When Causation Does Not Imply Correlation: Robust Violations Of The Faithfulness Axiom

Richard Kennaway

School of Computing Sciences, University of East Anglia

jrk@cmp.uea.ac.uk
http://www.cmp.uea.ac.uk/~jrk

Impetus for this paper

An apparent conflict between

- dynamical systems (especially control systems)
- causal analysis

Control systems characteristically create paradoxical patterns of correlation

- low or zero where there is a direct causal connection
- very high ($|c| \geq 0.95$) between some variables with only indirect causal connections, via zero-correlation links.

Classical causal analysis à la Pearl

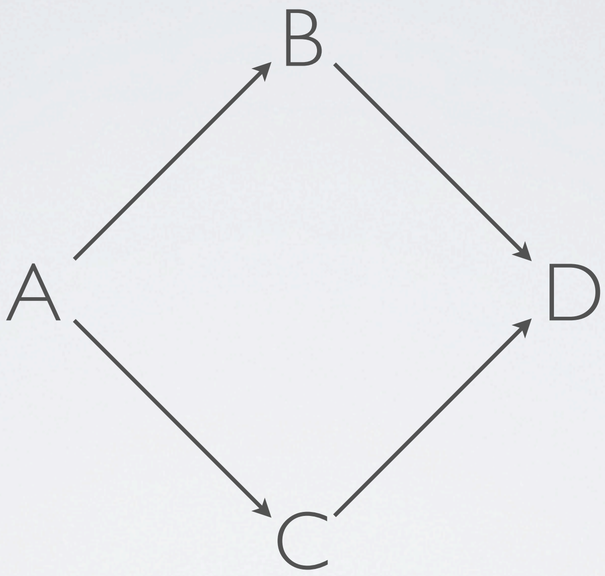
Pearl *Causality* (2000, 2009)
Spirtes, Glymour, Scheines *Causation, Prediction, and Search* (2001)

- The direct causal connections among the variables form a directed acyclic graph.

- The joint probability distribution factors as the product of the conditional distributions $P(X_i|\mathcal{X}_i)$ of each variable X_i given its immediate antecedents $\mathcal{X}_i$. (*Markov* assumption.)

- Every conditional correlation not implied by the Markov condition to be zero is non-zero. (*Faithfulness* assumption.)

Justification of the Faithfulness Assumption
"Causation implies correlation"



If the effects of A on D via B and via C exactly cancel out, then A and D could have zero correlation.

But this can only happen with probability zero.

Objections to faithfulness

The correlations, though non-zero, might be small, and the limited quantity of experimental data might fail to distinguish them from zero.

Answer: Get more data.

But here we will exhibit a large class of systems that show:

- zero correlations between directly connected variables
- very large correlations between indirectly connected variables, via zero-correlation links.
- stability — robust to changes in parameters

First counterexample

THEOREM. Suppose x is a variable dependent on time t.
If:

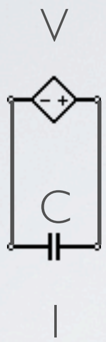
- x is differentiable
- x and dx/dt are bounded
- the correlation c(x,dx/dt) exists

then c(x,dx/dt) = 0, measured either over all time, or any finite interval where x takes the same value at both ends.

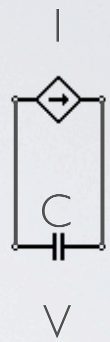
The proof is immediate from $\int x \, dx/dt \, dt = x^2/2$.

There's a corresponding result for discrete time series.

Concrete examples



Voltage V varied by the user causes current I



Current I varied by the user causes voltage V

For both circuits, $I = C\, dV/dt$

Correlation between V and I is zero.

Discrete version

Given a time series x_t, $t = \ldots -2, -1, 0, 1, 2, \ldots$

The two sequences:

- $(x_{t+1} + x_t)/2$
- $x_{t+1} - x_t$

have zero correlation.

