Causal Inference: prediction, explanation, and intervention

Lecture 3: Probabilistic causality

Samantha Kleinberg
samantha.kleinberg@stevens.edu

Why probabilistic causality?

- Non-deterministic relationships
  - Radioactive decay
- Incomplete knowledge
  - Even if food poisoning were deterministic, we still would not know all the conditions needed to bring it about
- Capture ***how much*** of a difference cause makes to effect
  - Difference between oxygen and lit match for house fire

Quick probability review

- Sample space: set of all possible outcomes
- Event: subset of the sample space

- Flipping a coin twice in row
  - $\Omega = \{HH,HT,TH,TT\}$
  - Event of getting exactly one tail $\{HT, TH\}$

Probability function

- For A being subsets of Ω
  - P(A)∈[0,1], for A in set of events
  - P(Ω)=1
  - For disjoint events A,B: P(A ∨ B) = P(A) + P(B)

- Ω = {HH,HT,TH,TT}

- P any combination of heads/tails = 1

- P(HH ∨ TT) = 1/2

Probabilities from data

How to determine probability of heads/tails from coin flip?

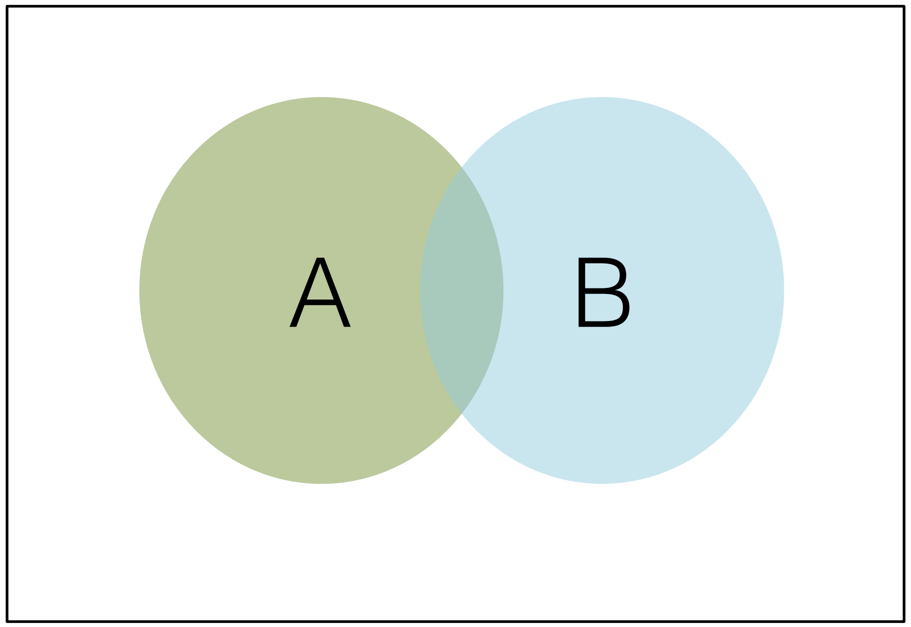
- Frequency of event x
  - #(x) = number of times it is observed
- Relative frequency
  - #(x)/N, where N is number of observations
- As N→∞, relative frequency →P(x)

Probabilities from data - challenges

- Sampling bias
  - E.g. survey FOX news or MSNBC viewers about president
- Confirmation bias
  - Ignoring some events
  - Coding events differently

Disjunction

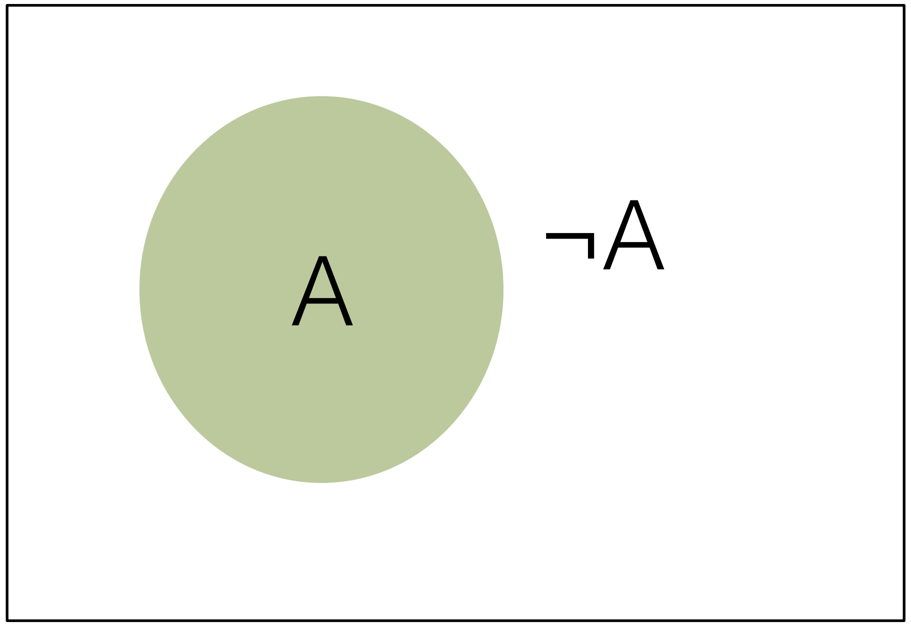
- Addition rule
  - P(A∨B)=P(A)+P(B)-P(A^B)
- Mutually exclusive events
  - P(A∨B)=P(A)+P(B)
  - Flipping H∨T versus increasing unemployment or decreasing interest rates



Complement rule

$1 = P(A) + P(\neg A)$

$P(\neg A) = 1 - P(A)$



Dependence/independence

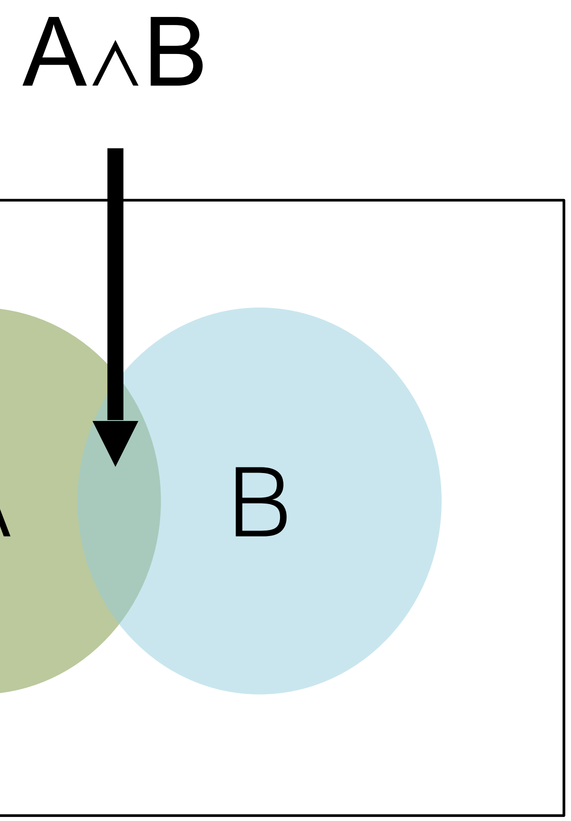
- Events that are independent
  - Flipping heads and then tails
  - Day of week and whether a patient had a heart attack (probably?)

