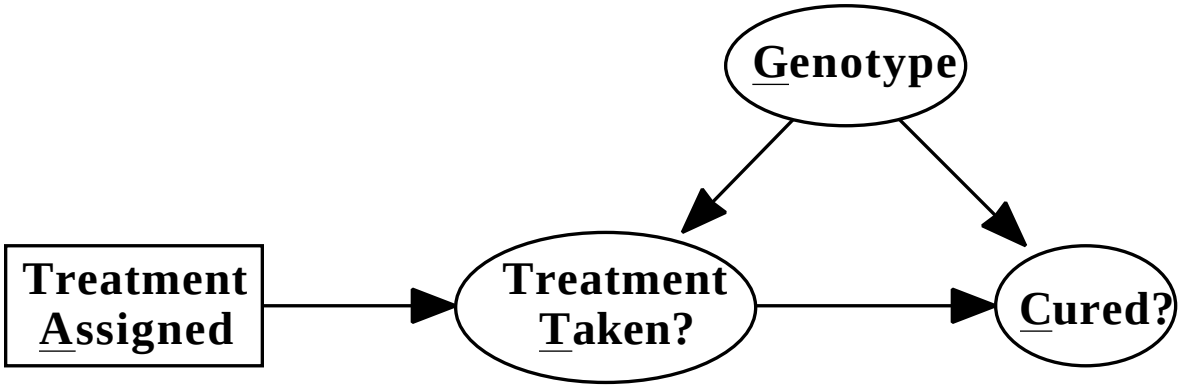
An uncertain variable *x* is <u>unresponsive</u> to *D* <u>in worlds limited by</u> variables *Y* if $Y[s,d_i] = Y[s,d_j] \Rightarrow x[s,d_i] = x[s,d_j]$ **for all** d_i **,** d_j **, and states of the world *s*. That is, *x* necessarily assumes the same instance under all acts that yield no change in the instance for *Y*.**
It is possible that *x* can still be relevant to *D* given *Y*.

Limited Unresponsiveness

State of the World			Treatment Assigned		Treatment Not Assigned	
			Treatment?	Cured?	Treatment?	Cured?
1:	complier	helped	taken	cured	not taken	not cured
2:	complier	hurt	taken	not cured	not taken	cured
3:	complier	cured	taken	cured	not taken	cured
4:	complier	not cured	taken	not cured	not taken	not cured
5:	defier	helped	not taken	not cured	taken	cured
6:	defier	hurt	not taken	cured	taken	not cured
7:	defier	cured	not taken	cured	taken	cured
8:	defier	not cured	not taken	not cured	taken	not cured
9:	taker	cured	taken	cured	taken	cured
10:	taker	not cured	taken	not cured	taken	not cured
11:	not taker	cured	not taken	not cured	not taken	not cured
12:	not taker	not cured	not taken	cured	not taken	cured
13:	impossible		taken	cured	taken	not cured
14:	impossible		taken	not cured	taken	cured
15:	impossible		not taken	not cured	not taken	cured
16:	impossible		not taken	cured	not taken	not cured

Cured? is unresponsive to Treatment Assigned in worlds limited by Treatment Taken?, and that is enforced by the mapping variable.

Cause (with respect to Decisions)