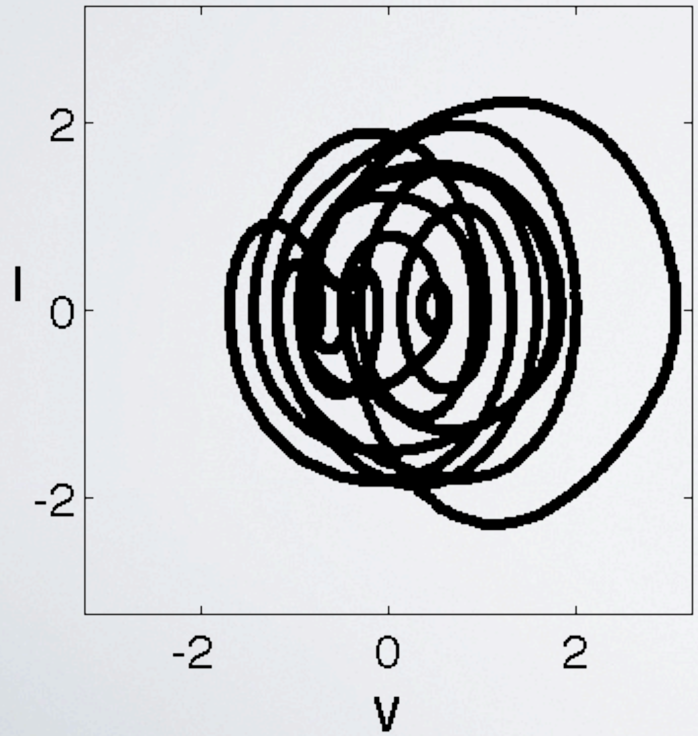
For example, monthly average bank balance and monthly change in bank balance.

If you're not getting richer or poorer, these have zero correlation.

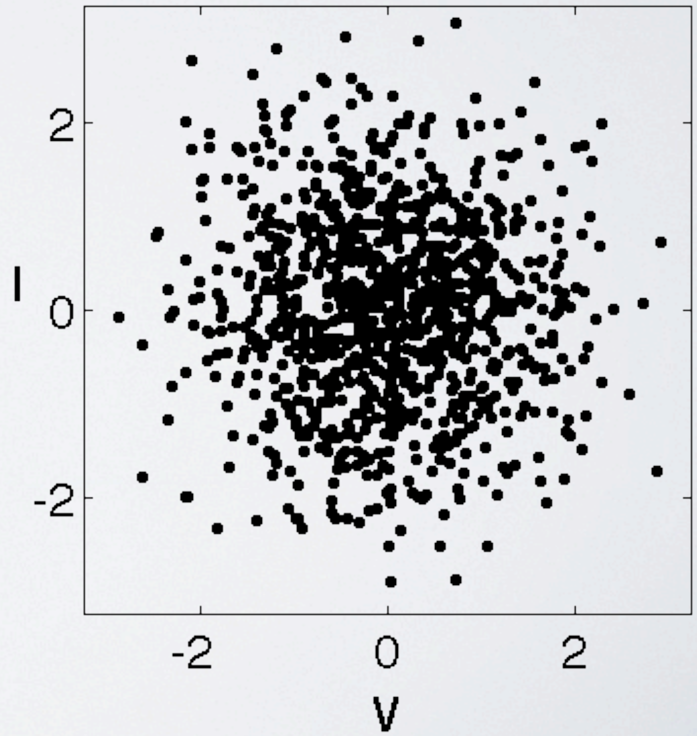
Robustness of the counterexamples

No tweaking of parameters will make these correlations non-zero.

Sampled on a fine timescale you can see trajectories.



Sampled on a coarse timescale you cannot.



What do methods of causal analysis have to say?

Pearl's book does not deal with dynamical systems or time series. Neither does the SGS book.