- Events that are dependent
  - Vice presidential candidate and presidential nominee
  - Diagnostic test being positive and whether patient has a disease

Conjunction

- For independent events

  $P(A \wedge B) = P(A)P(B)$



Conditional probability

$P(A \wedge B) = P(A|B)P(B)$, where $P(B) > 0$

Note, if A and B independent, $P(A \wedge B) = P(A)P(B)$

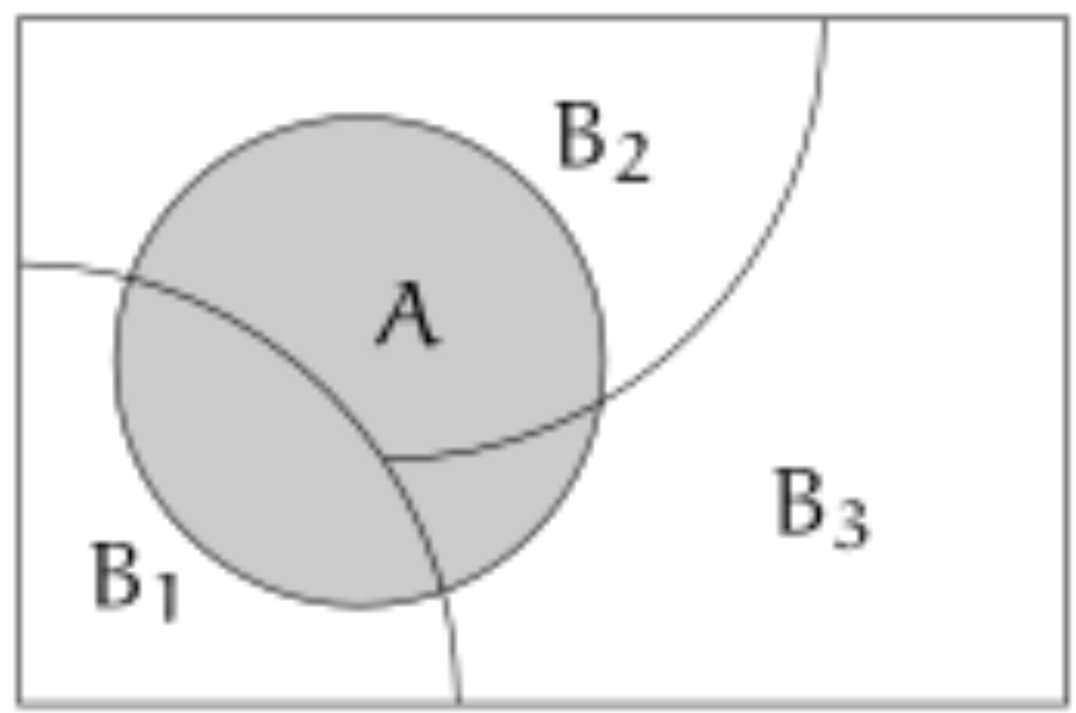
Then: $$P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$$

Law of total probability

$$P(A) = P(A \wedge B_1) + P(A \wedge B_2) + \ldots + P(A \wedge B_3)$$

$$= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_n)P(B_n)$$

$$= \sum_{i=1}^{n} P(A|B_i)P(B_i)$$



Also called marginal probability, and this process is marginalization

Bayes rule

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

$$P(A \wedge B) = \boxed{P(A|B)P(B) = P(B|A)P(A)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Example of Bayes rule

- Probability of disease (d) is 0.001
- But diagnostic test not foolproof
  - $P(+|d) = 0.98$
  - $P(+|\neg d) = 0.1$

Given a positive test, what's probability patient has disease?

