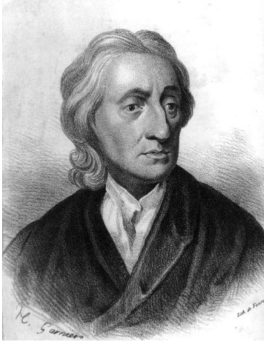
Causality and directed acyclic graphs

Get to know and have some intuition about

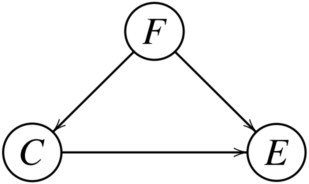
- Causality in philosophy (of science)
- Conditional independence relations
- Causal discovery algorithms
- Confounds and back doors



David Hume (1711-1776)

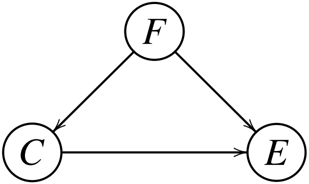
Confounds and back-doors

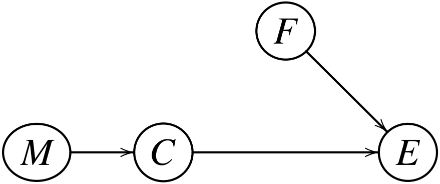
How can you be sure to infer a causal effect from C to E when it is possible that there is another variable that causes the relation between C and E?



Confounds and back-doors

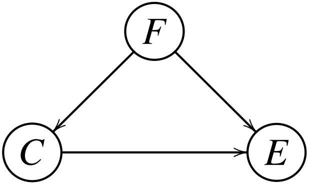
How can you be sure to infer a causal effect from C to E when it is possible that there is another variable that causes the relation between C and E?

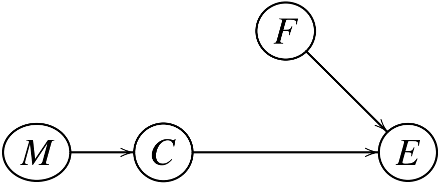




Confounds and back-doors

How can you be sure to infer a causal effect from C to E when it is possible that there is another variable that causes the relation between C and E?





In experimental studies (randomization) there are no confounds because all influence of other variables is non-systematic.

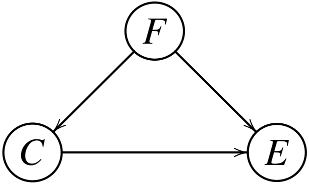
Confounds and interventions

No-confounding, causal definition

Denote by $P(y \mid do(x))$ the probability of $Y = y$ under the hypothetical intervention $X = x$. Then X and Y are not confounded if and only if

$$P(y \mid do(x)) = P(y \mid x)$$

for all x and y, and $P(y \mid x)$ is the conditional probability.



Confounds and back-doors

Combined	E	$\neg E$		Recovery rate
Drug (C)	20	20	40	50%
No drug ($\neg C$)	16	24	40	40%
	36	44	80	

$P(E \mid C) > P(E \mid \neg C)$

Males	E	$\neg E$		Recovery rate
Drug (C)	18	12	30	60%
No drug ($\neg C$)	7	3	10	70%
	25	15	40	

$P(E \mid C) < P(E \mid \neg C)$

Females	E	$\neg E$		Recovery rate
Drug (C)	2	8	10	20%
No drug ($\neg C$)	9	21	30	30%
	11	29	40	

$P(E \mid C) < P(E \mid \neg C)$

Confounds and back-doors

Simpson's Paradox is when you find in this example:

Confounds and back-doors

Simpson's Paradox is when you find in this example:

1. Drug works for the whole population: $P(E \mid C) > P(E \mid \neg C)$

Confounds and back-doors

Simpson's Paradox is when you find in this example:

1. Drug works for the whole population:
   $P(E \mid C) > P(E \mid \neg C)$
2. Drug does not work for females:
   $P(E \mid C, F) < P(E \mid \neg C, F)$