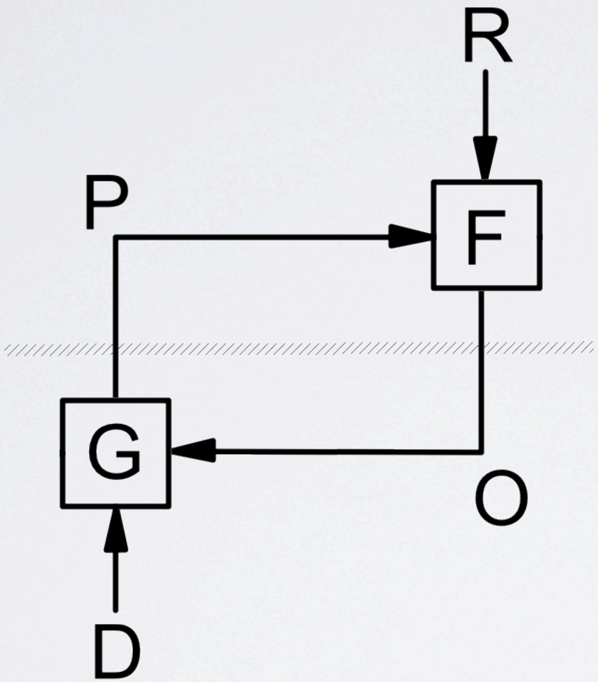
Some later work does. I'll come back to that after the second counterexample to faithfulness: control systems.

What is a control system?

A device for:

- maintaining some variable (the **perception**)
- at or close to some value (the **reference**)
- in spite of other influences on it (the **disturbances**)
- by generating some **output** that has an effect on the perception
- defined by a **control law** that determines the output from the perception and reference.

General form of all control systems



Controller

Environment

F is the control law.
G is the environment.

Examples of control systems

A room thermostat.

- Perception: actual temperature at the sensor
- Reference: value set on the dial
- Disturbances: external temperature, people in the room, etc.
- Output: turning the heater on and off
- Control law: turn the heater on when the room is too cold, off when it is too warm.

Examples of control systems

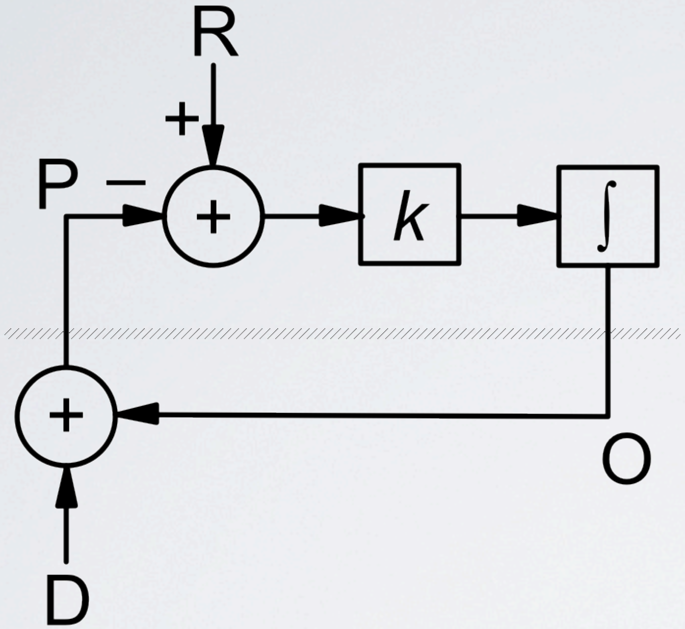
A cruise control.

- Perception: actual speed of car
- Reference: speed set by driver
- Disturbances: wind, gradient, tyre wear, etc.
- Output: flow of fuel to the engine
- Control law: usually a PID law:

  fuel flow rate =

  $$a(R-P) + b\int(R-P)dt + c\, d(R-P)/dt$$

A specific control system



$E = R - P$ (the **error**)