Example continued

$$P(d \mid +) = \frac{P(+ \mid d)P(d)}{p(+)}$$

$$P(\neg d) = 1 - P(d)$$

$$= \frac{P(+ \mid d)P(d)}{P(+ \mid d)P(d) + P(+ \mid \neg d)P(\neg d)}$$

$$= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.1 \times (1 - 0.001)}$$

$$= \frac{0.00098}{0.10088} \approx 0.01$$

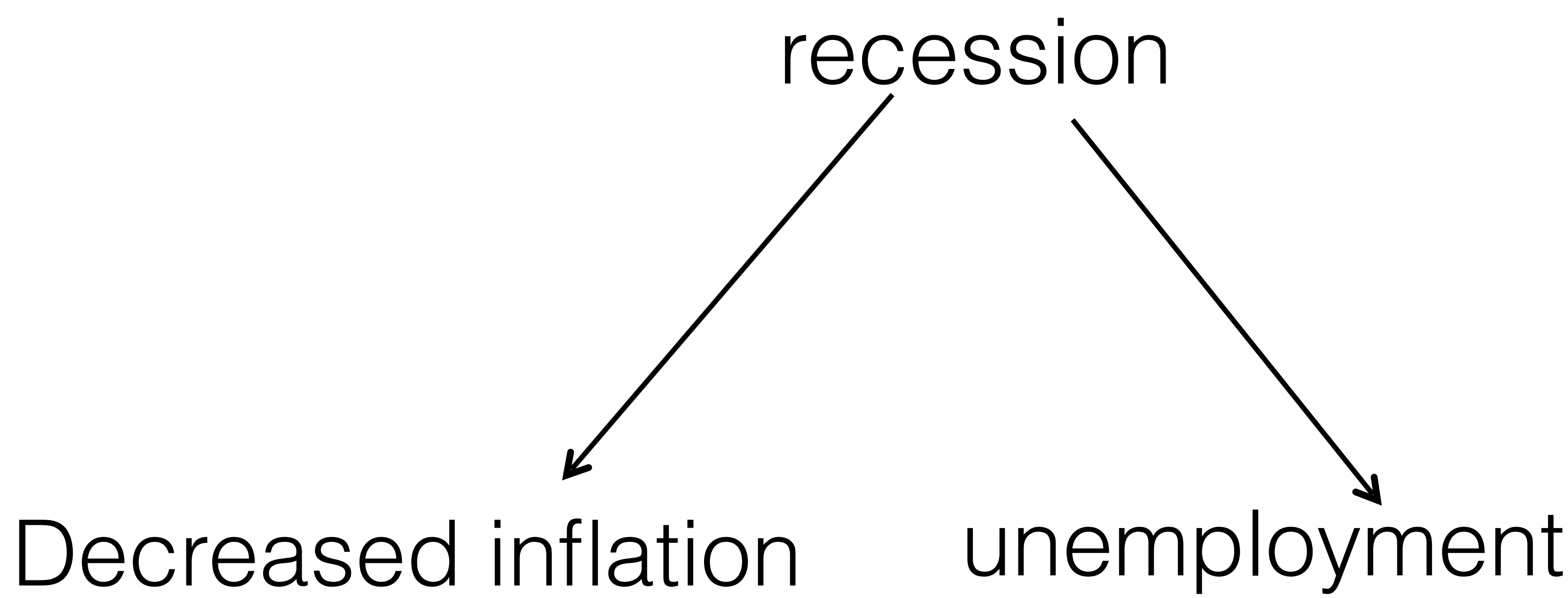
Probabilistic causality: basic idea

C causes E if

$P(E|C)>P(E)$

Or $P(E|C)>P(E|\neg C)$

Probability raising and common causes



Screening off

How to weed out non-causes that raise probability

- Yellowed fingers and lung cancer
- Falling barometer and rain

C *screens off* A from B if

$P(B|A \wedge C)=P(B|C)$

Screening off and causal chains

Disease →medication →side effect

$P(S|M,D)=P(S|M)$

Common cause principle

C is common cause of A and B if C is earlier than both and:

1. $P(A \wedge B|C)=P(A|C)P(B|C)$
2. $P(A \wedge B|\neg C)=P(A|\neg C)P(B|\neg C)$
3. $P(A|C)>P(A|\neg C)$
4. $P(B|C)>P(B|\neg C)$

Reichenbach

If C and E are dependent either C causes E, E causes C or they have a common cause (that is earlier than both) that screens them off from one another

Problem: no single common cause

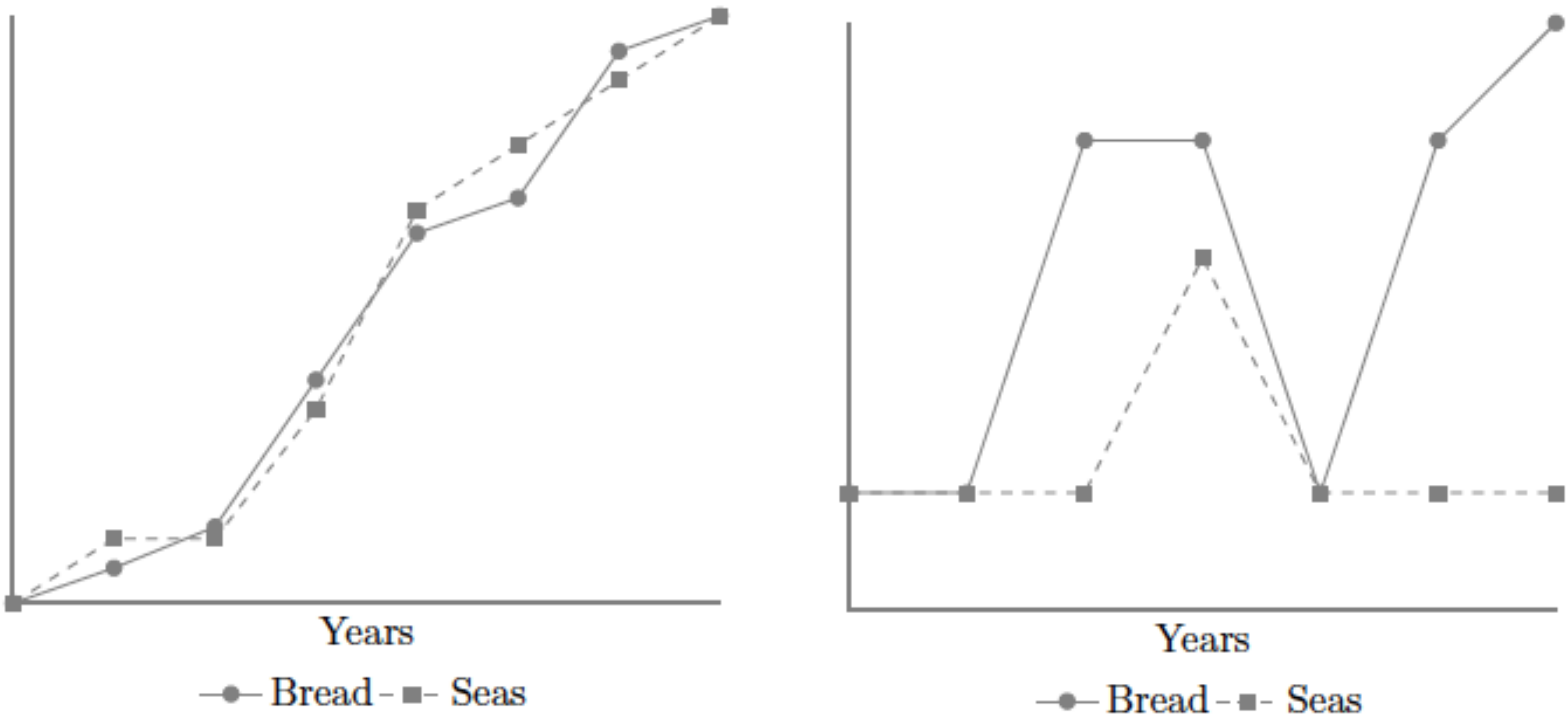
Alice and Bob both like to take AI (i) classes that meet in the afternoon (t), so they're likely to be in the same classes