Confounds and back-doors

Simpson's Paradox is when you find in this example:

1. Drug works for the whole population:
   $P(E \mid C) > P(E \mid \neg C)$
2. Drug does not work for females:
   $P(E \mid C, F) < P(E \mid \neg C, F)$
3. Drug does not work for males:
   $P(E \mid C, \neg F) < P(E \mid \neg C, \neg F)$

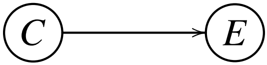
Confounds and back-doors

Simpson's Paradox is when you find in this example:

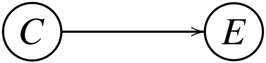
1. Drug works for the whole population:
   $P(E \mid C) > P(E \mid \neg C)$
2. Drug does not work for females:
   $P(E \mid C, F) < P(E \mid \neg C, F)$
3. Drug does not work for males:
   $P(E \mid C, \neg F) < P(E \mid \neg C, \neg F)$

It is unclear which table to use: the combined table or the two separate tables

Confounds and back-doors

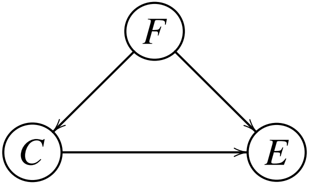


Confounds and back-doors



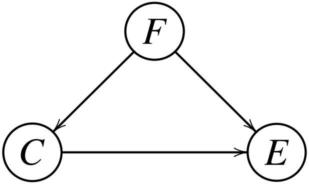
1. Infer the effect of C (drug) on E (recovery)

Confounds and back-doors



1. Infer the effect of C (drug) on E (recovery)
2. F (gender) influences both (is a common cause of) C and E

Confounds and back-doors



1. Infer the effect of C (drug) on E (recovery)
2. F (gender) influences both (is a common cause of) C and E

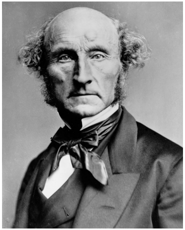
F (gender) is a confound!

Causal relation and laws

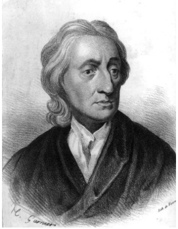
John Stuart Mill's conditions

- A always co-occurs with B
- A occurs before B
- There is no alternative explanation for the co-occurrence of A and B

But this was unsatisfactory because A does not always occur with B; no universality.

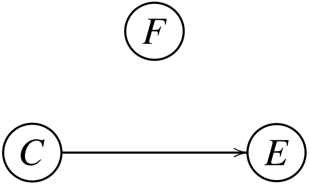


John Stuart Mill (1806-1973)

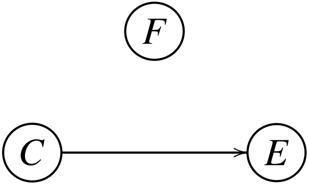


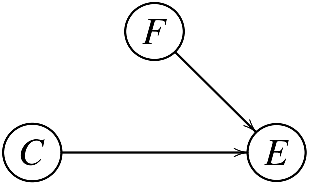
David Hume (1711-1776)

Interventions and invariance



Interventions and invariance

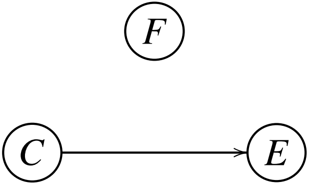


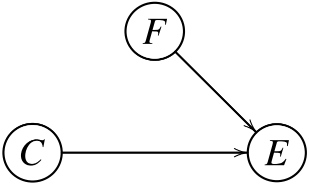


Interventions and invariance

- What if we only used C to determine E?

Interventions and invariance





- What if we only used C to determine E?
- What if we used C and F to determine E?

