Solving problems using imperfect advice

Davin Choo

24 March 2023

What is a prime factor of this?

66

What is a prime factor of this?

$$66 = 2 \times 3 \times 11$$

What is a prime factor of this?

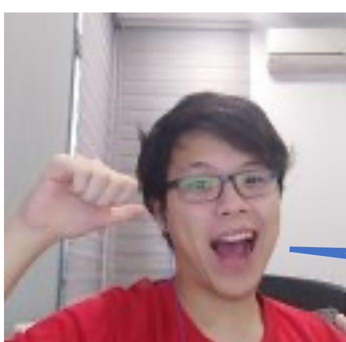
$$66 = 2 \times 3 \times 11$$

101291

What is a prime factor of this?

66 = 2 x 3 x 11

101291



It is a product of two numbers

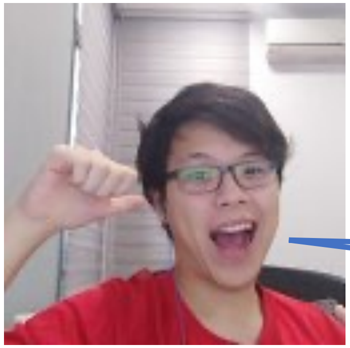
What is a prime factor of this?

66 = 2 x 3 x 11

101291

Idea: Let's just output the smaller factor

- sqrt(101291) = 318.262470298
- Brute force:
  - Is 2 a factor? ✗
  - Is 3 a factor? ✗
  - Is 4 a factor? ✗
  - …



It is a product of two numbers

What is a prime factor of this?

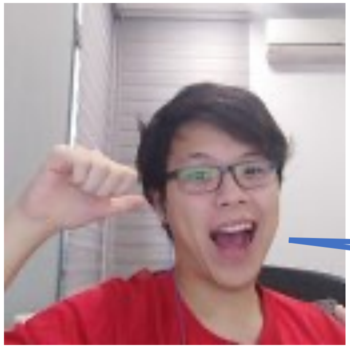
66 = 2 x 3 x 11

101291

Idea: Let's just output the smaller factor

- sqrt(101291) = 318.262470298
- Brute force:
  - Is 2 a factor? ✗
  - Is 3 a factor? ✗
  - Is 4 a factor? ✗
  - …
- Smarter: Sieve of Eratosthenes

(If you can generically solve this efficiently, you can break RSA)

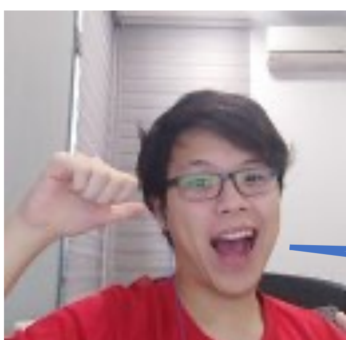


It is a product of two numbers

What is a prime factor of this?

66 = 2 x 3 x 11

101291



Let me help you!

(If you can generically solve this efficiently, you can break RSA)

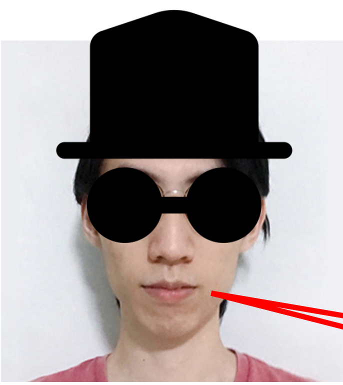
It is a product of two numbers

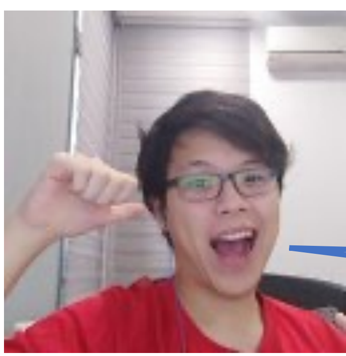
https://scholar.google.com/citations?user=f5-f2c8AAAAJ&hl=en

What is a prime factor of this?

66 = 2 x 3 x 11

101291





Let me help you!

(If you can generically solve this efficiently, you can break RSA)

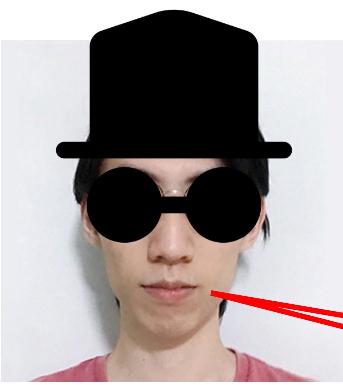
It is a product of two numbers

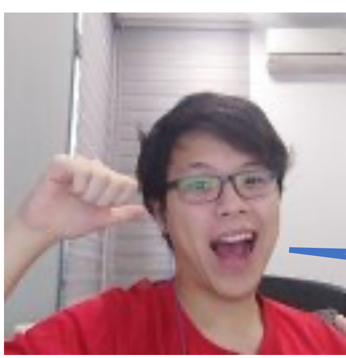
https://thenounproject.com/icon/spy-2680743/

What is a prime factor of this?

66 = 2 x 3 x 11

101291





One of the factors is ≈500

(If you can generically solve this efficiently, you can break RSA)

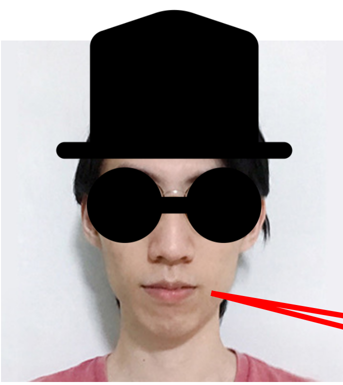
It is a product of two numbers

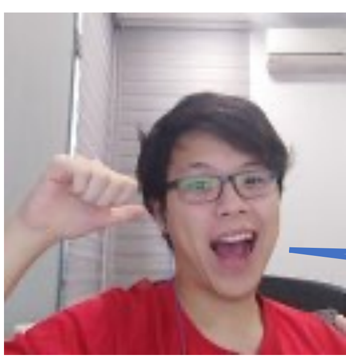
What is a prime factor of this?

66 = 2 x 3 x 11

101291

101291/500 = 202.582 ⇒ ✗





One of the factors is ≈500

(If you can generically solve this efficiently, you can break RSA)

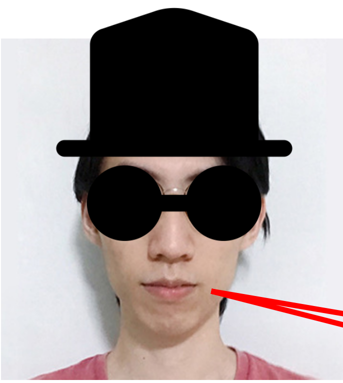
It is a product of two numbers

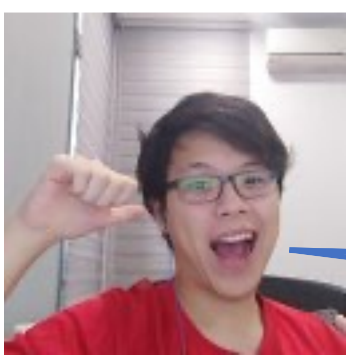
What is a prime factor of this?

$66 = 2 \times 3 \times 11$

101291

101291/500 = 202.582 ⇒ ✗
101291/501 = 202.177644711 ⇒ ✗





One of the factors is ≈500

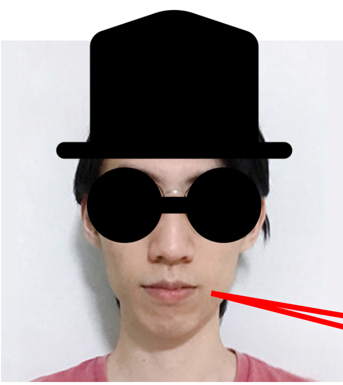
(If you can generically solve this efficiently, you can break RSA)

It is a product of two numbers

What is a prime factor of this?

$$66 = 2 \times 3 \times 11$$

$$101291$$



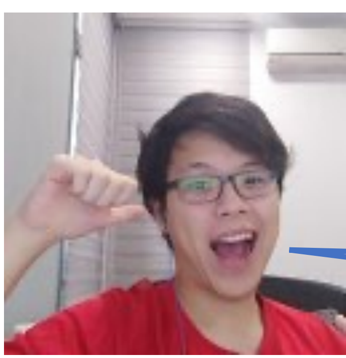
$101291/500 = 202.582 \Rightarrow$ ✗

$101291/501 = 202.177644711 \Rightarrow$ ✗

$101291/499 = 202.987975952 \Rightarrow$ ✗

One of the factors is ≈500

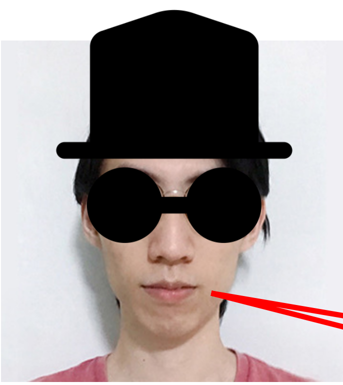
(If you can generically solve this efficiently, you can break RSA)



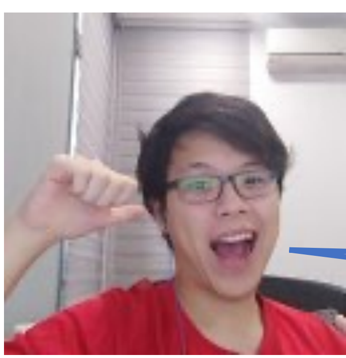
It is a product of two numbers

What is a prime factor of this?

$$66 = 2 \times 3 \times 11$$
$$101291 = 199 \times 509$$



101291/500 = 202.582 ⇒ ✗
101291/501 = 202.177644711 ⇒ ✗
101291/499 = 202.987975952 ⇒ ✗
...
101291/509 = 199 ⇒ Found it! ✓



One of the factors is ≈500

(If you can generically solve this efficiently, you can break RSA)

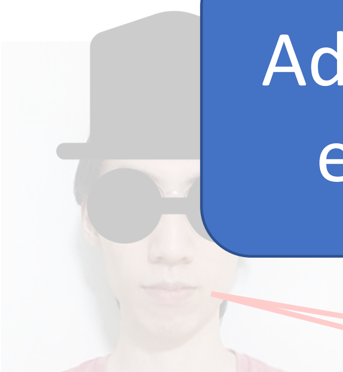
It is a product of two numbers

What is a prime factor of this?

$$66 = 2 \times 3 \times 11$$

$$101291 = 199 \times 509$$

Advice can help solve hard problems, especially if advice is high quality!