$P(A|B)>P(A)$

$P(A|B \wedge i)>P(A|i)$

$P(A|B \wedge t)>P(A|t)$

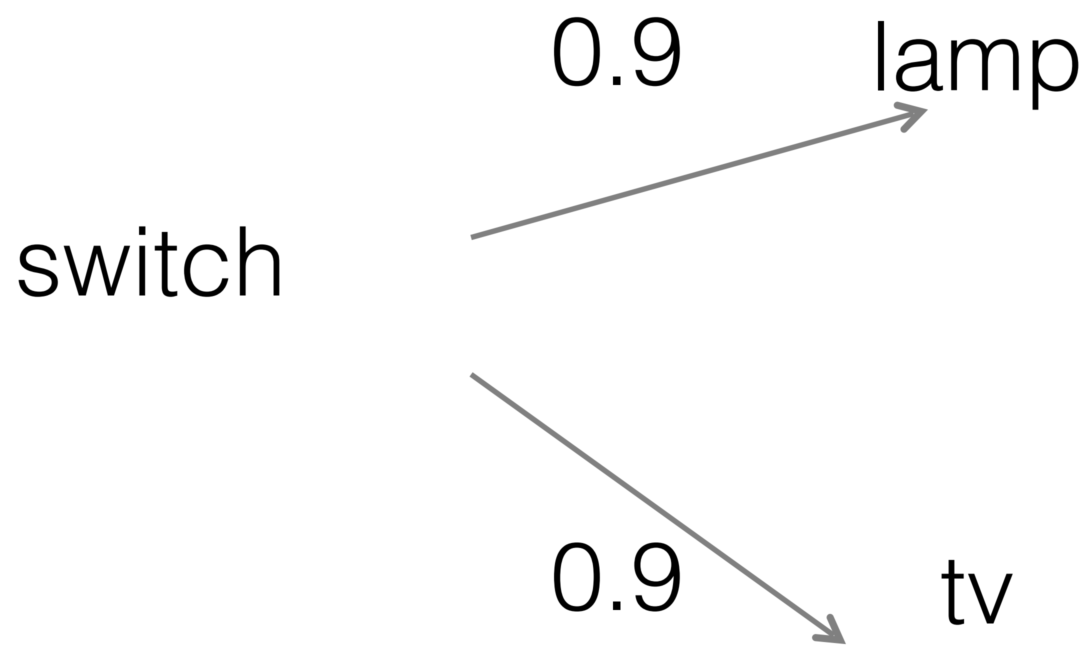
Problem: Nonstationarity



Sober, E. (2001). Venetian sea levels, British bread prices, and the principle of the common cause. *The British Journal for the Philosophy of Science*, *52*(2), 331

Problem: Indeterminism

If lamp on, tv also on, so $P(L|T)=1$



Does switch screen off L and T?

Suppes

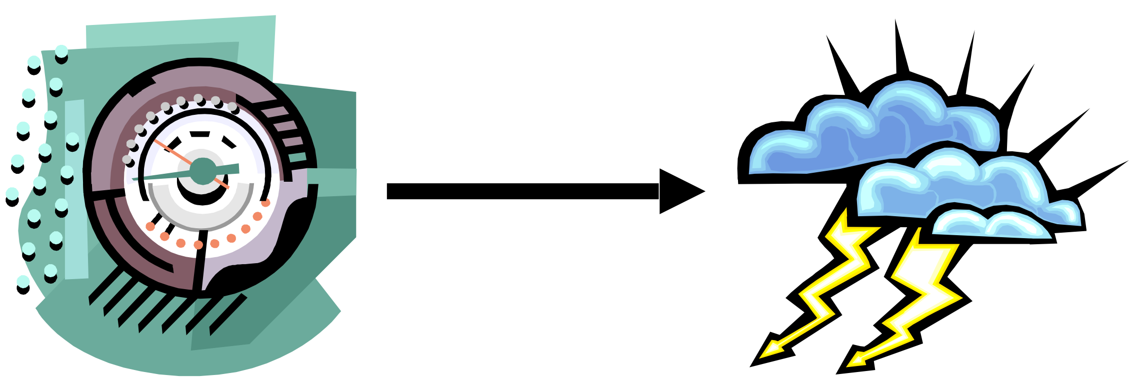
- Causes raise probability of their effects
- Causes are temporally prior to their effects

How can we identify spurious causes?

Prima facie causes

C is a **prima facie cause** of E iff:

- C is earlier than E
- $P(C) > 0$
- $P(E|C) > P(E)$



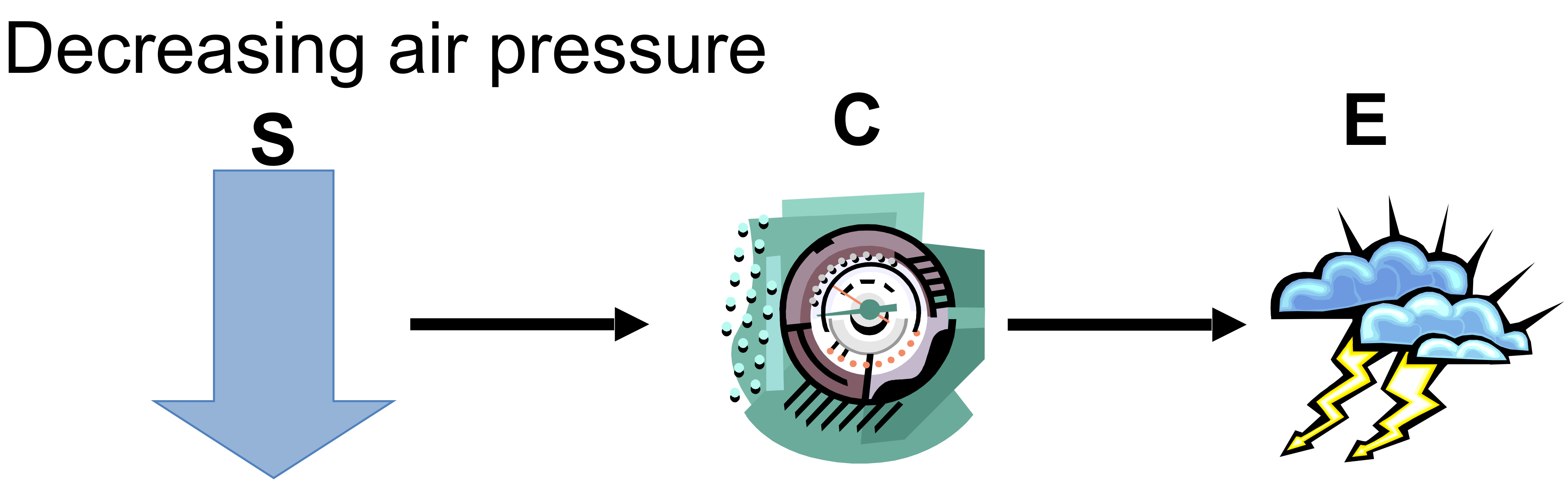
Spurious causes

- Many non-causal correlations
- Main idea: finding things earlier than prima facie cause that account equally well for effect

Spurious

C is **a spurious cause** of E iff:

- C is prima facie cause of E
- Exists S earlier than C such that:
  - $P(C,S) > 0$
  - $P(E|C,S)=P(E|S)$
  - $P(E|C,S) \geq P(E|C)$



Why is this insufficient?