Given uncertain variables U, decisions D, and $x \in U$,

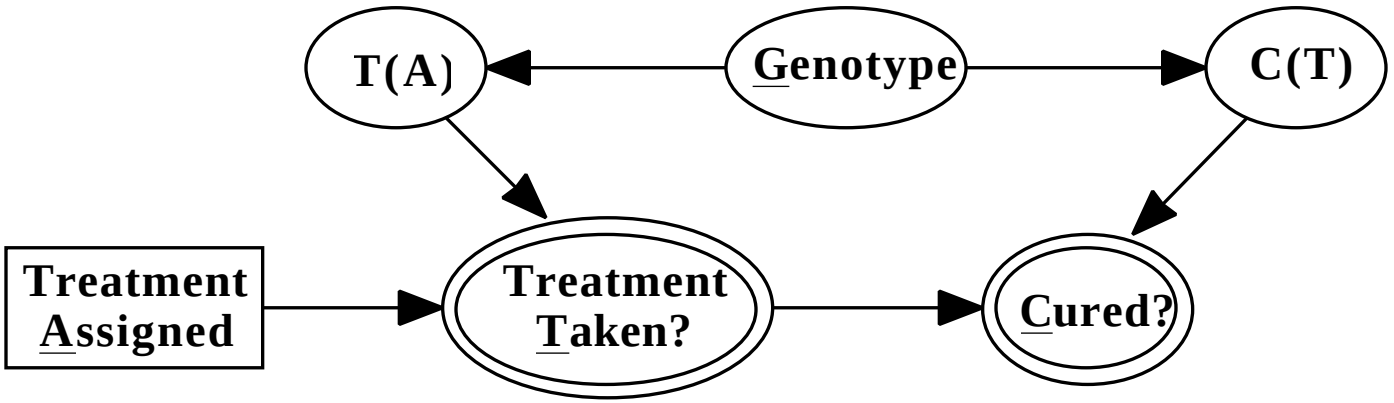
$C \subseteq D \cup U \setminus \{x\}$ is said to be a <u>cause</u> for x <u>with respect to</u> D
if C is a minimal set of variables such that
x is unresponsive to D in worlds limited by C.

With respect to Treatment Assigned,

- **Treatment Taken? is caused by Treatment Assigned**
- **Cured? is caused by Treatment Taken?**
- **Cured? is caused by Treatment Assigned**
- **Genotype is caused by Ø**

Usually, a cause is a minimal set blocking the paths from decisions.

Mapping Variable Definition of Cause



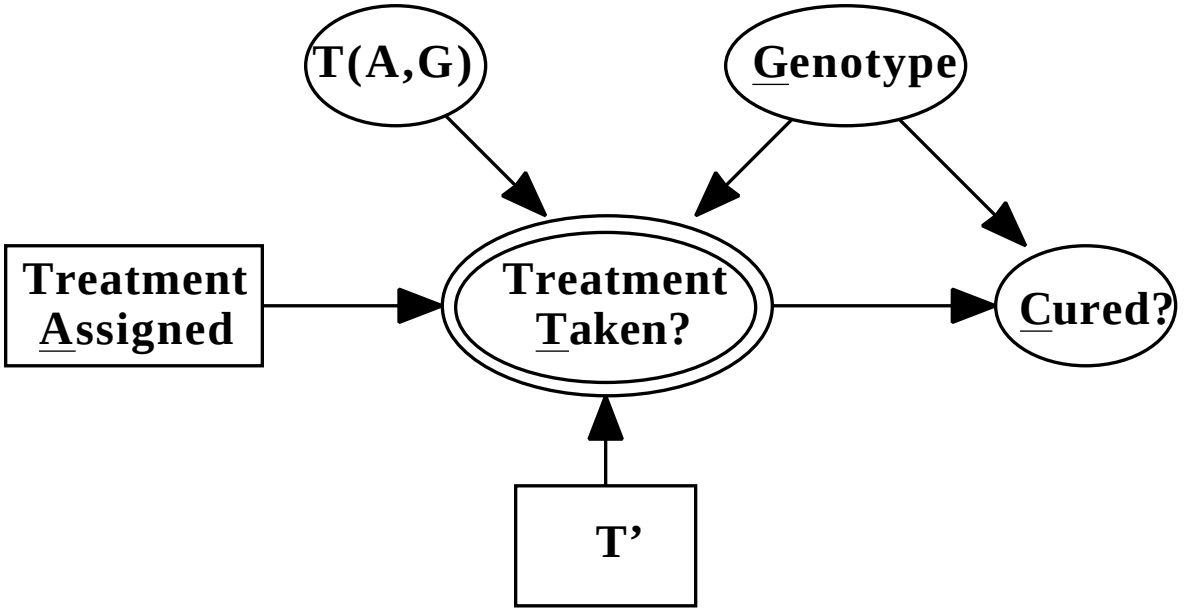
Given uncertain variables U, decisions D, $x \in U$, and $Y \subseteq D \cup U$,
x is unresponsive to decisions D in worlds limited by Y
if and only if $x(Y)$ is unresponsive to D;