101291/509 = 199 ⇒ Found it! ✓

One of the factors is ≈500

(If you can generically solve this efficiently, you can break RSA)

It is a product of two numbers

Searching sorted array

- Problem

Given a sorted array A on n numbers, locate $x^* \in A$

- Solution

Binary search: $\mathcal{O}(\log n)$ queries

Without further assumptions, $\Omega(\log n)$ queries

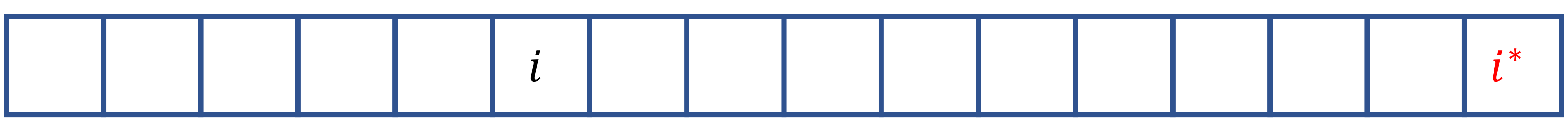
Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

  Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$



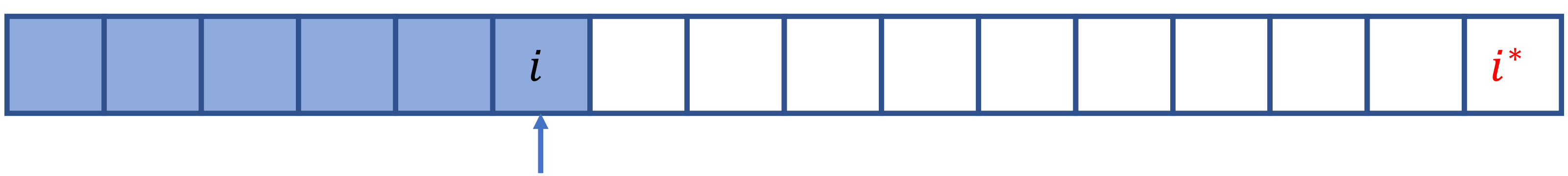
Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

  Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$
- Linear probe



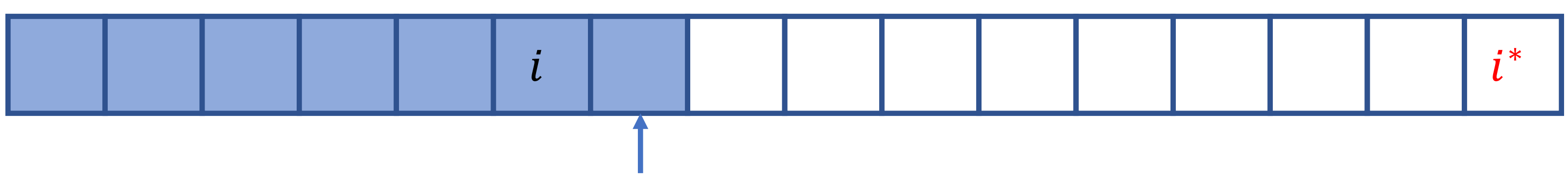
Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$
- Linear probe



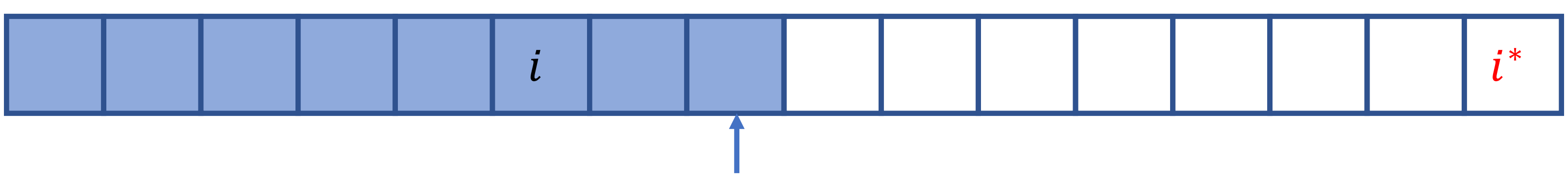
Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

  Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$
- Linear probe



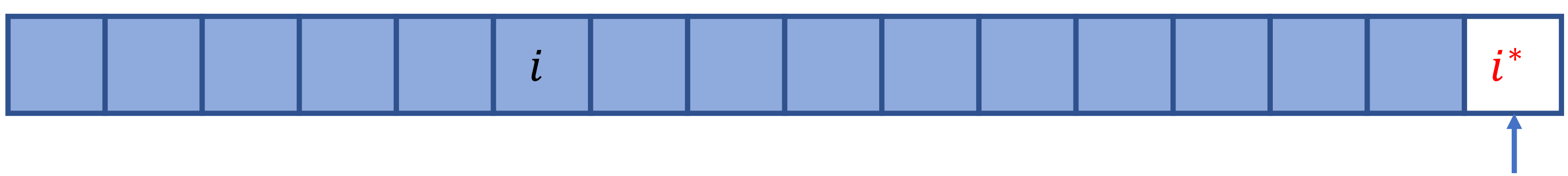
Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

  Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$
- Linear probe

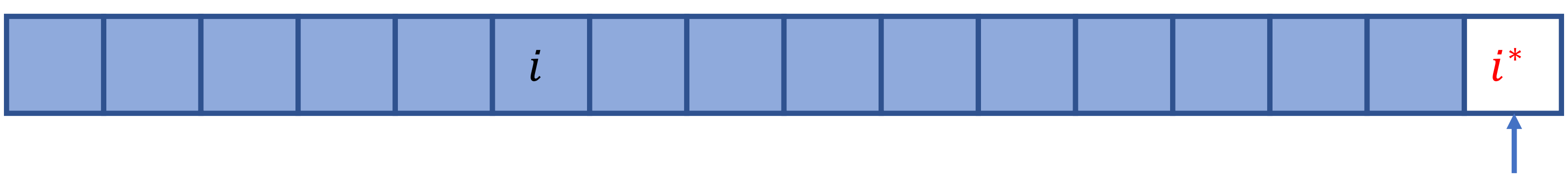


Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

  Given a sorted array A on n numbers, locate $x^* \in A$
- Advice: index $i \in [n]$
- Linear probe: Good when $i \approx i^*$; $\mathcal{O}(n)$ worst case

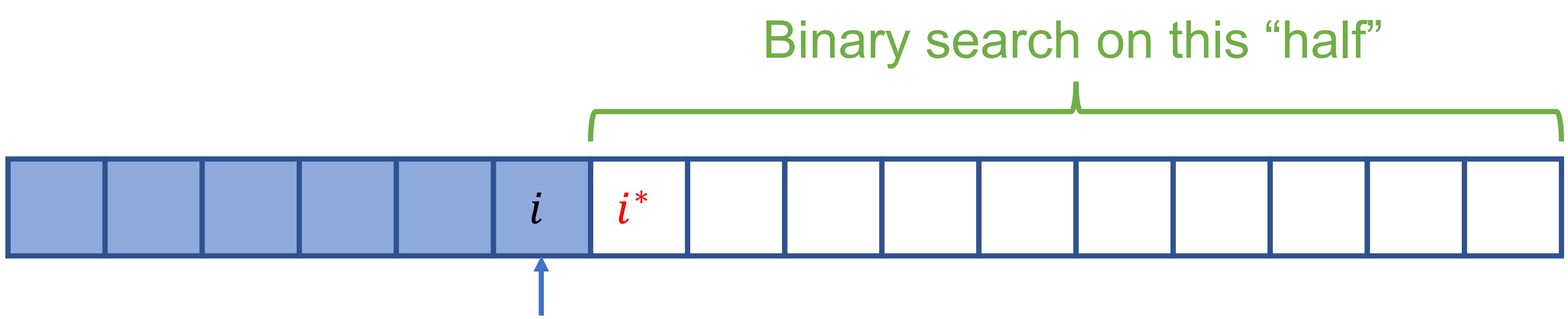


Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

  Given a sorted array A on n numbers, locate $x^* \in A$
- Advice: index $i \in [n]$
- Linear probe: Good when $i \approx i^*$; $\mathcal{O}(n)$ worst case
- Prune then binary search



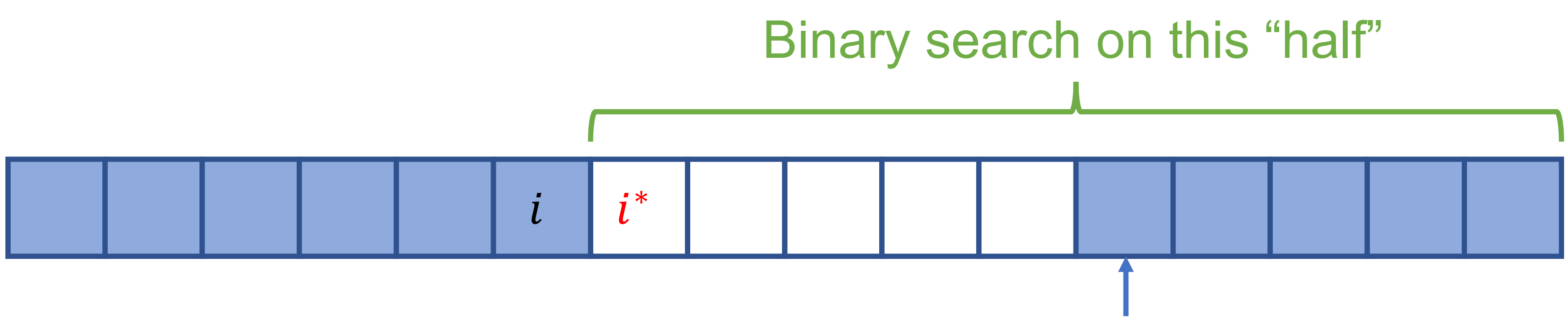
Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$
- Linear probe: Good when $i \approx i^*$; $\mathcal{O}(n)$ worst case
- Prune then binary search



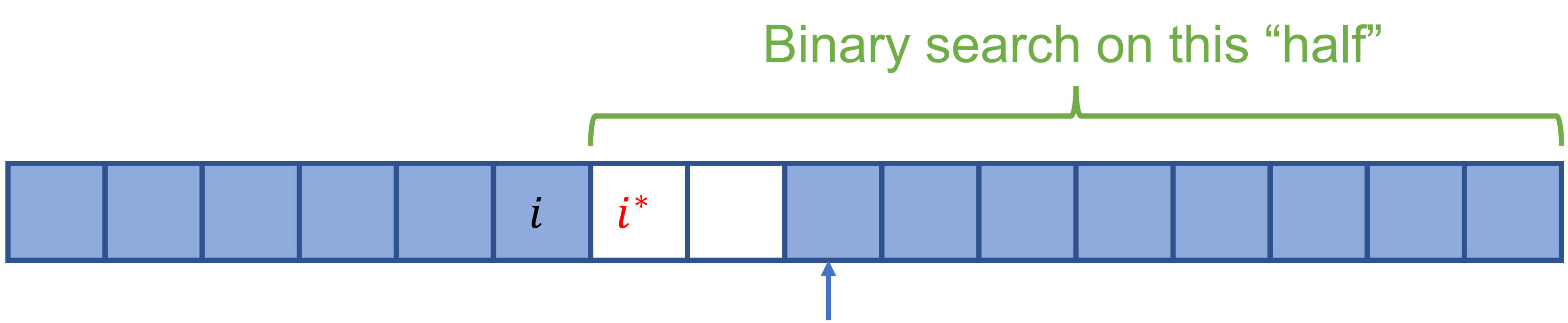
Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

  Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$
- Linear probe: Good when $i \approx i^*$; $\mathcal{O}(n)$ worst case
- Prune then binary search

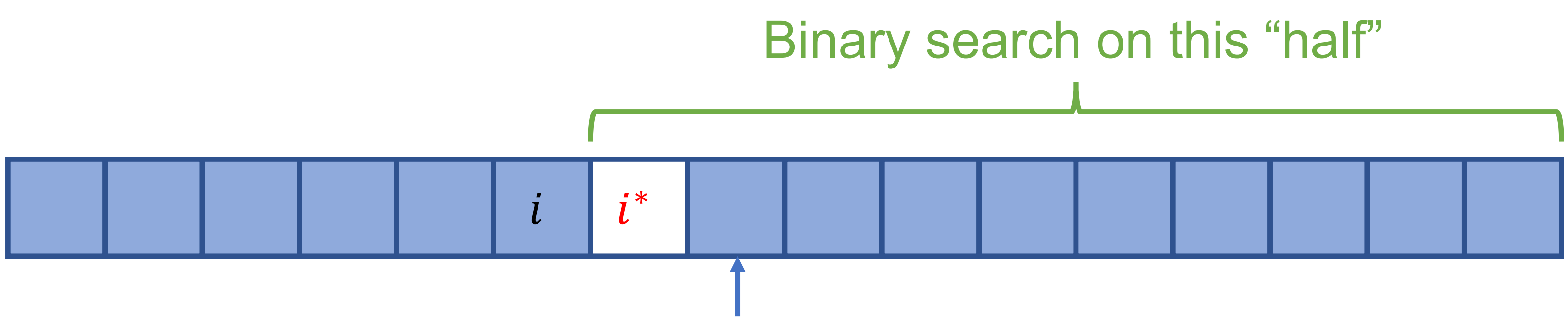


Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

  Given a sorted array A on n numbers, locate $x^* \in A$
- Advice: index $i \in [n]$
- Linear probe: Good when $i \approx i^*$; $\mathcal{O}(n)$ worst case
- Prune then binary search



Searching sorted array with advice

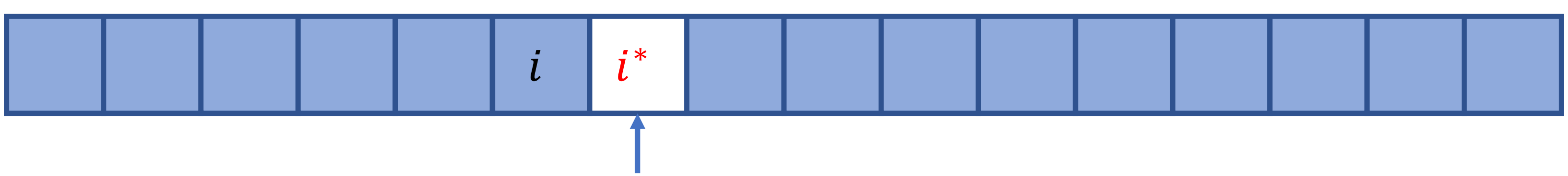
Suppose $A[i^*] = x^*$

- Problem

  Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$
- Linear probe: Good when $i \approx i^*$; $\mathcal{O}(n)$ worst case
- Prune then binary search: Doesn't exploit $i \approx i^*$

  (Still worst case $\mathcal{O}(\log n)$)



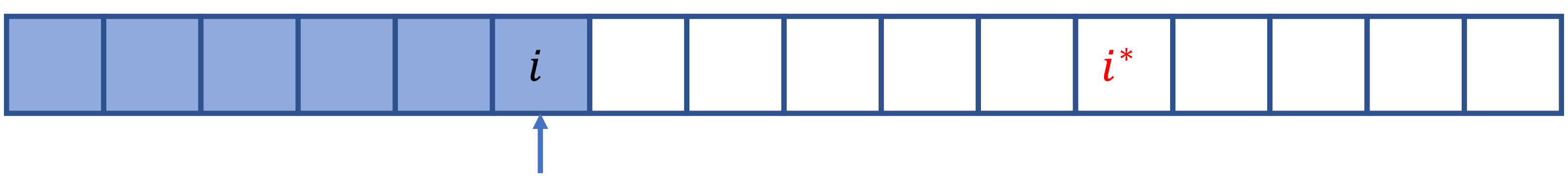
Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

  Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$
- Idea: Exponential search from $\mathrm{A}[i]$, then binary search within bounds



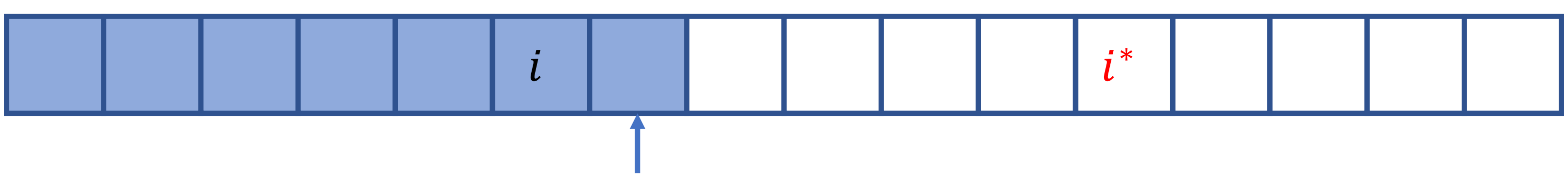
Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$
- Idea: Exponential search from $\mathrm{A}[i]$, then binary search within bounds



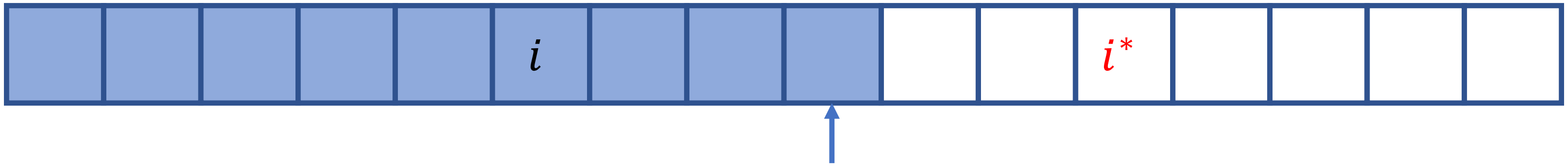
Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

  Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$
- Idea: Exponential search from $\mathrm{A}[i]$, then binary search within bounds



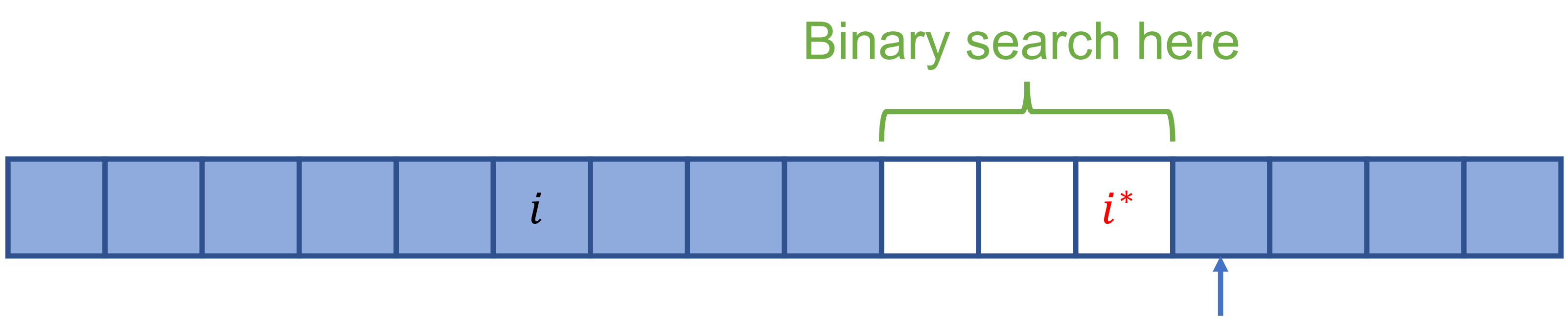
Searching sorted array with advice

Suppose $A[i^*] = x^*$

- Problem

Given a sorted array A on n numbers, locate $x^* \in A$

- Advice: index $i \in [n]$
- Idea: Exponential search from $\mathrm{A}[i]$, then binary search within bounds $\mathcal{O}(\log |i - i^*|)$ suffices!



Desiderata

$\mathcal{O}(\log |i - i^*|)$

1. When advice is "perfect"

   Perform as well as just the oracle

2. When advice is "garbage"

   Perform as well as state-of-the-art without advice

3. Provable guarantees and "efficient"

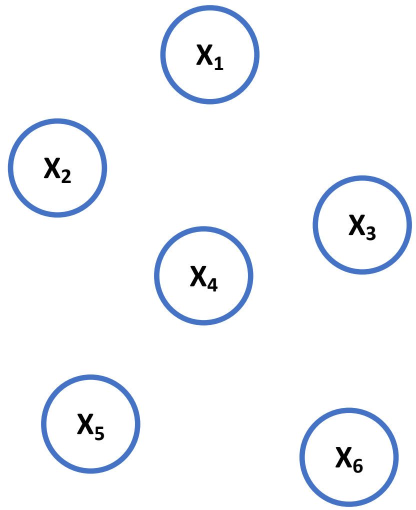
Cute idea… but is it useful?

https://www.facebook.com/idealistxavier/photos/a.112131290688897/344800484088642/?paipv=0&eav=Afb8y52jLpa70_fgwy7IdoKBZ_ILT6WA6684hlWwhoyTDWM3cPnRjG3aVjhPeLqpuOM&_rdr

Case study:
Causal graph discovery via interventions

Health factors

How are they related?



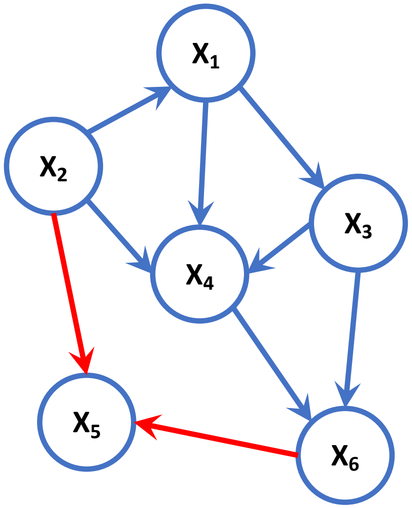
Example:
X_2 = Blood glucose
X_6 = Blood pressure

Image credit: https://thenounproject.com/icon/doctor-1285618/

Health factors

How are they related?