Cause is spurious if

1. We have an event that eliminates cause's effectiveness for predicting the effect
2. S is at least as good a predictor of effect as cause alone

Genuine causes

- A prima facie cause that is not spurious is genuine

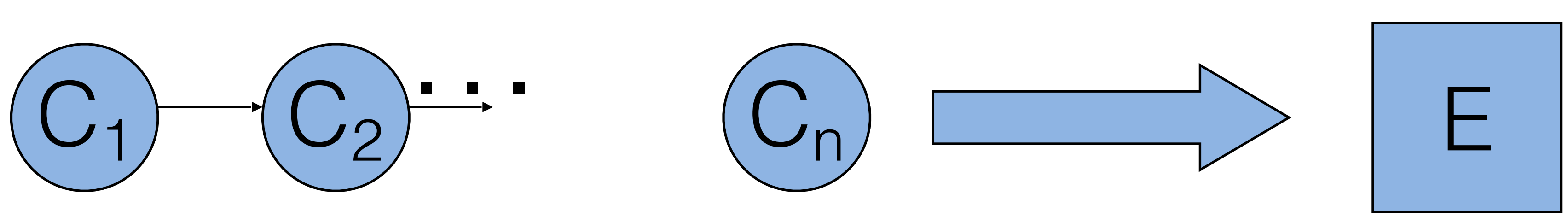
ε-spurious

- Before, $P(A|B,C)=P(A|C)$
- What if not exactly =, but still small effect?
- B is ε-spurious cause of A iff:
  - Same conditions as before, but:
  - $|P(A|B,C)-P(A|C)| < \varepsilon$

Difficulties with probabilistic causality?

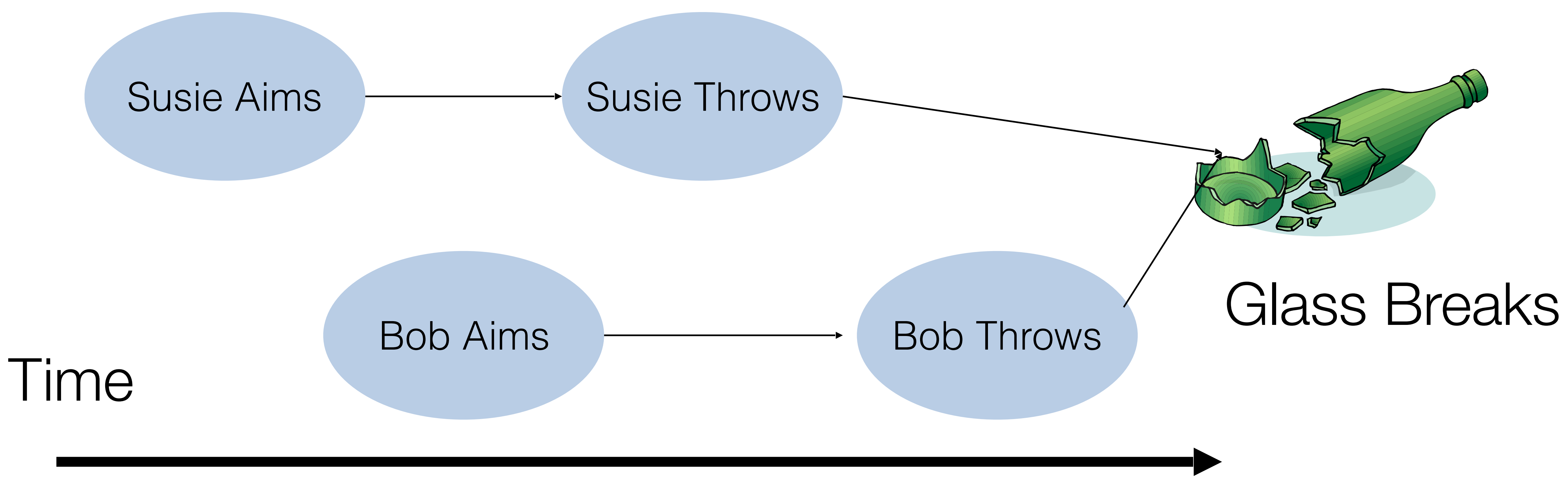
Problem: Causal chain

- Each element in chain produces next with P=1
- What happens?
- Each element spurious aside from first – always an earlier cause to account for effect



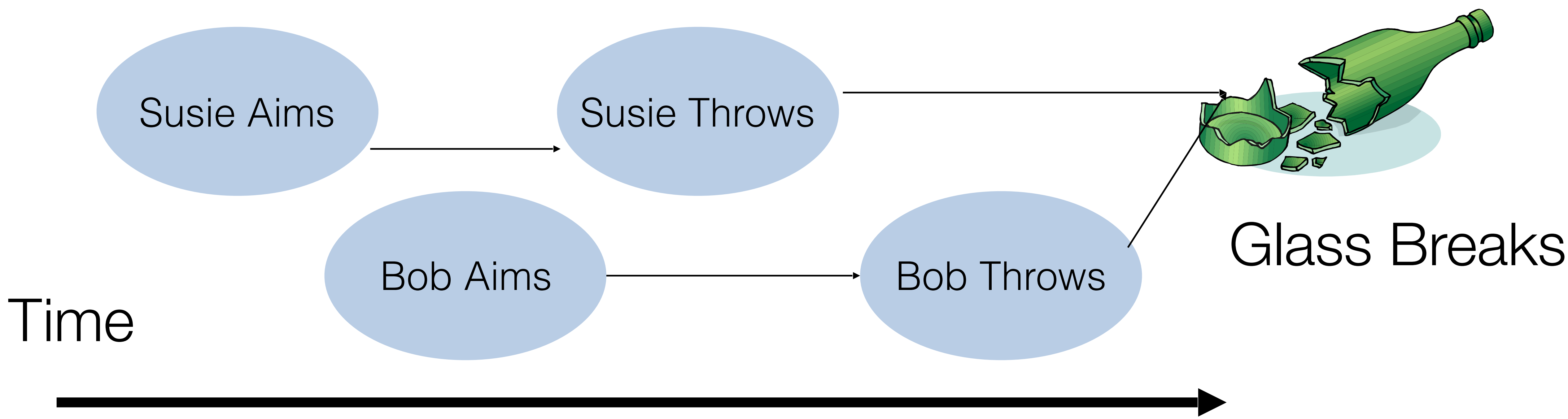
Problem: Overdetermination

- Bob and Susie throwing rocks at glass bottle
- Rocks hit simultaneously, but Susie threw first