$C \subseteq D \cup U \setminus \{x\}$ is a cause for x with respect to D
if and only if C is a minimal set of variables such that
$x(C)$ is unresponsive to D.

Thus, with respect to Treatment Assigned,

- **Treatment Taken? is caused by Treatment Assigned**
- **Cured? is caused by Treatment Taken?**
- **Cured? is caused by Treatment Assigned**
- **Genotype is caused by Ø**

What is a set variable?



We can now define the "setting" operation.
We define an atomic intervention on uncertain variable *x* as a decision *x'* such that

- **all other uncertain variables are unreponsive to *x'* in worlds limited by *x***
- ***x'* has an alternative corresponding to each possible state of *x* plus one more, "idle"**
  - **when *x'* is set to a state then *x* takes on that state**
  - **when *x'* is "idle", there is no manipulation of *x***

Pearl's Causal Theory Revisited

Given uncertain variables U,
DAG G with nodes corresponding to U,
and decision variables D comprising exactly one atomic interention x'
for each $x \in U$,

if for each $x \in U$,

$\mathrm{Pa}(x) \cup \{x'\}$ are "direct causes" for x with respect to D
then the relationships among the variables are characterized by

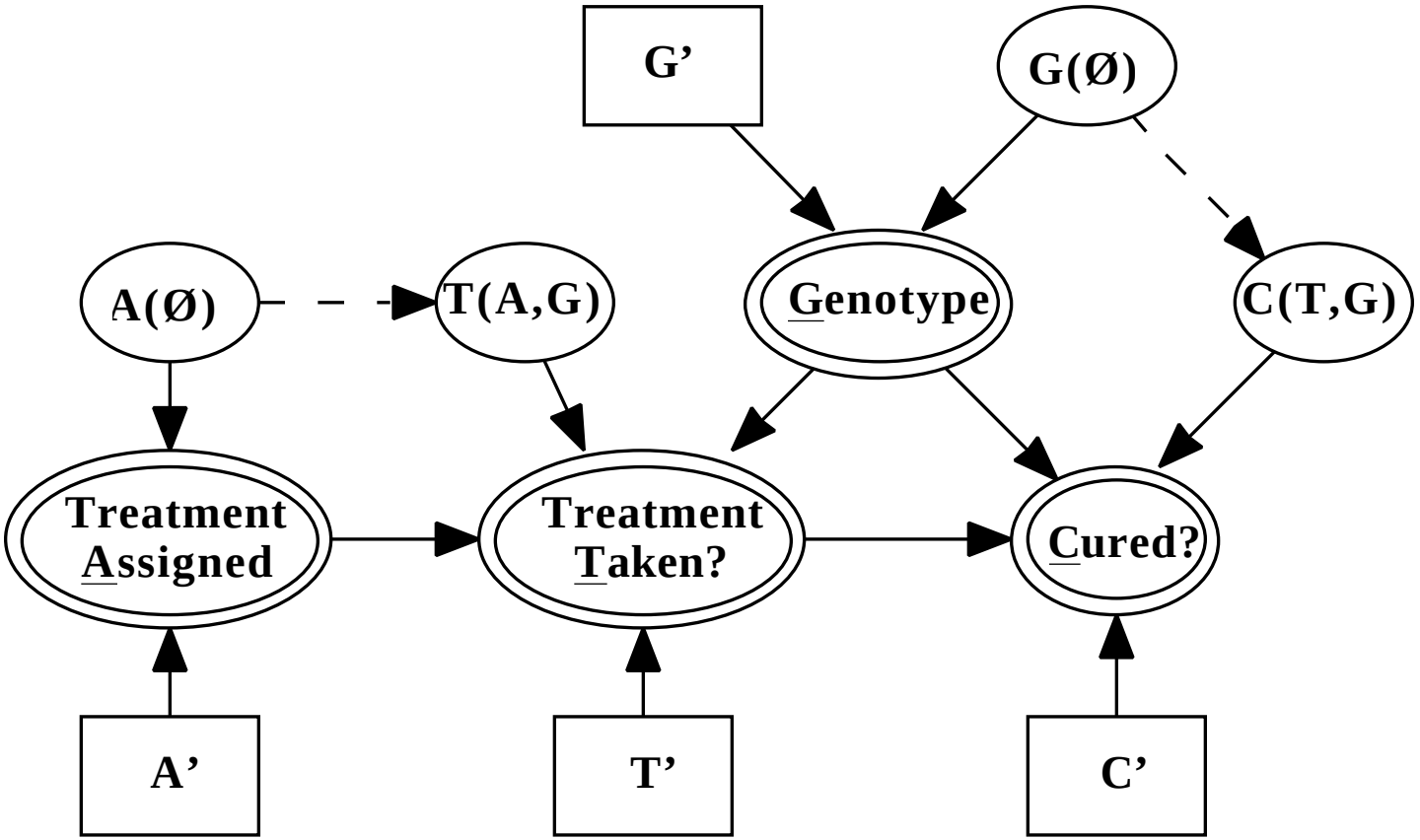
$$x = f_x\ (\ \mathrm{Pa}(x),\ x',\ x(\mathrm{Pa}(x), x')\ )\ \text{for all}\ x \in U,$$