Arrows denote acyclic **causal** relationships



Example:
X_2 = Blood glucose
X_6 = Blood pressure

Example scenario:

Low levels of X_2 and X_6 **leads to** low levels of X_5 (say, cancer cell counts)

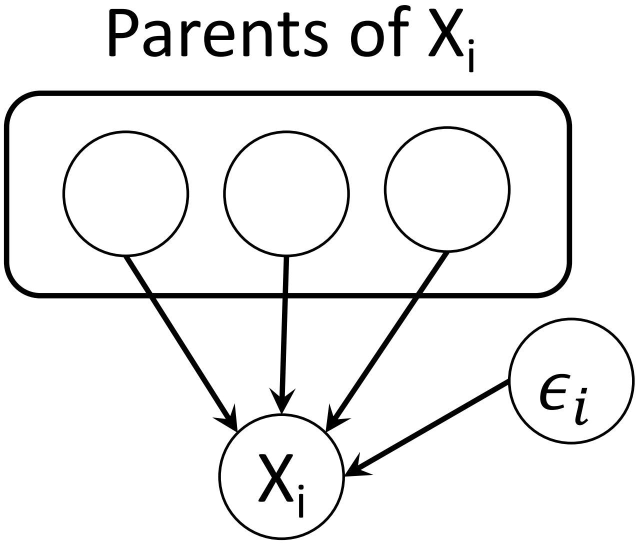
Suppose high levels of X_5 is known to cause disease Y but there is no available treatment to suppress X_5. Then, given the above causal relationships, a possible treatment is to suppress the levels of X_2 and X_6

Modelling causal relations

"**We may regard the present state of the universe as the effect of its past and the cause of its future…**" – Pierre Simon Laplace, *A Philosophical Essay on Probabilities*, 1814

$$X_i = f_i(pa_i, \epsilon_i)$$

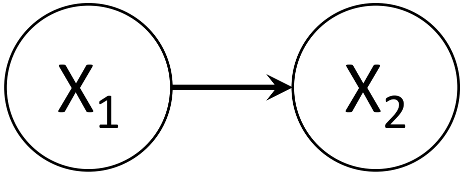
The value of each variable X_i is function f_i of the values taken by its parents pa_i and some noise ϵ_i

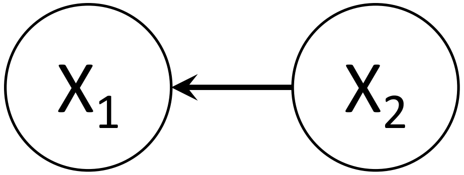


https://en.wikipedia.org/wiki/Laplace%27s_demon

Which model generated this data?

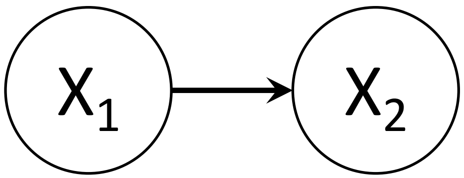
X_1	-0.27	0.29	0.37	-0.09	0.34	0.33	0.30	-1.34	0.68
X_2	-0.10	1.65	0.47	1.92	2.04	1.67	0.11	-3.58	1.97



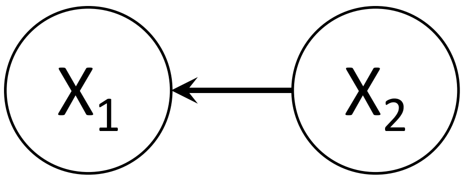


Which model generated this data?

X_1	-0.27	0.29	0.37	-0.09	0.34	0.33	0.30	-1.34	0.68
X_2	-0.10	1.65	0.47	1.92	2.04	1.67	0.11	-3.58	1.97



- $X_1 = \epsilon_1$
- $X_2 = a \cdot X_1 + \epsilon_2$



- $X_1 = b \cdot X_2 + \epsilon_3$
- $X_2 = \epsilon_4$

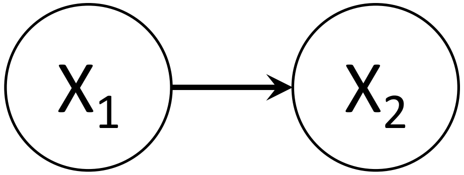
Simple linear relationship between variables

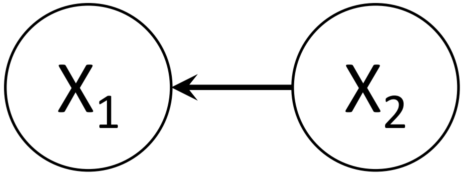
a and b are (hidden) positive constants

ϵ's are independent Gaussian terms with mean 0

Two equivalent causal models

X_1	-0.27	0.29	0.37	-0.09	0.34	0.33	0.30	-1.34	0.68
X_2	-0.10	1.65	0.47	1.92	2.04	1.67	0.11	-3.58	1.97





- $X_1 = \epsilon_1 \sim N(0,1)$
- $X_2 = X_1 + \epsilon_2 \sim N(0,2)$
- $\epsilon_1 \sim N(0,1)$
- $\epsilon_2 \sim N(0,1)$

- $X_1 = \frac{1}{2} \cdot X_2 + \epsilon_3 \sim N(0,1)$
- $X_2 = \epsilon_4 \sim N(0,2)$
- $\epsilon_3 \sim N\left(0, \frac{1}{2}\right)$
- $\epsilon_4 \sim N(0,2)$

Two equivalent causal models

X_1	-0.27	0.29	0.37	-0.09	0.34	0.33	0.30	-1.34	0.68
X_2	-0.10	1.65	0.47	1.92	2.04	1.67	0.11	-3.58	1.97

X_1 X_2

So what?
Who cares?

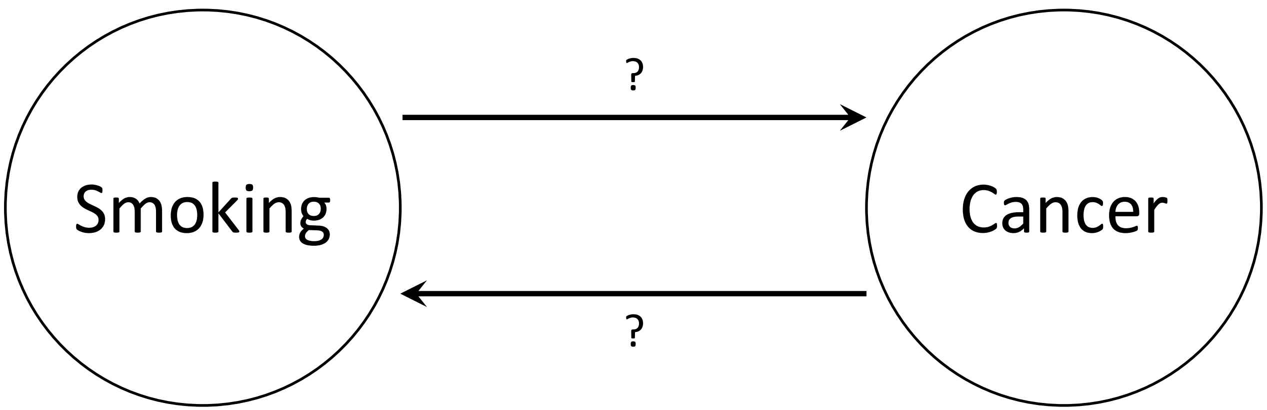
- $X_1 = \epsilon_1 \sim N(0, 1)$
- $X_2 = X_1 + \epsilon_2 \sim N(0, 2)$
- $\epsilon_1 \sim N(0, 1)$
- $\epsilon_2 \sim N(0, 1)$

- $X_1 = \frac{1}{2} \cdot X_2 + \epsilon_3 \sim N(0, 1)$
- $X_2 = \epsilon_4 \sim N(0, 2)$
- $\epsilon_3 \sim N\left(0, \frac{1}{2}\right)$
- $\epsilon_4 \sim N(0, 2)$

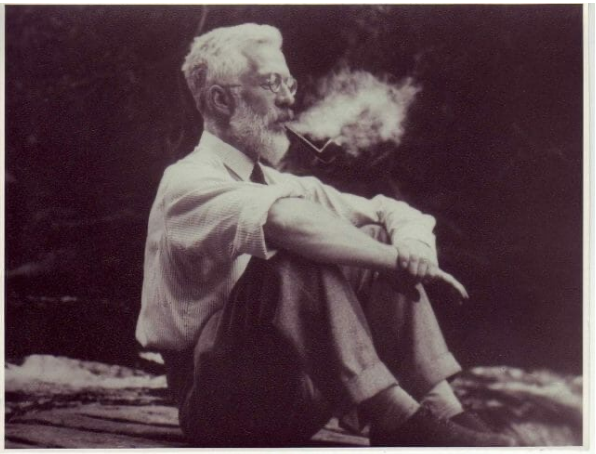
Example from: https://youtu.be/rE6IMfSkOU0?t=849. See https://github.com/csquires/6.S091-causality for full course

Smoking

Smoking	Yes	Yes	Yes	No	No	No	...
Cancer	No	Yes	Yes	No	No	Yes	...



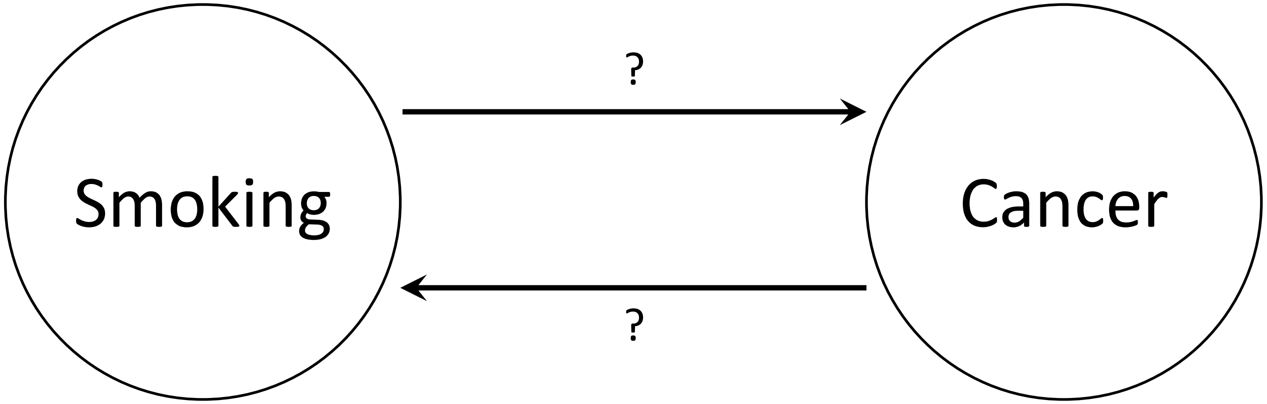
Smoking



Ronald Fisher

Fisher’s letter to Nature, 1958:

“The curious associations with lung cancer found in relation to smoking habits do not, in the minds of some of us, lend themselves easily to the simple conclusion that the products of combustion reaching the surface of the bronchus induce, though after a long interval, the development of a cancer... **Such results suggest that an error has been made, of an old kind, in arguing from correlation to causation**...”