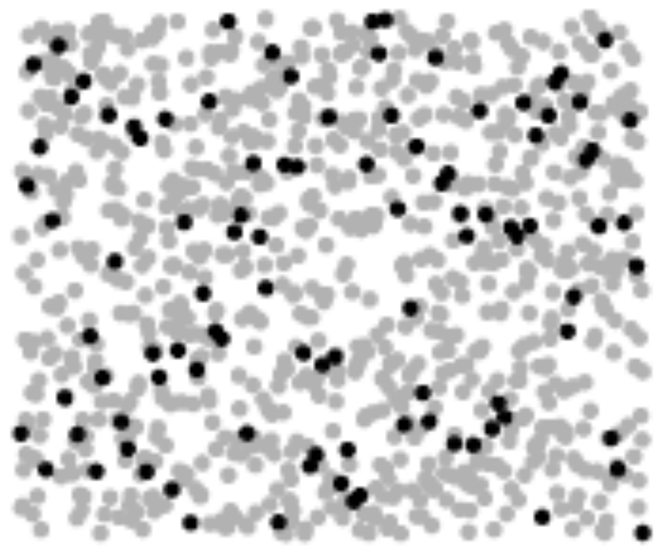


Problem: Preemption

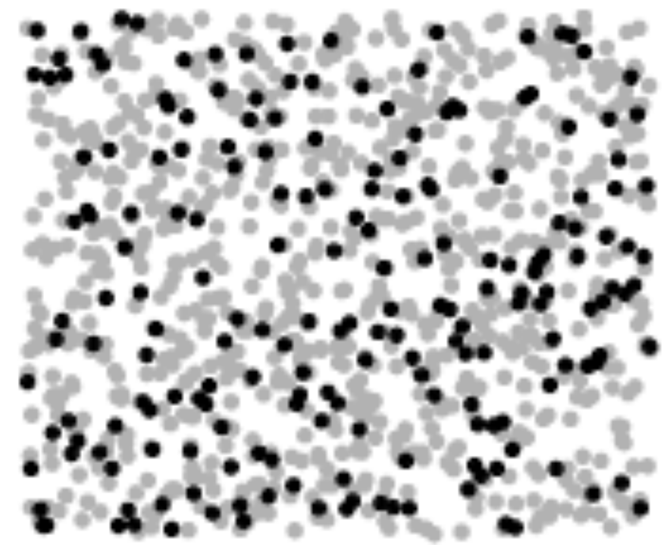
- Bob's rock arrives first and breaks glass
- But Susie still threw first and has perfect aim (breaks glass with P=1)



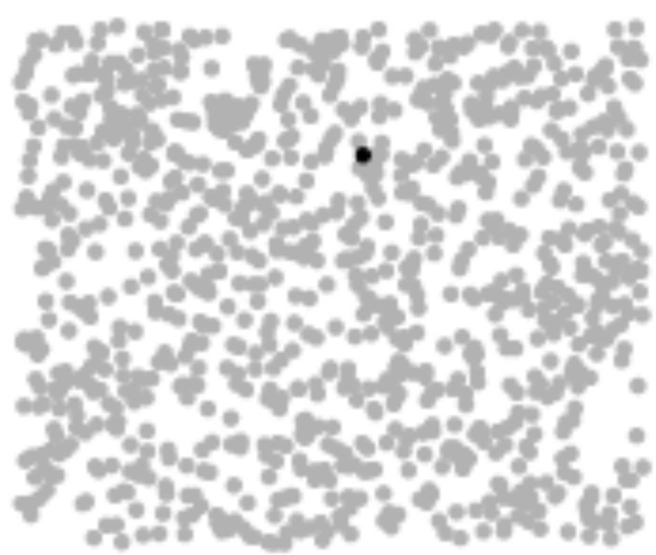
Problem: effect size



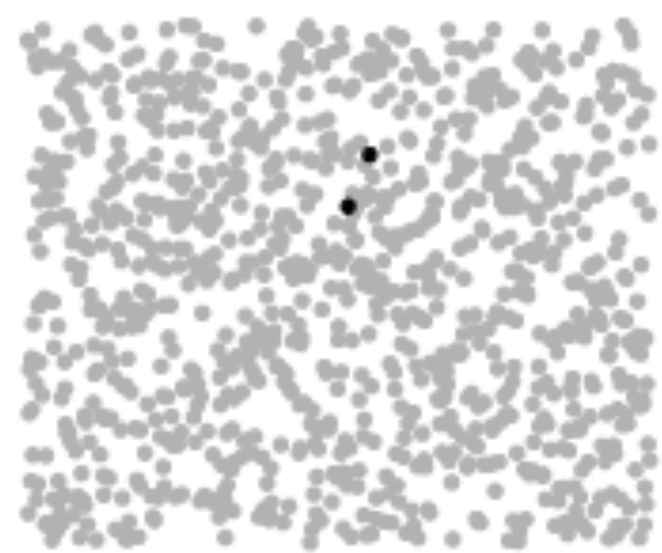
(a) Probability .1



(b) Probability .2



(c) Probability .0000001



(d) Probability .0000002

Journal club: today’s paper

Eells

- Instead of single event making something spurious, average over contexts
- Separated type/token
  - Probability in general != probability in single case

Background contexts

- Holding fixed other factors, does cause still raise probability of effect?
- Idea is to hold fixed everything relevant to the effect that's not caused by C

Causal background contexts

- n factors relevant to E independently of C
- $K=2^n$ ways of holding these factors fixed
- Background contexts are K_i's such that:
  - $P(K_i \wedge C)>0$
  - $P(K_i \wedge \neg C)>0$

E.g. factors A,B,C, one K_i = {A, ¬B,C}

Comparison to Suppes

- Suppes factors: earlier than prima facie cause
- Eells: factors earlier than effect, not caused by cause

Context unanimity

- For C to be positive cause, must raise probability of E with respect to **every** background context
  - $P(E|C \wedge K_i) > P(E|\neg C \wedge K_i)$ for all K_i

- Can't lower probability or remain same

- What about medications with rare side effects?

Four types of causal factors

Note: all relative to population

1. Positive
   $P(E|C \wedge K_i) > P(E|\neg C \wedge K_i)$ for all K_i
2. Negative
   $P(E|C \wedge K_i) < P(E|\neg C \wedge K_i)$ for all K_i
3. Neutral
   $P(E|C \wedge K_i) = P(E|\neg C \wedge K_i)$ for all K_i
4. Mixed
   Neither positive, negative, or neutral

Strength of relationship

- Not just relevant/not relevant, but how relevant
- Weed out factors that are always positively relevant, but make small difference to probability
- Average degree of causal significance

$$\sum_i [P(E|C \wedge K_i) - P(E|\neg C \wedge K_i)]P(K_i)$$

Example

- To assess whether smoking causes lung cancer, would hold fixed perhaps genetic factors (G), and asbestos exposure (A)
  - Both relevant to LC but independent of S
- Test whether P(LC) raised by S in each context $\{A \wedge G, A \wedge \neg G, \neg A \wedge G, \neg A \wedge \neg G\}$