where f_x is a deterministic function satisfying $x = \mathrm{x}$ if $x' = \mathrm{set}(\mathrm{x})$.

Exogenous variables $\{e_x\}$ have been replaced by the unresponsive (to D) mapping variables $\{x(\mathrm{Pa}(x), x')\}$.

Pearl's Causal Theory

The assumptions implicit in Pearl's causal theory can now be made explicit. There is an implicit decision parent for every uncertain variable, and the exogenous variables are mapping variables:



The dashed arcs are not present under the Causal Markov Assumption.

Pearl's Causal Theory--A Simpler Approach?

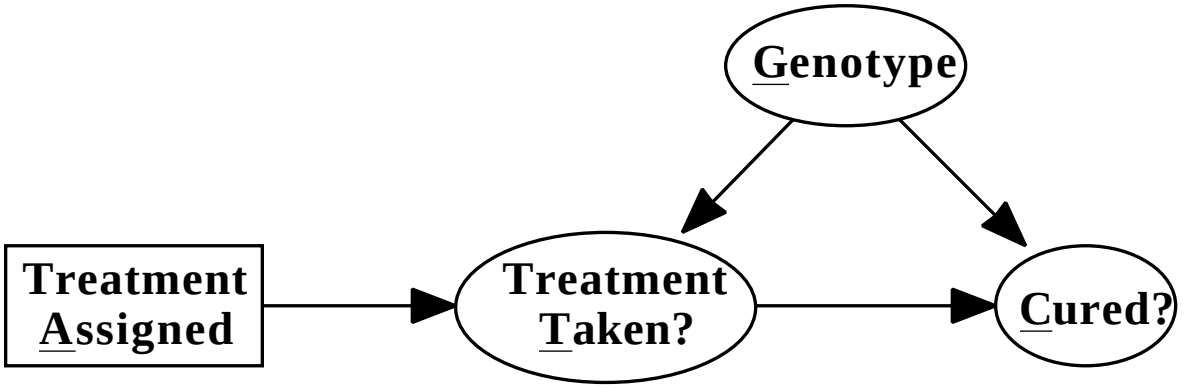
One problem with Pearl's causal theory is that all of the relationships in the causal graph are assumed to be causal.