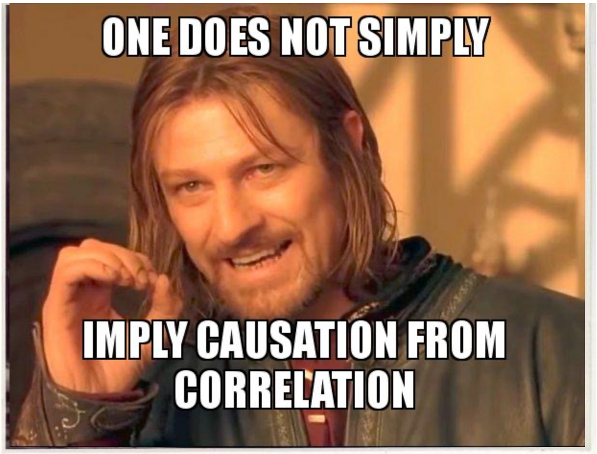


Fisher’s letter to Nature, 1958: https://www.nature.com/articles/182596a0.pdf
Nice blog post: https://priceonomics.com/why-the-father-of-modern-statistics-didnt-believe/ (picture of Fisher from this link)

Smoking

Fisher's letter to Nature, 1958:

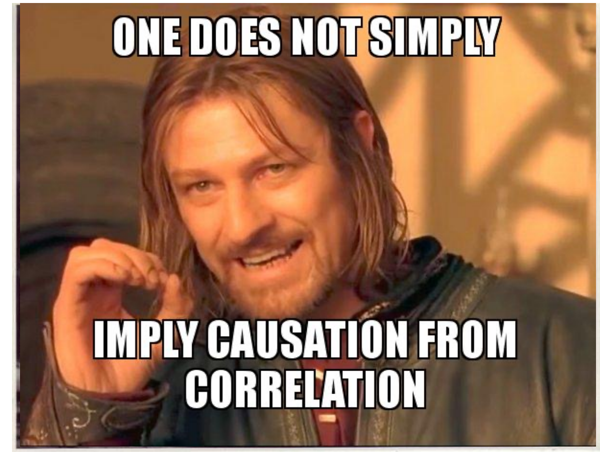
"The curious associations with lung cancer found in relation to smoking habits do not, in the minds of some of us, lend themselves easily to the simple conclusion that the products of combustion reaching the surface of the bronchus induce, though after a long interval, the development of a cancer... **Such results suggest that an error has been made, of an old kind, in arguing from correlation to causation**..."



Ronald Fisher

From Google image search: https://makeameme.org/meme/one-does-not-5aed73

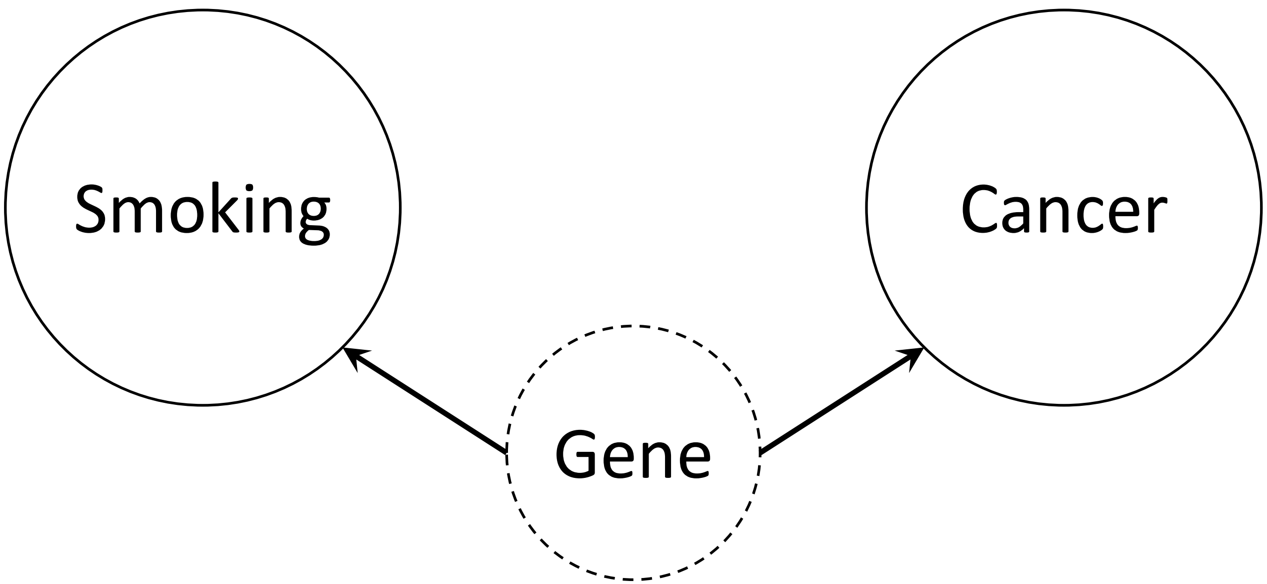
Smoking



Ronald Fisher

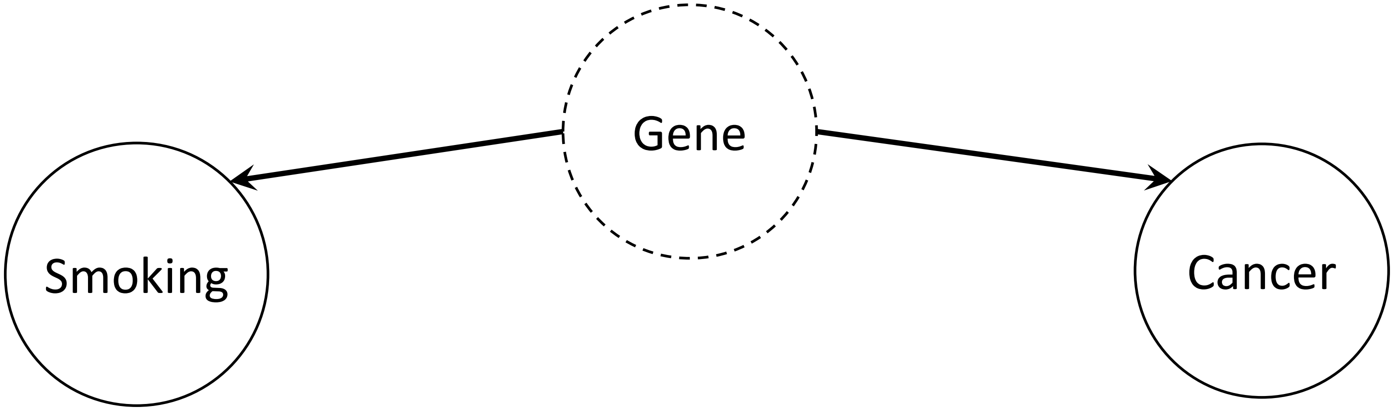
Fisher's letter to Nature, 1958:

"… Such results suggest that an error has been made, of an old kind, in arguing from correlation to causation, and that the possibility should be explored that the different smoking classes, non-smokers, cigarette smokers, cigar smokers, pipe smokers, etc., have adopted their habits partly by reason of their personal temperaments and dispositions, and are not lightly to be assumed to be equivalent in their **genotypic composition**…"

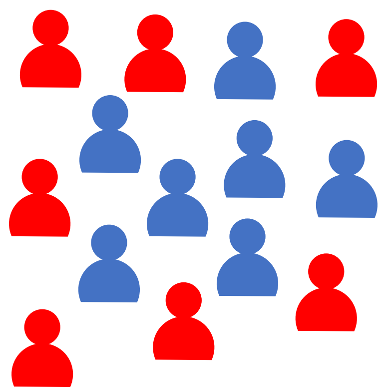


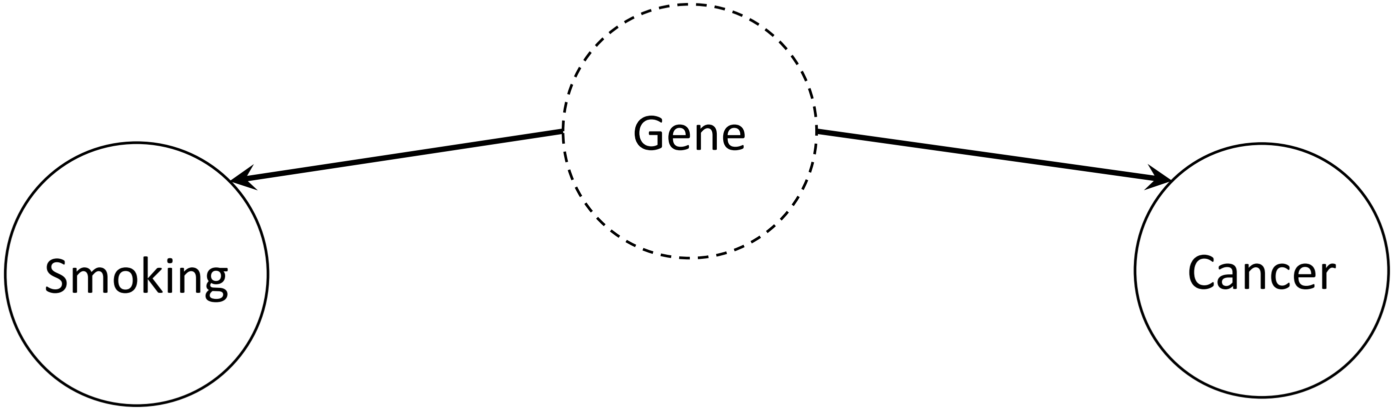
Maybe there's an unmeasured confounder?