Difficulties

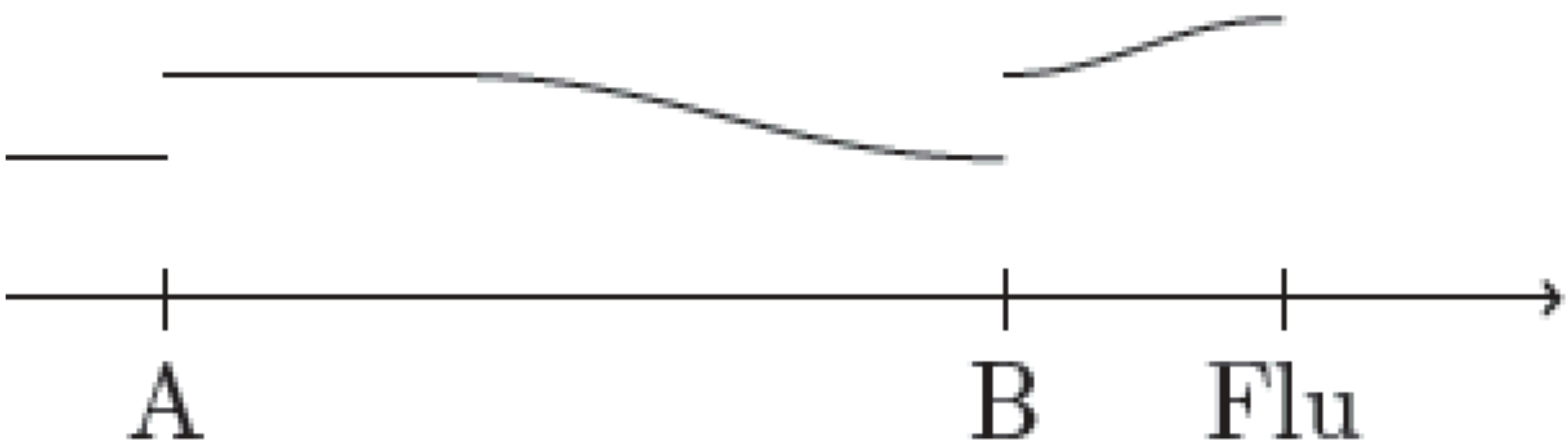
- Circular
  - Need to hold fixed causally relevant factors… to figure out if something is causally relevant
- Indirect
  - What happens in a causal chain? (not holding fixed other effects of cause)

Token causality

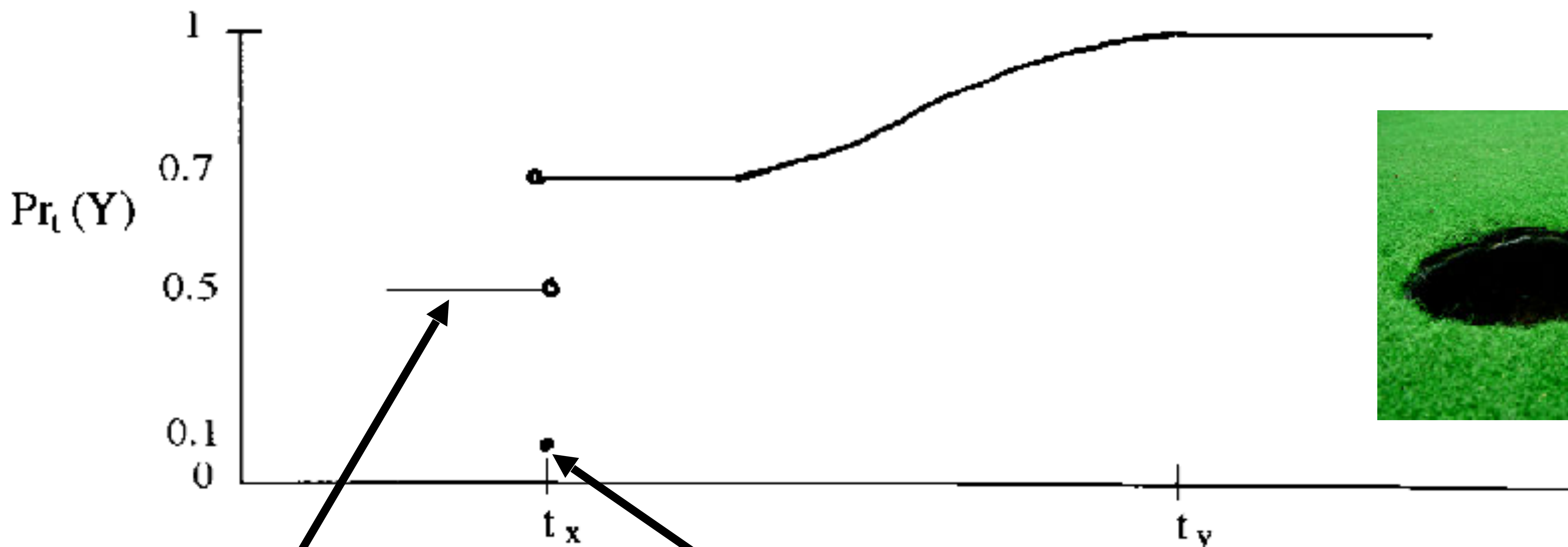
- Difference between P(LC|S) for individual and population
- Timing: lit match and house fire today versus match lit weeks ago
- Differing roles: Positive causes at type level may be negative at token level

Eells on token causality

- Type and token are distinct
- Look at how probability changes over time after cause
- Can distinguish between probability in general and probability on specific occasion
  - Note: this lets us update explanation to be consistent with observation


A
B
Flu

Probability trajectory



- P(Y) changes at x
- After x P(Y) is high
- P is higher than before x
- P remains high until y

Was Because Of!

Ways of answering "why"?

Not just caused/didn't cause, but what was role of cause in effect's occurrence?

Anne drives drunk and crashes car. What's significance of her drinking for the crash?