A related problem is that it assumes the causal structure as a primitive. This is because

- **any variables downstream from the manipulated variable are responsive to the intervention**
- **any other variables are unresponsive to the intervention**

Finally, as we saw before, there are problems with set variables.

Why Define Cause with respect to Decisions?



- Obtain an unambiguous definition based on clear subjective judgements (responsiveness, well-defined decisions)
- Concentrate on determining (and manipulating) cause for those uncertain quantities you can affect
- No need to assume all dependencies are causal
- Any setting decisions you wish to hypothesize are permitted, including multiple distinct setting decisions for the same uncertain variable
- Well represented by graphical models

- Thus, although the decision approach might appear more complex at first, it is simpler than the setting approach!

Reasoning Tasks

The assumptions and data needed to construct and learn a causal model depend on the reasoning tasks to be performed.

- **Passive observation:**
  Suppose you have no ability to affect any of the variables you are observing, but you want to perform inference based on those observations. This might arise when someone else is making the decisions.

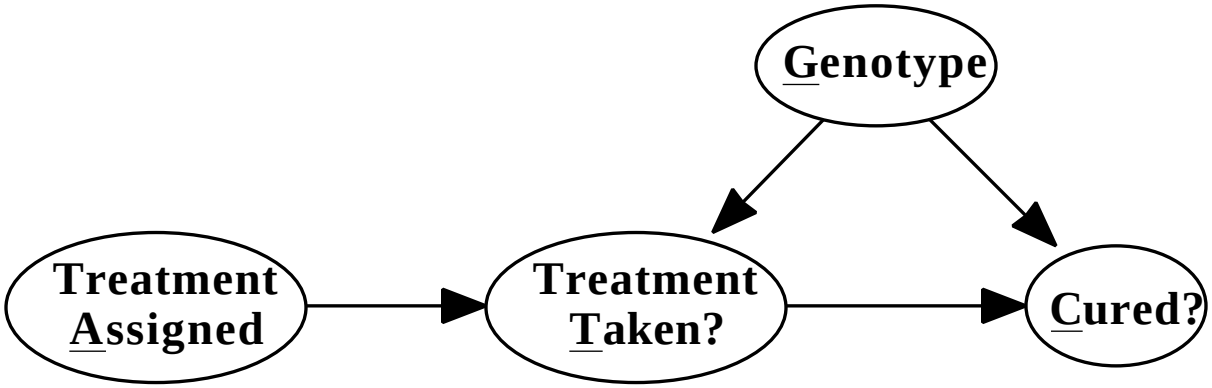
- **Effect of cause (or decision):**
  Suppose you are considering a set of acts/interventions and want to predict the effects or consequences of those acts.

- **Cause of effects:**
  Suppose you have the opportunity to learn something about the effects of one or more possible acts before you must commit to a particular act. This arises in troubleshooting situations (sequential diagnosis and repair), and, in general, whenever diagnostic information is gathered before taking action.

Passive Observations

Suppose you have no ability to affect any of the variables you are observing, but you want to perform inference based on those observations. This might arise when someone else is making the decisions.