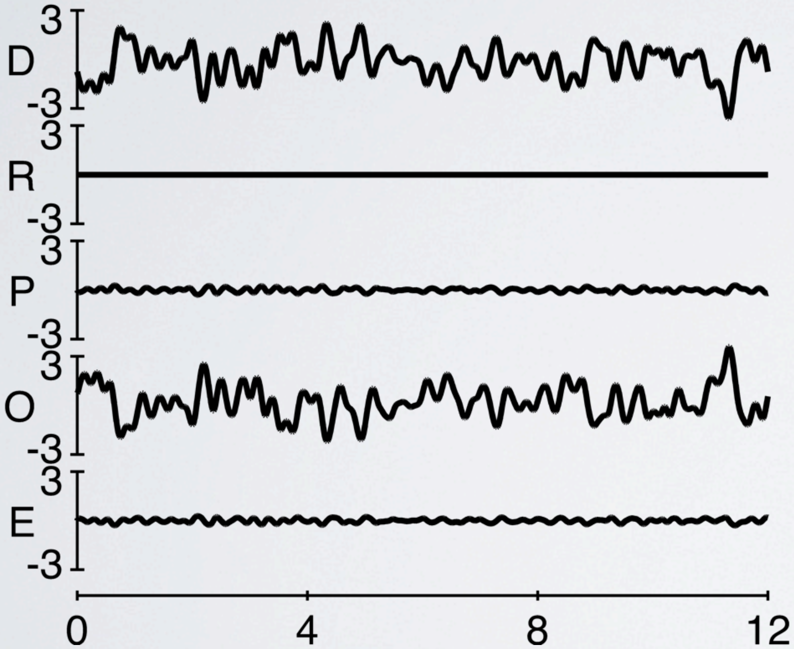
$O = k \int E \, dt$

$P = O + D$

What happens when D and R are random smooth waveforms?