Could be

- Because of
- Despite
- Independently of

y is because of x if

- P(y) changes after x
- P(y) is high after x
- P(y) is higher than it was before x
- P(y) <u>remains</u> high until it occurs

Four relations

Despite

- y is Y <u>despite</u> x if the probability of Y is lowered after $x_{t'}$

Because

- y is Y <u>because</u> ***of*** x if the probability of Y is increased after $x_{t'}$ and <u>remains increased</u> until y_t

Autonomously

- y is Y <u>autonomously</u> ***of*** x if the probability of Y changes at $x_{t'}$, this probability is high, but then decreases before y_t

Independently

- y is Y <u>independently</u> ***of*** x if the probability of Y is the same just after $x_{t'}$ as it is just before $x_{t'}$

Poison and antidote

$P(D|O)=0.8$ in 30 minutes

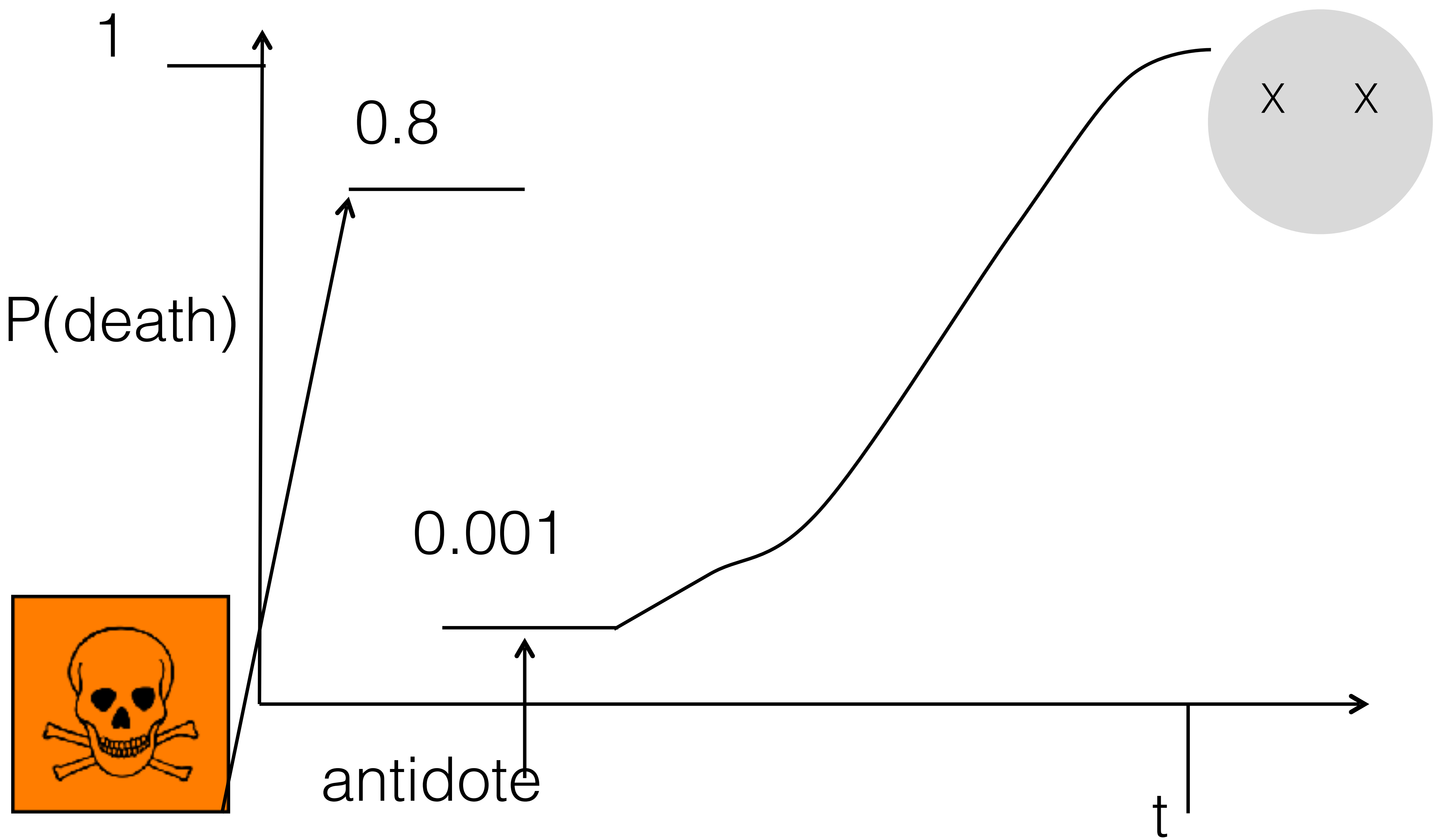
$P(D|A \wedge O)=0.001$ if within 10 minutes of O

$P(D|A \wedge \neg O)=0.001$

Are A and O positive, negative, neutral or mixed? Better handled by Suppes or Eells?

Can impact on a specific instance be different than that at type level? When?

Poison and antidote



Why are background contexts needed?

- What if poison and antidote are administered at same time?
- What about exposure that provides resistance to later bacteria?

More generally

- Need to account for temporally intermediate factors that affect outcome – independently of the cause
- Want to focus on impact of cause on probability, separating it from impact of other factor

Background factors

- Not relevant
  - Things caused by cause
  - Golf ball’s motion near cup
- Relevant
  - Factors interacting with cause, or causes of effect (not caused by cause)
  - Radon exposure for lung cancer

Token causal relevance

x is relevant to y if it happened either because of or despite x, after holding fixed two sets of factors

1. Factors that occur, are not token caused by x being X and are type-level causally relevant to y's being Y holding fixed how things are before these factors actually occur.
2. Factors that token occur in the particular case, are not token caused by x being X and that interact with X with respect to Y holding fixed what is actually true before x_t.

Caveats

- How can we know all the important factors?
- How can we find the probability trajectory?

Another approach

- Strength of relationship at token level proportional to that at type level
  - Radiation exposure and death / Marie Curie's radiation exposure causing her death

- Something is a type-level cause because it's often a token cause
  - If poison is fatal with P=0.8, have good reason to believe that when someone is poisoned and dies, this is another instance of death by poison

Connecting principle

If C is a causal factor for producing E in population P of magnitude m, then the support for the hypothesis that C at time t1 caused E at time t2 given that they both token occur in P at these times is m.

Sober, E., & Papineau, D. (1986). Causal factors, causal inference, causal explanation. Proceedings of the Aristotelian Society, Supplementary Volumes, 60, 97-136

Connecting principle

- C and E are types of events, but occur at specific times/places

$$m = \sum_i [P(E \mid C \wedge K_i) - P(E \mid \neg C \wedge K_i)] P(K_i)$$

- S(c caused e|c, e occur in P)=m

But what population?

- Sober: most specific
    - In explaining illness, if we know age/weight, the population is people with those characteristics
    - If only know gender, population is all men or women
- In practice..
    - Never have arbitrarily specific data
    - Would uniquely specify an individual if taken to extremes

Also…

- What if we do not know cause happened for certain?
- What about timing?

Recap of first section

- Main approaches
  - Regularities
  - Counterfactuals
  - Probabilistic causality
  - Omitted: interventionist
- Two problems
  - Type level
  - Token level

Looking forward

- Next few weeks
  - Graphical models
  - Bayes nets, DBNs
  - Granger causality
  - Other computational methods
- Then
  - Experiment design, applications

For next time…

Be sure to read discussion paper!