This situation could be represented as

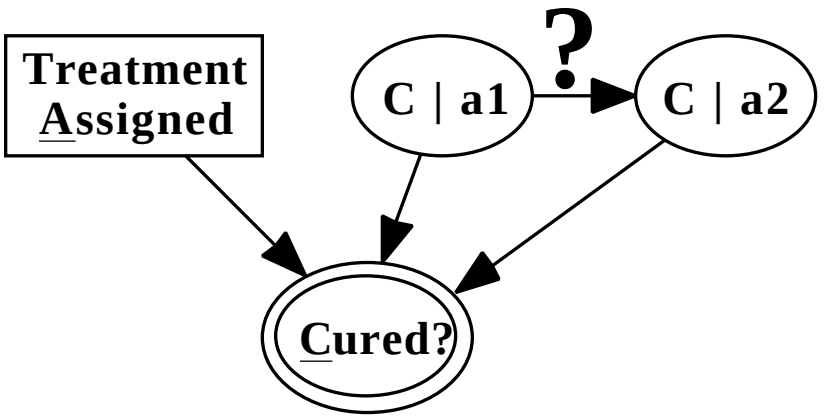


and all you need is the joint distribution and belief network.

Effect of Cause (or Decision)

Suppose you are considering a set of acts/interventions and want to predict the effects or consequences of those acts.

It is sufficient to have the marginal distributions of the relevant mapping variables:



This could be learned by conducting prospective randomized controlled trials, in which treatment assignment is independent of the features of the case. These are difficult to manage, and subject to numerous biases.

It could also be learned when the effect satisfies Pearl's identifiability condition, or when the Causal Markov Assumption is satisfied. In those cases, the causal graph and joint distribution would be sufficient.

Causal Implications of Conditional Independence

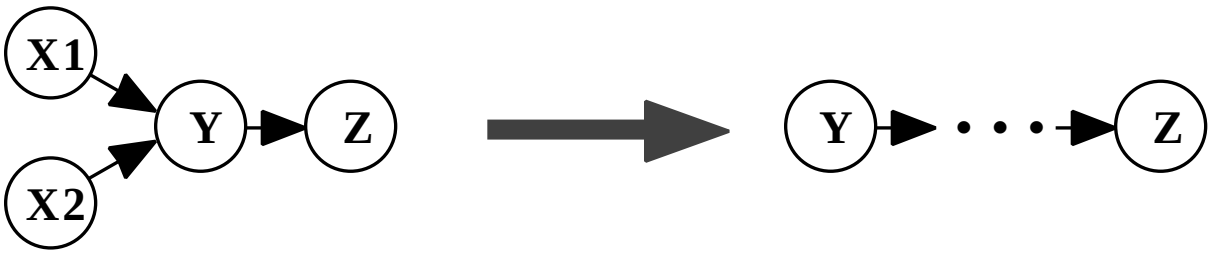
When you agree to the following three conditions:

- **All graphs for your domain are causal;**
- **Causal Markov assumption:**
  d-separation in a causal graph => independence
- **Faithfulness: independence => d-separation in a causal graph**

If the following conditions hold:

- **X1 is independent of X2**
- **X1 depends on X2 given Y**
- **Z depends on Y**
- **(X1, X2) is independent of Z given Y**

then in every causal graph containing X1, X2, Y, and Z, and any other variables, there is a directed path from Y to Z.



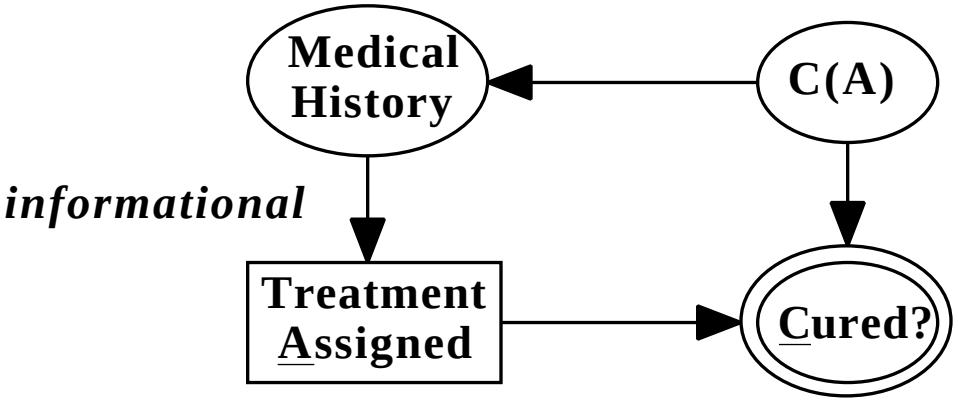
Therefore, Y causes Z.