From Google image search: https://makeameme.org/meme/one-does-not-5aed73

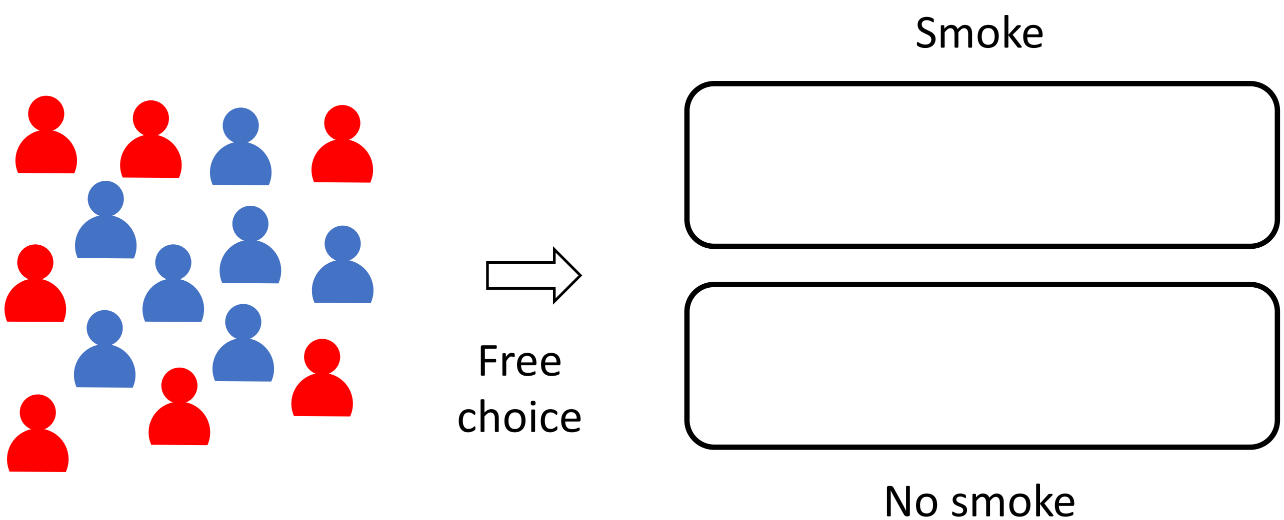


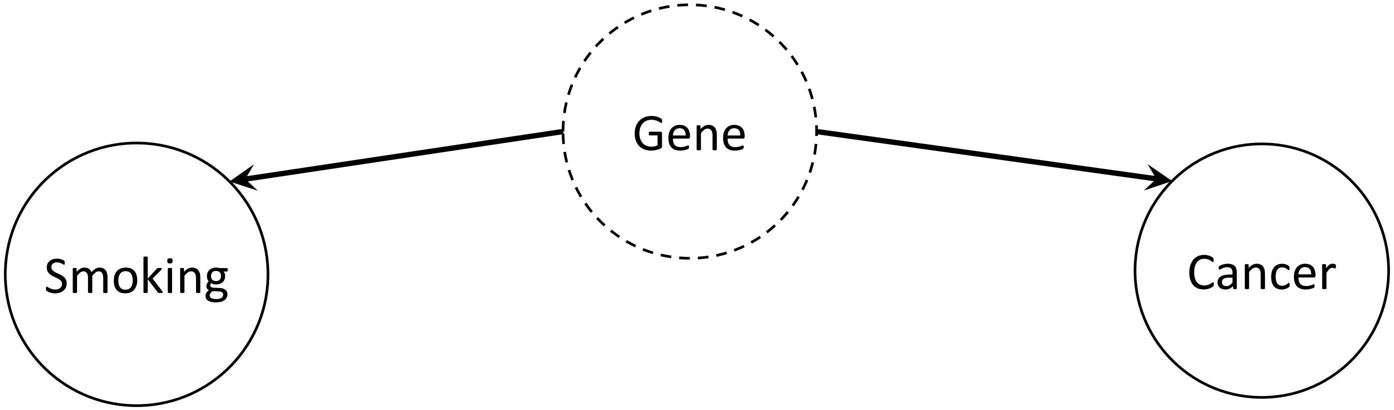
Hypothesis: There are two types of people in the world



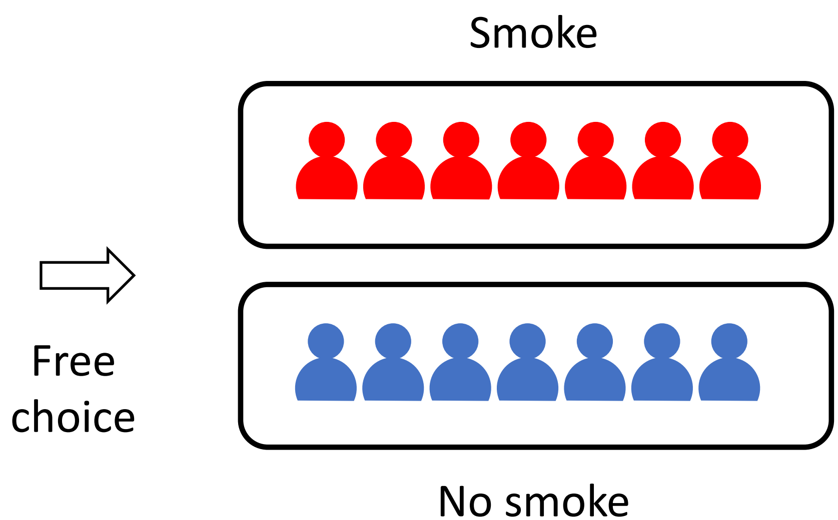


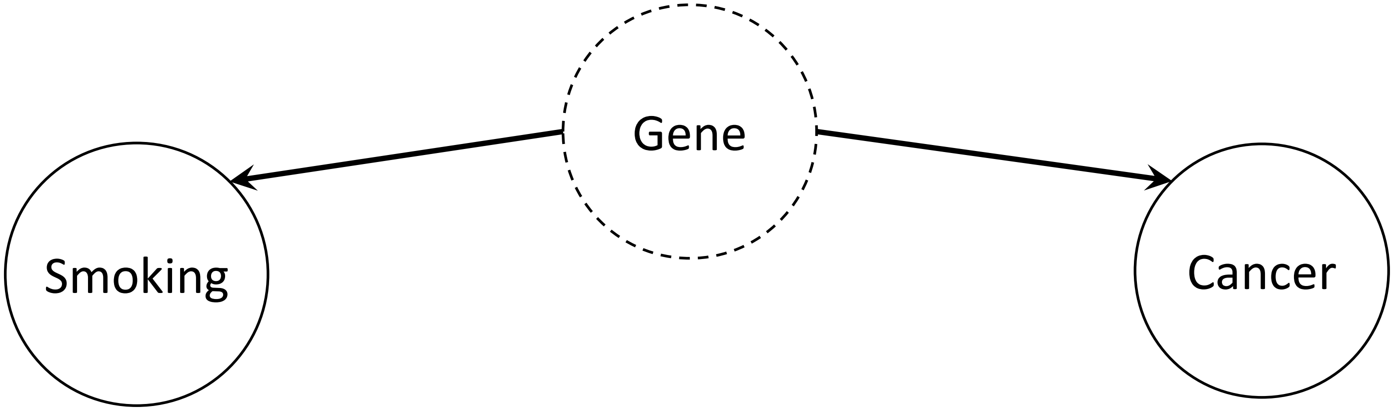
Hypothesis: There are two types of people in the world



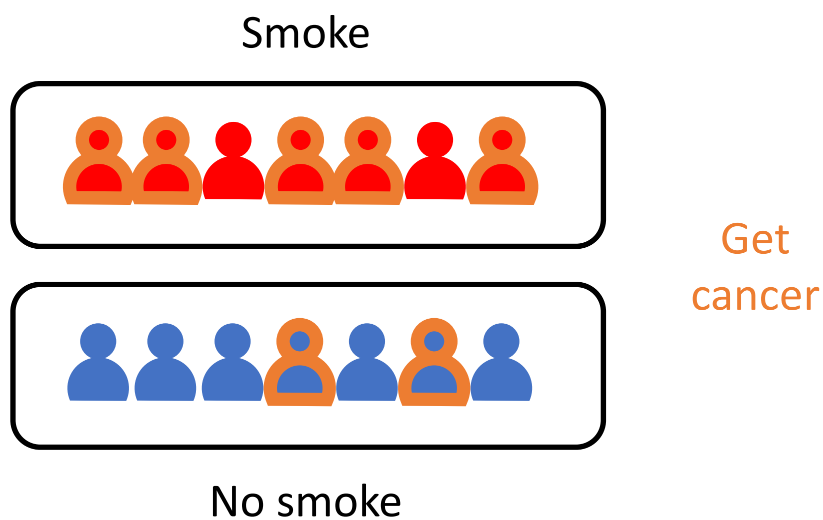


Hypothesis: There are two types of people in the world



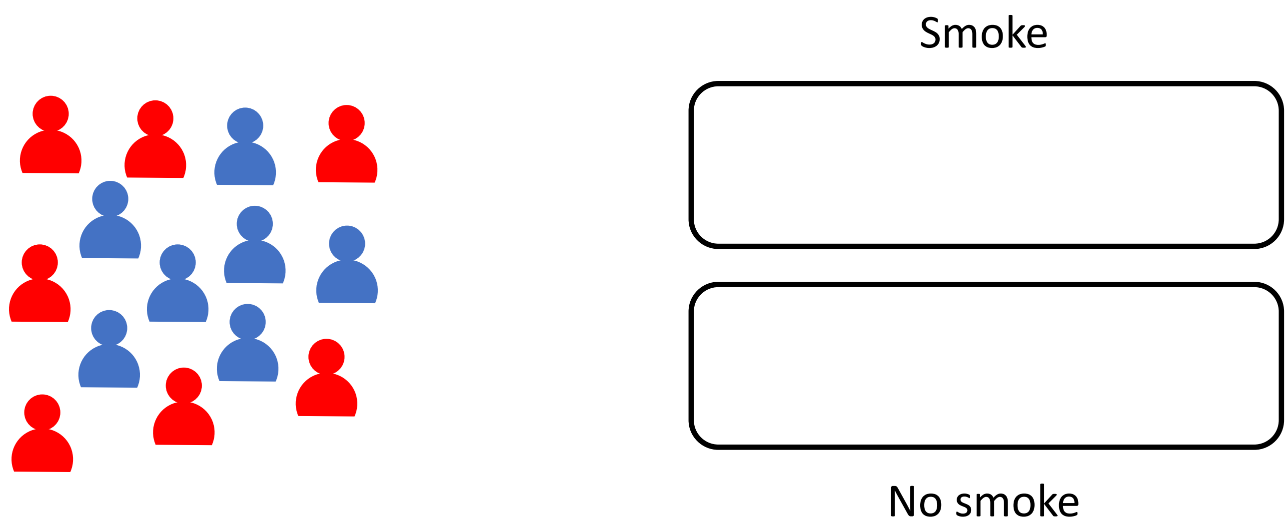


Hypothesis: There are two types of people in the world



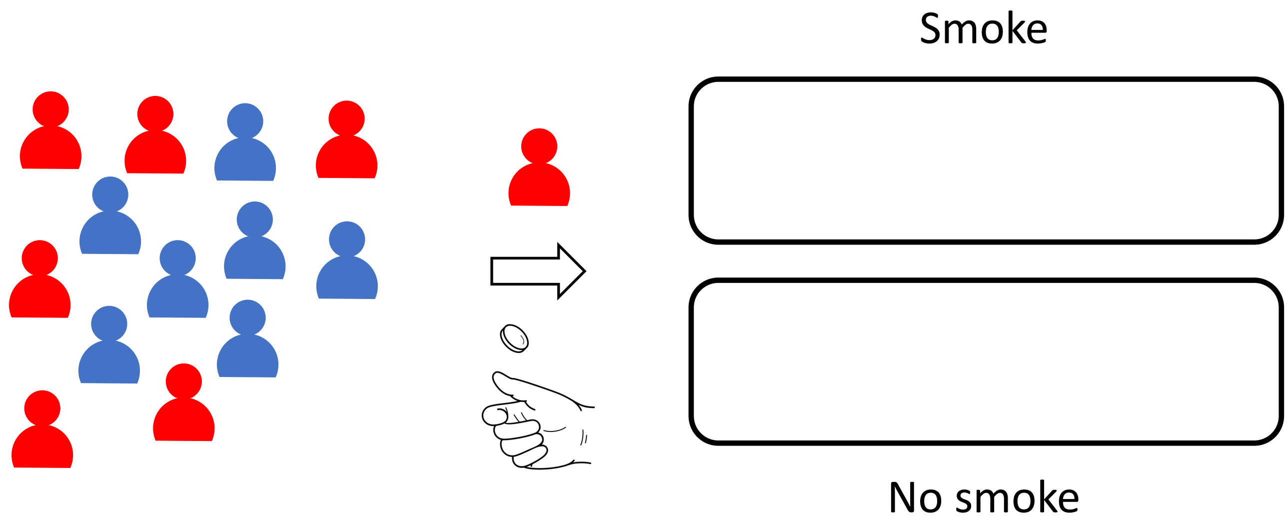
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