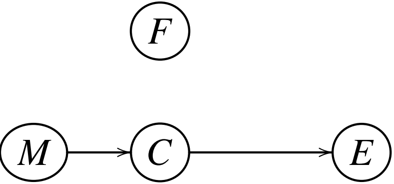
Interventions and invariance

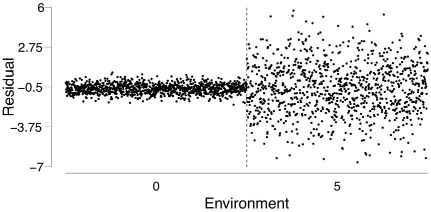
- What if we only used C to determine E?
- What if we used C and F to determine E?
- Would the residuals of E (after regressions) be the same in both cases?

Interventions and invariance

- What if we only used C to determine E?
- What if we used C and F to determine E?
- Would the residuals of E (after regressions) be the same in both cases?
- No! So we can check whether for some set of nodes the residuals are the same for different subsets *and manipulations*

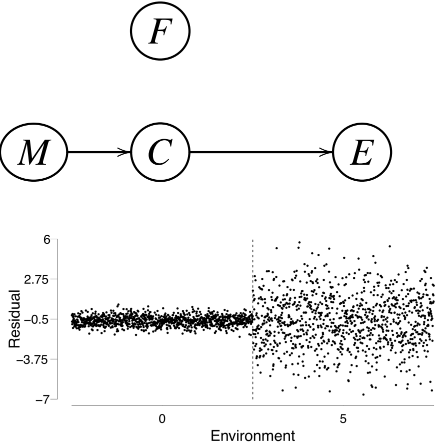
Interventions and invariance

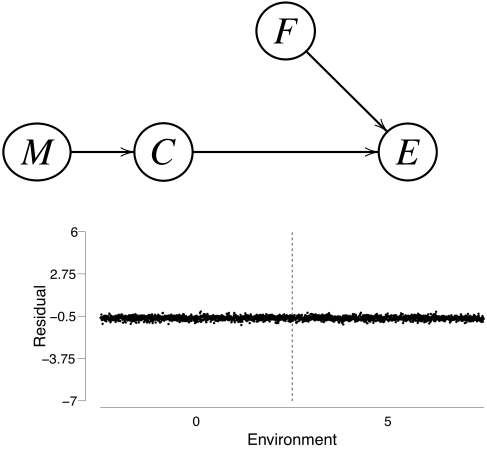




Kossakowski, Maas, Waldorp (2020, in revision)

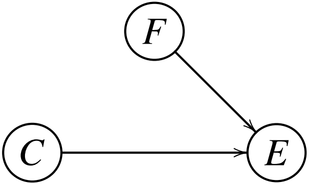
Interventions and invariance



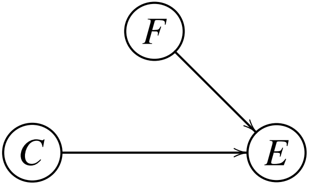


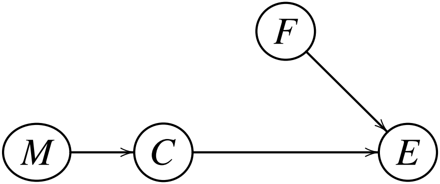
Kossakowski, Maas, Waldorp (2020, in revision)

Interventions and invariance

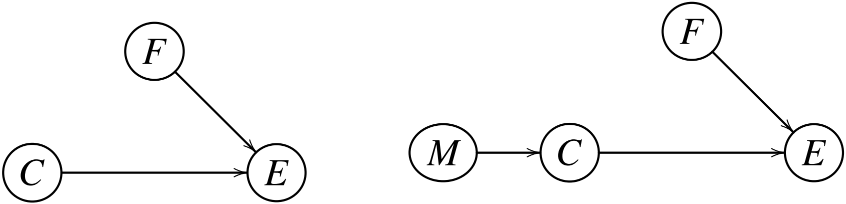


Interventions and invariance





Interventions and invariance



Whether we control C using M or not, if we have both F and C and the residuals remain the same, then we have the correct set of direct causes.