Cause of Effect

Suppose you have the opportunity to learn something about the effects of one or more possible acts before you must commit to a particular act.

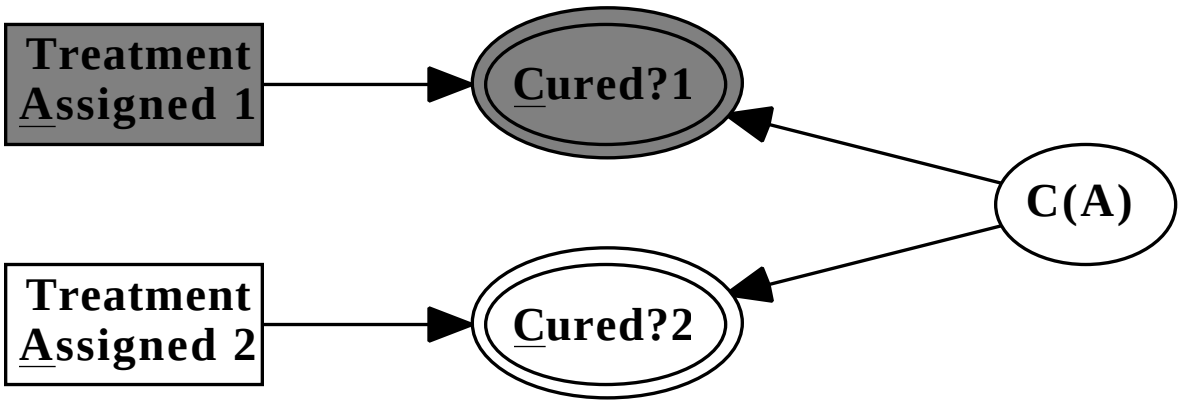
This can take several forms.

One is to take the best action in light of diagnostic information.



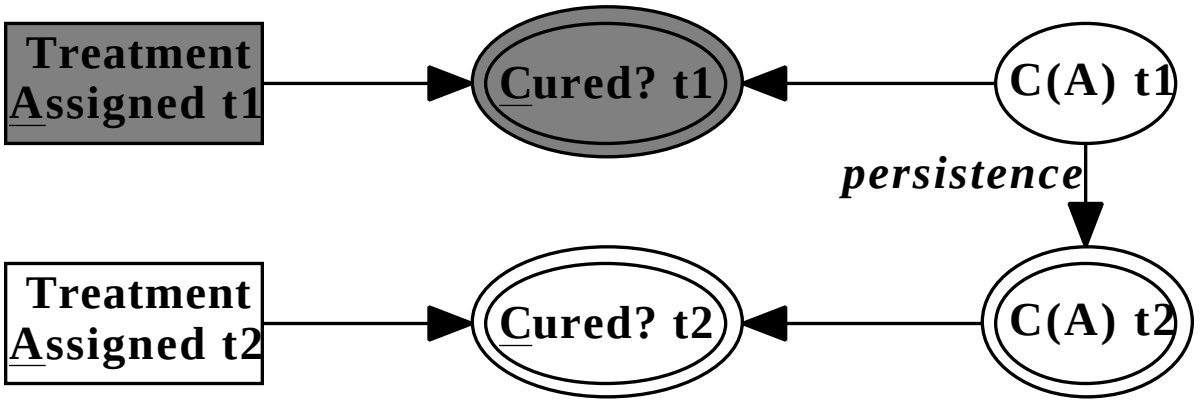
Cause of Effect (Continued)

Another situation involves counterfactual reasoning,
either prospectively ("If I were to observe this effect of my action what would happen if I instead were to . . . ?")
or retrospectively ("Given that I have observed this effect of my action what would have happened if I had instead . . . ?")