Correlations among the variables

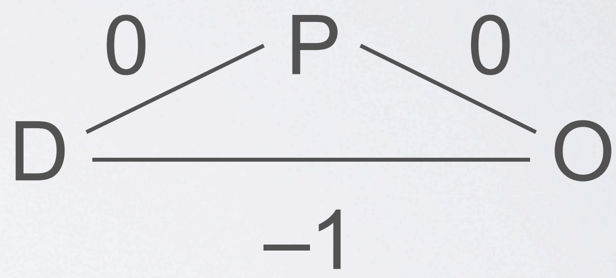
Random D, zero R



	O	P	D
O	1	0.002	-0.999
P		1	0.043
D			1

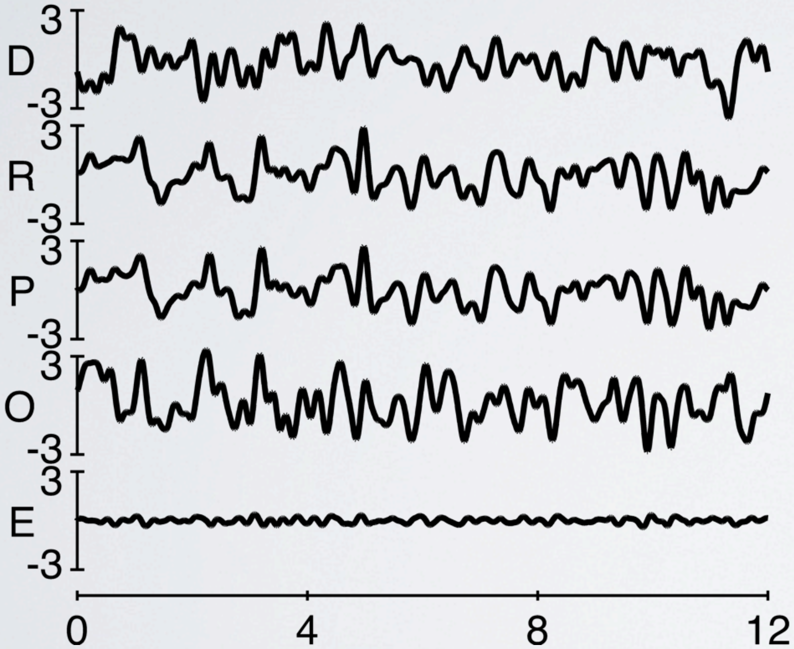
Causal relationships

Correlations



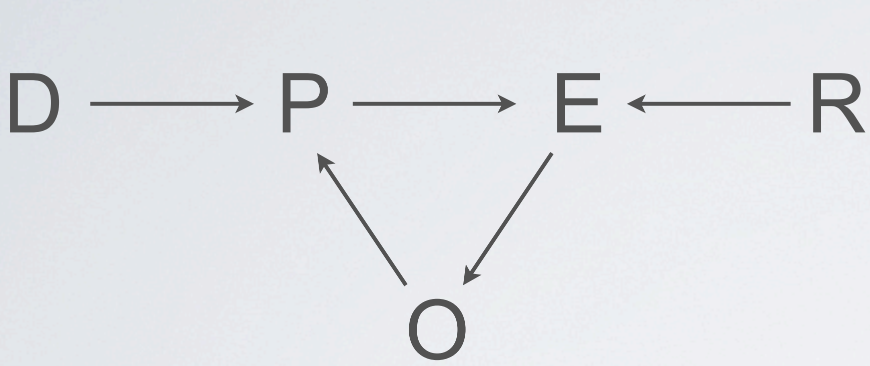
Correlations among the variables

Random and independent
D and R

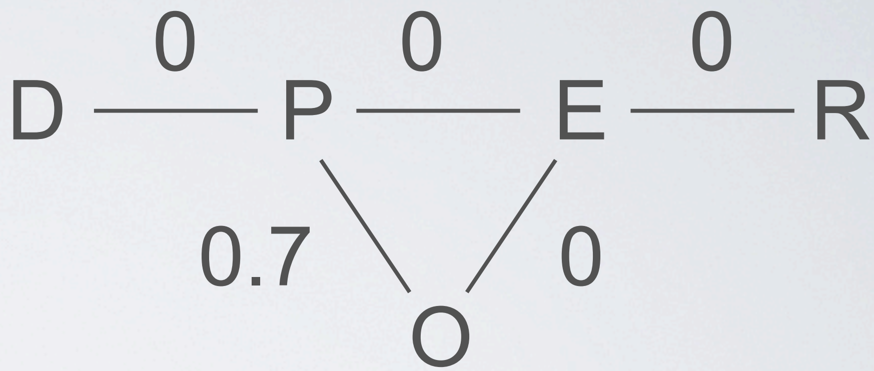


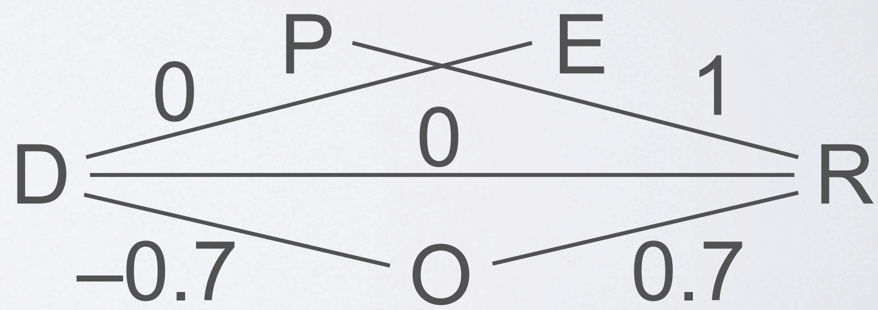
	O	P	R	E	D
O	1	0.718	0.717	-0.002	-0.718
P		1	0.998	-0.027	-0.031
R			1	0.039	-0.032
E				1	-0.024
D					1

Causal relationships



Correlations





Other correlations

Since P = O+D, the correlation between P and O+D is identically 1. If you discovered this correlation without knowing that P = O+D, it would look important.

But D is not easily measurable: it consists of *all* influences on P besides O.

Suppose $D = D_0 + D_1$, where you can measure D_0 but not D_1. D_1 is unmeasured exogenous noise.

What will be the correlation between P and $O+D_0$?

Effect of unmeasured disturbances

Assume:

- D_0 and D_1 are independent
- $D = D_0 + D_1$
- $var(D_1) = \mathbf{0.1}\ var(D)$

In other words, you can account for 90% of the variance of D.

	O	P = O+D	$O+D_0$	D_0
O	1	0.002	0.308	-0.947
P		**1**	**0.132**	0.042
$O+D_0$			1	0.012
D_0				1

A small amount of noise destroys most of the correlation.

What does causal analysis have to say about these systems?

Out of scope of Pearl and SGS, because:

- they don't handle dynamical systems or time series
- they assume acyclic causal relationships — but control systems are always cyclic.

Other people have tried to extend their methods in both directions.