Randomized controlled trials

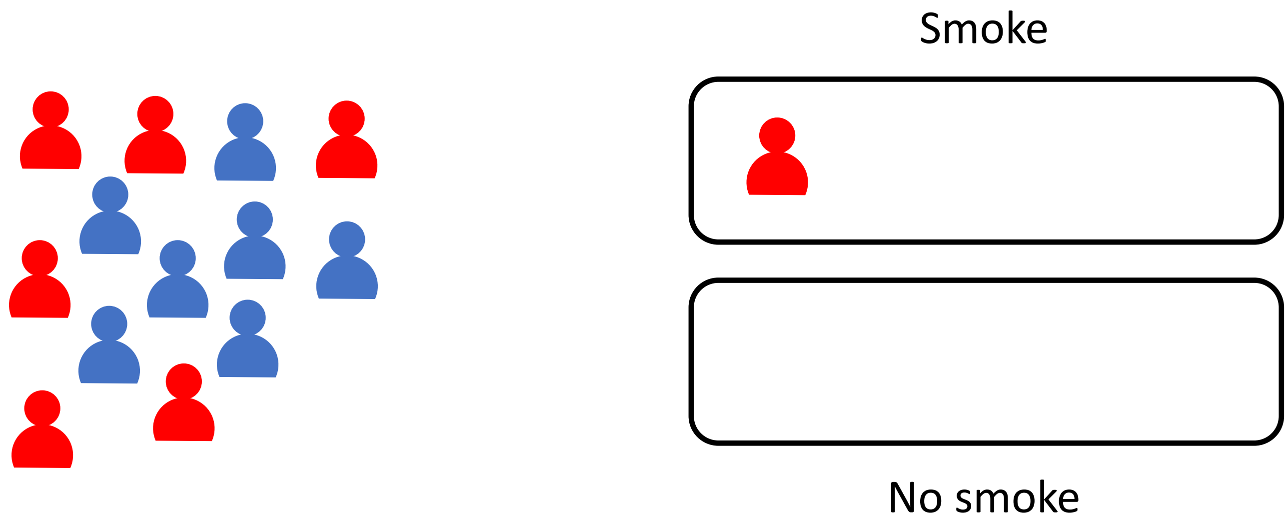
- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



https://thenounproject.com/icon/coin-flipping-2307580/

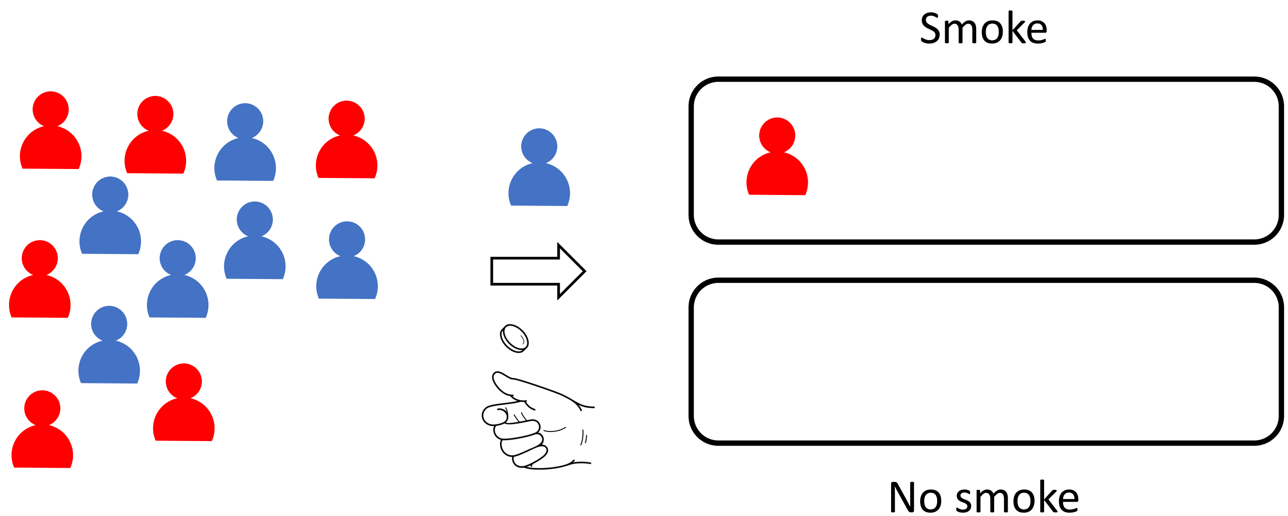
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



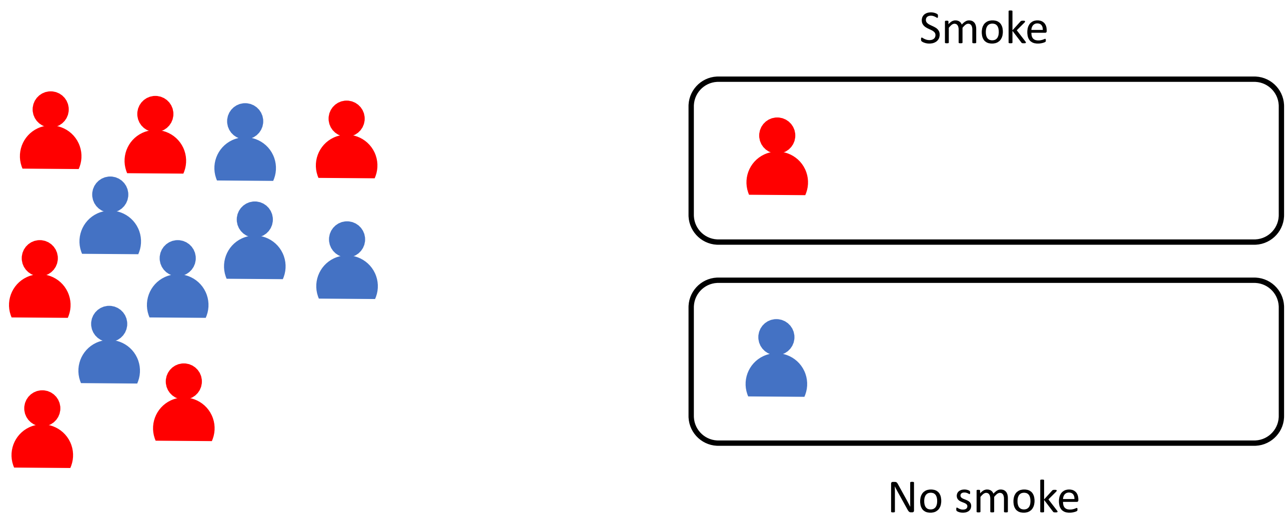
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



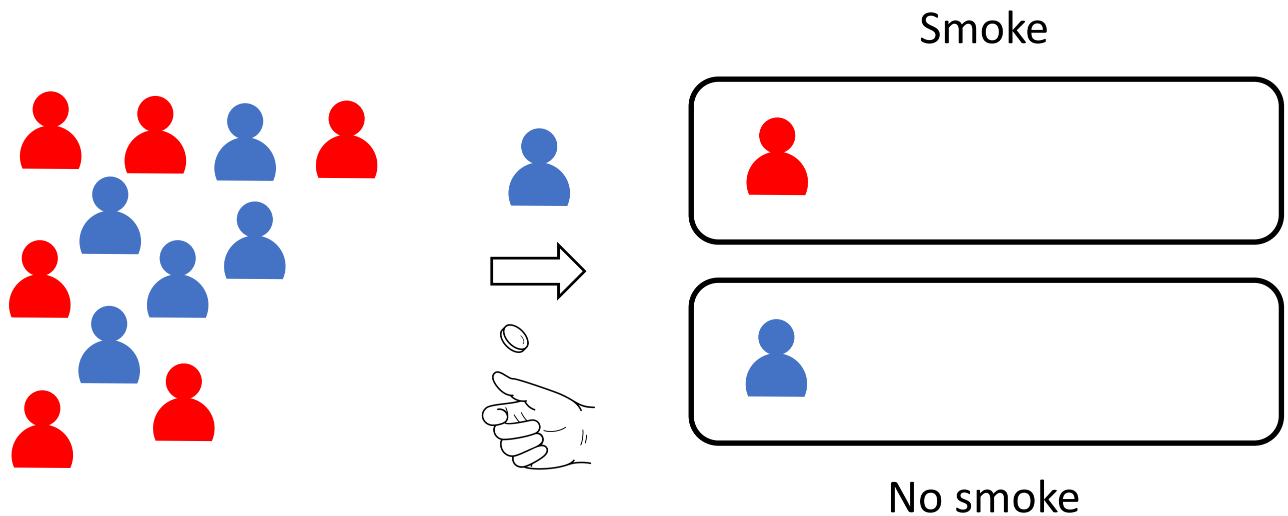
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



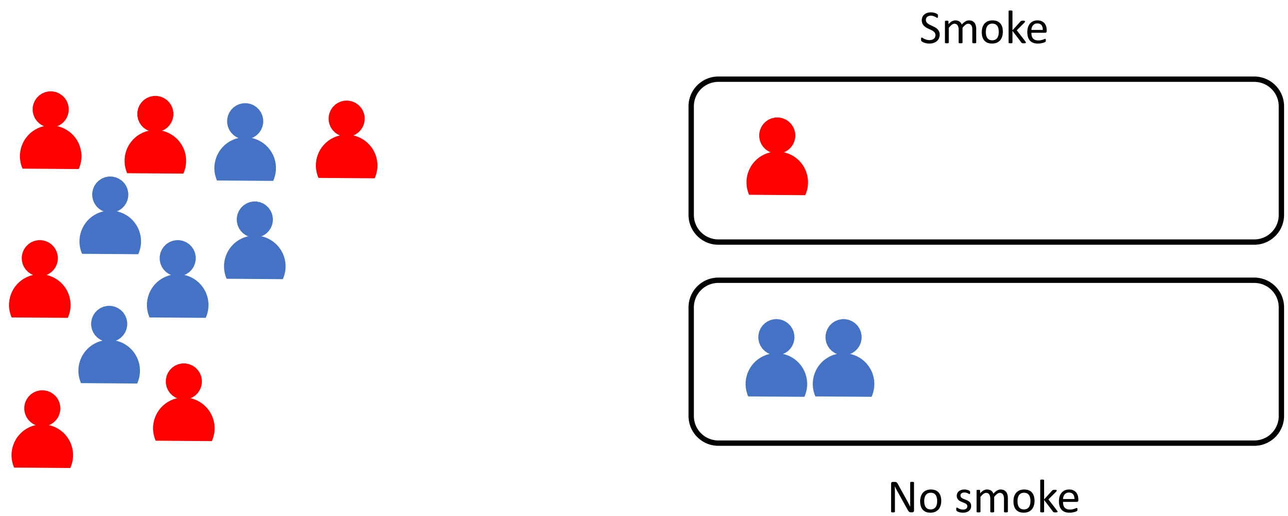
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