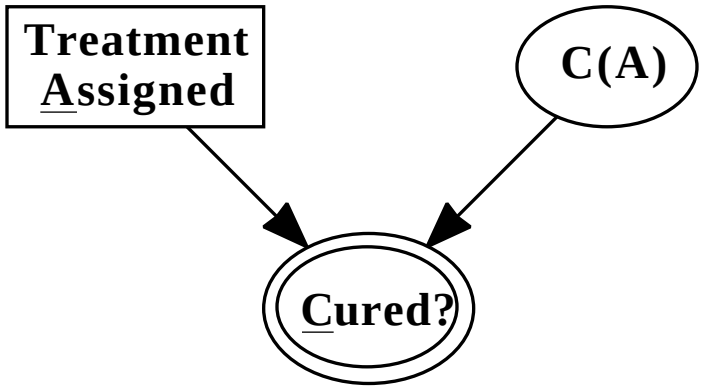
Cause of Effect (Continued)

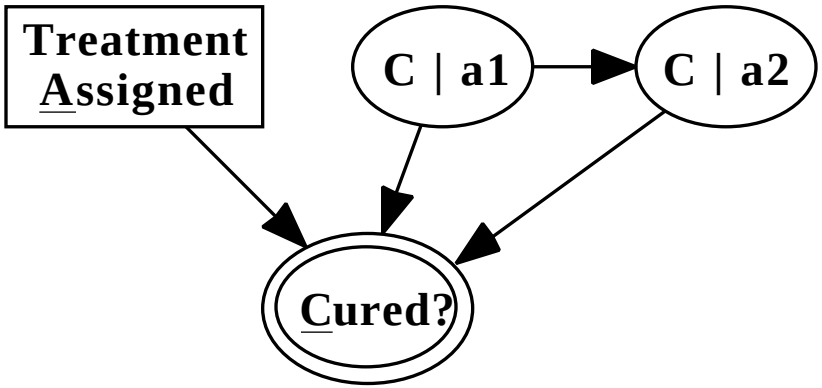
Finally, we can sequentially diagnose and repair a persistent system. This arises in troubleshooting situations (sequential diagnosis and repair).



Cause of Effect

To predict the effect of a decision after gathering information, you need the full distribution for the mapping variables.





This is just about impossible to acquire unless

- **You have a model for the dependency among mapping variables, such as independence.**

or

- **You can learn from a situtation with a reversible intervention-- a stable (persistent) system in which the interventions introduce changes within an acceptable range where the effects of actions are reversible and reproducable.**
  This might arise when designing hardware or software.

Task Summary

Tasks:	Passive Observation	Effect of Cause	Cause of Effect
Prior Knowledge			
Joint Distribution	√	no	no
+ Causal Graph	√	√	no
+ Causal Markov	√	√	no
Causal Graph with Marginal Distrib. for Mapping Vars.	√	√	no
Causal Graph with Full Joint Distrib. for Mapping Vars.	√	√	√
Data Source			
Passive Observation	√	sometimes	no
+ Causal Graph	√	sometimes	no
Random. Controlled Trials	√	√	no
Reversible Interventions	√	√	√

Summary

Depending upon the reasoning task, the model needs vary:

- **joint distribution of passive observations**
  is needed for passive inference from observations
- **marginal distributions of mapping variables, as obtainable from randomized controlled trials or, in some cases and under certain assumptions, passive observation accompanied by assertions of causal structure**
  are needed to predict the effects of actions
- **joint distributions of mapping variables, as obtained from reversible interventions**
  are needed to troubleshoot or act in light of information

Thus even though the "counterfactuals" (joint mapping variables) are always there in theory, we might be able to get by with much less to perform our reasoning tasks.

Savage's framework provides a strong foundation for causal model-building with clear subjective judgements.
This is a simpler approach than using functional equations and a causal theory as primitive elements.