Lacerda et al. (2008)
Discovering cyclic causal models by independent components analysis

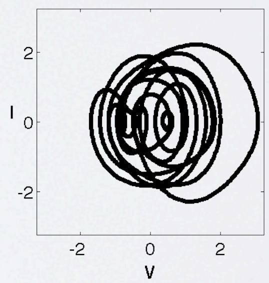
Allow time series with cyclic dependencies:

$$x_i(t+1) = \sum_j a_j x_j(t) + \text{noise}$$

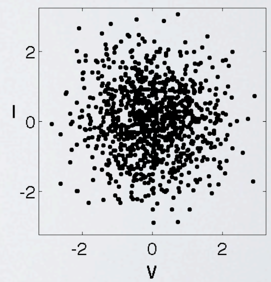
provided that in this equation $a_i \neq 1$.

But this rules out any relationship $x = \int y \, dt$.
Discrete time version is $x(t+1) = x(t) + y(t)$

They also suggest sampling on a timescale *longer* than the settling time. This rules out the possibility of seeing informative trajectories.



shorter



longer

“Equilibration” is known to be problematic

Iwasaki & Simon (1994) *Causality and model abstraction*

Recognises that causal relationships in dynamical systems can appear different, depending on the timescale of measurement relative to the timescale of equilibration processes in the system, but does not grapple with the particular difficulties posed by control systems.

More on equilibration

Dash 2003, 2005: considers interaction between Pearl's *do* operator and equilibration. Recommends, in effect, sampling the system on a timescale shorter than its equilibration time. Cf. engineering technique of hitting the system with a hammer (more technically, applying a spike or step function to an input and observing the transient response).

But this does not necessarily help if you're still looking at correlations — they can still be paradoxical.

Correlations for fast inputs

Slow disturbance
(as before)

	O	P	D
O	1	0	–1
P		1	0
D			1

Fast disturbance

	O	P	D
O	1	0	0
P		1	1
D			1

Correlations for fast inputs

Slow disturbance
and reference
(as before)

	O	P	R	E	D
O	1	0.7	0.7	0	-0.7
P		1	1	0	0
R			1	0	0
E				1	0
D					1

Fast disturbance
and reference

	O	P	R	E	D
O	1	0	0	0	0
P		1	0	-0.7	1
R			1	0.7	0
E				1	-0.7
D					1

In all cases, there are still zero correlations associated with direct causal connections.

Itani et al. (2008)
Structure learning in causal cyclic networks

Generalises the Markov condition to cyclic causal graphs, by defining a concept of *consistency* of a total distribution with the per-node conditional distributions.

Imposes a condition that there is a unique total distribution having the consistency property.

Uniqueness fails for the control systems shown here.

Why are control systems so problematic for causal analysis?

The whole purpose of a control system is to actively destroy the very information that current causal analysis methods depend on.

They maintain a variable at a specified level, *regardless of the other causal influences on that variable*.

How important is this?
When will you encounter control systems?

Living systems are full of control systems. They have to be, to be alive, i.e. to actively preserve their form in spite of environmental influences.

In the biological and social sciences, the default assumption has to be that control systems are present. Methods of data analysis must take this into account.

What is to be done?

Interventional experiments can elucidate the true causal structure.

But only if correctly performed.

What does *do*(P=p) mean, when P is the controlled variable of a control system? It means acting with enough force to override the system's own attempts to keep P equal to R.

If you succeed, you have driven the system into an abnormal operating regime. What you observe may not tell you anything about its normal functioning.

Testing for a controlled variable

When you intervene on a controlled variable P, you are really applying a disturbance: introducing a new variable X and a causal arrow from X to P.

If the disturbance is of a size and speed that the system can control against, P will hardly change. That is a sign that P is under control. If something else changes instead, that is a sign that it may be the output of that control system.

This is called the Test for the Controlled Variable (Powers 1974, 1998).

The Test for a Controlled Variable

If you know that there is a causal effect of A upon B,
but however you set A, B does not change,
then B is a controlled variable.

Some other variable C must be having an effect on B that cancels out the effect of A.

C must result from B and some reference value D.

Simple example of the TCV

Place a candle near the room thermostat.

The temperature of the thermostat does not change, even though placing the candle near anything else would warm it up.

The rate of energy supply to the room drops. The rest of the room gets colder.

Research question

Vast quantities of data are being gathered and subjected to various sorts of causal analysis:

- Gene expression arrays can test the levels of thousands of genes at once.
- Real-time neuroimaging techniques.

In practice, are these experiments vulnerable to the issues presented here?

Can the TCV be scaled up and applied to such data?