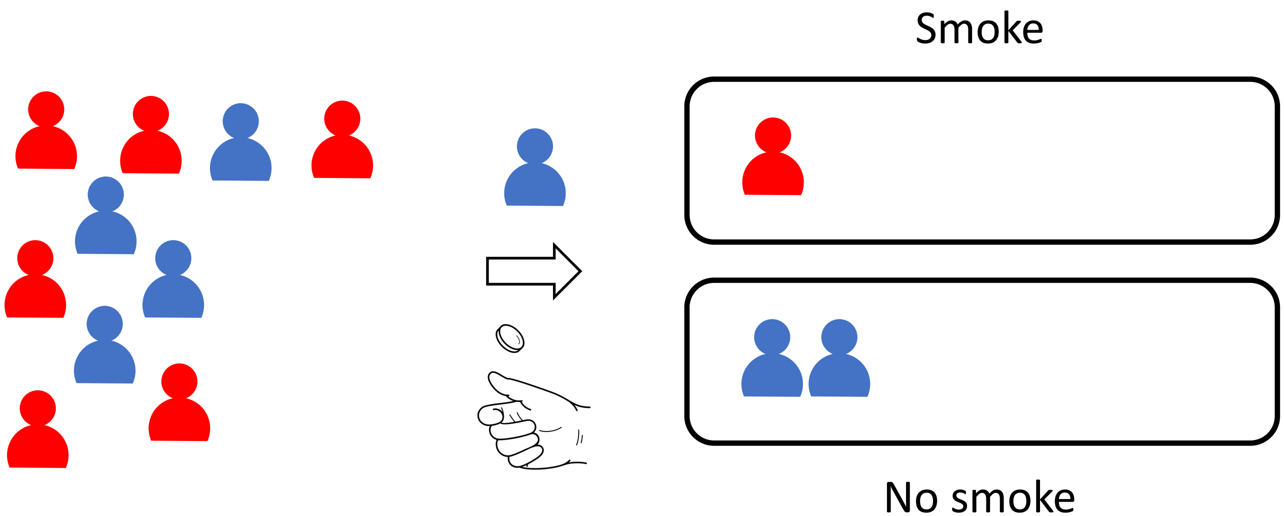
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



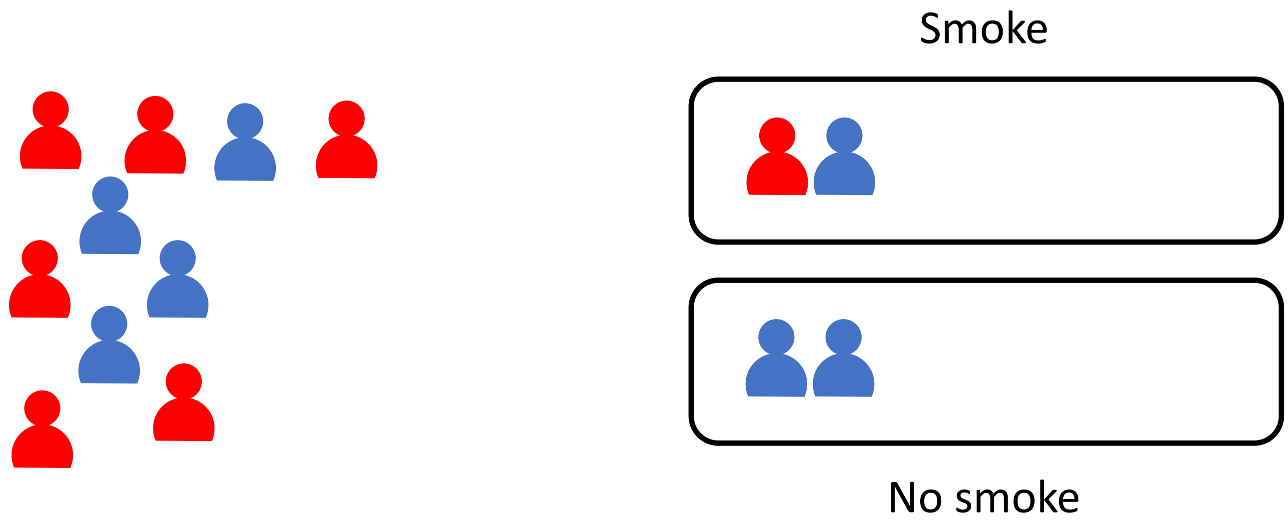
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



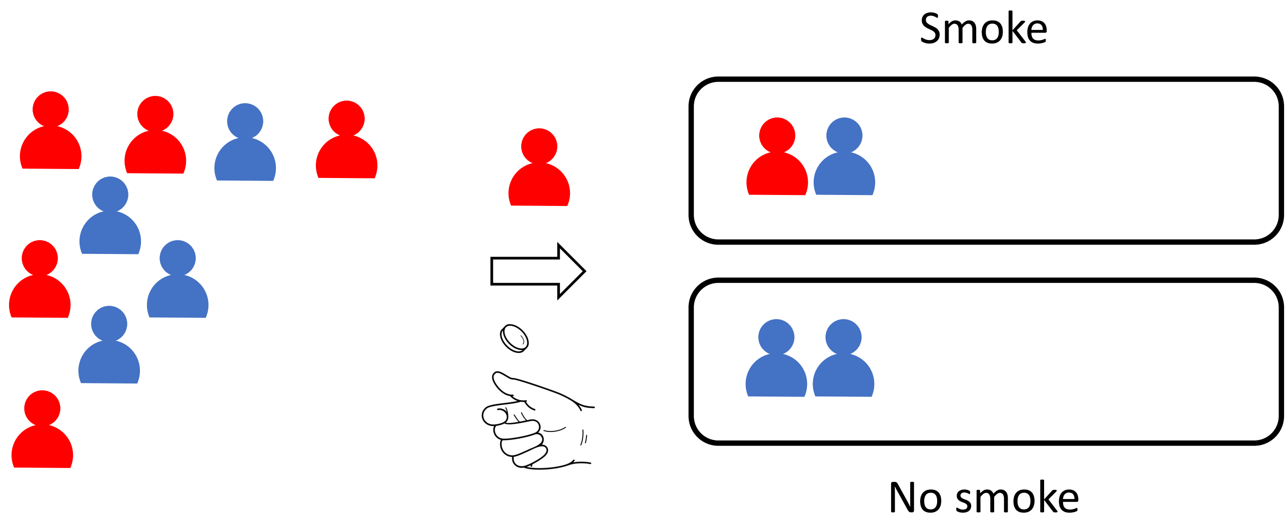
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



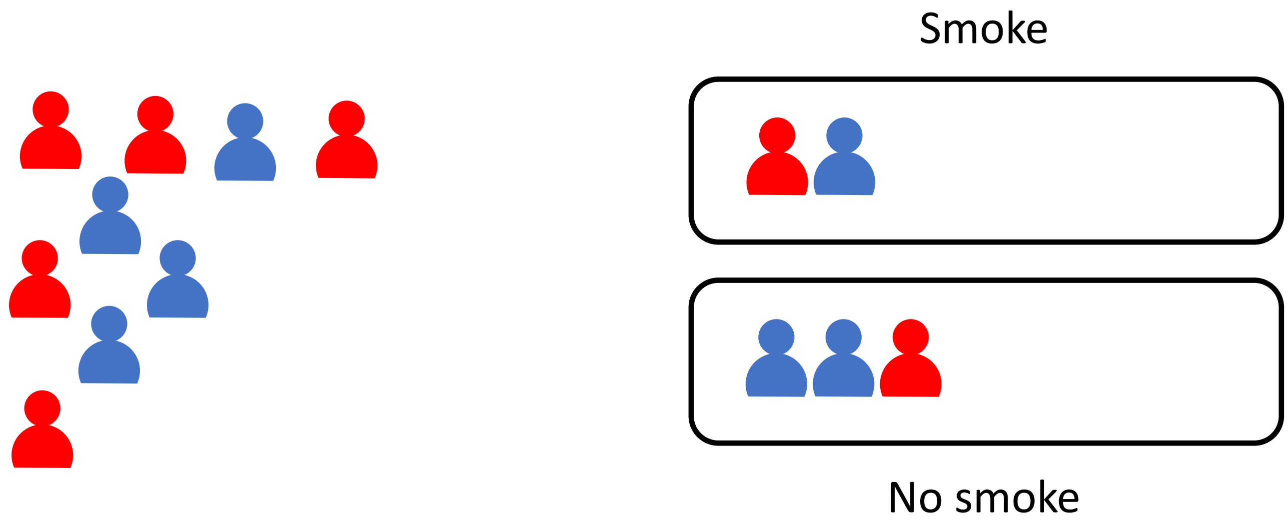
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



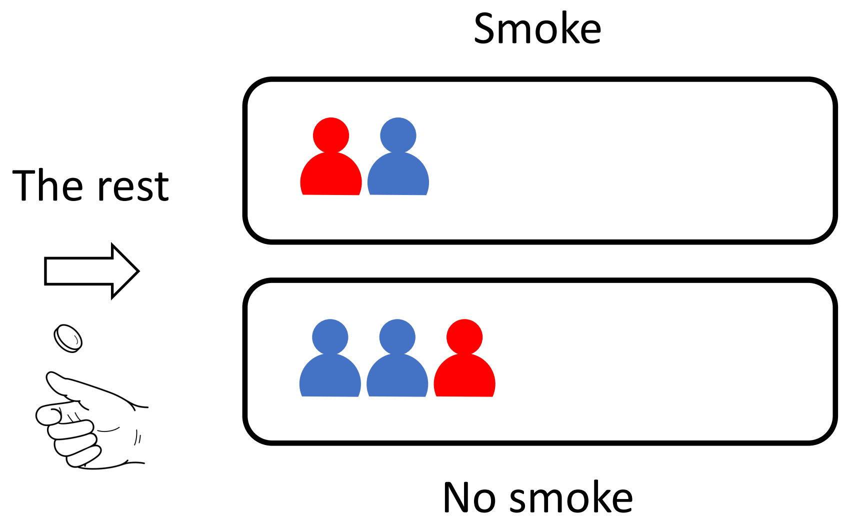
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



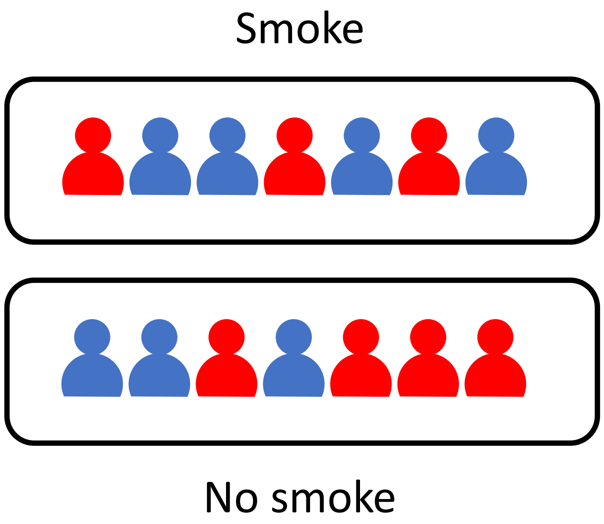
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



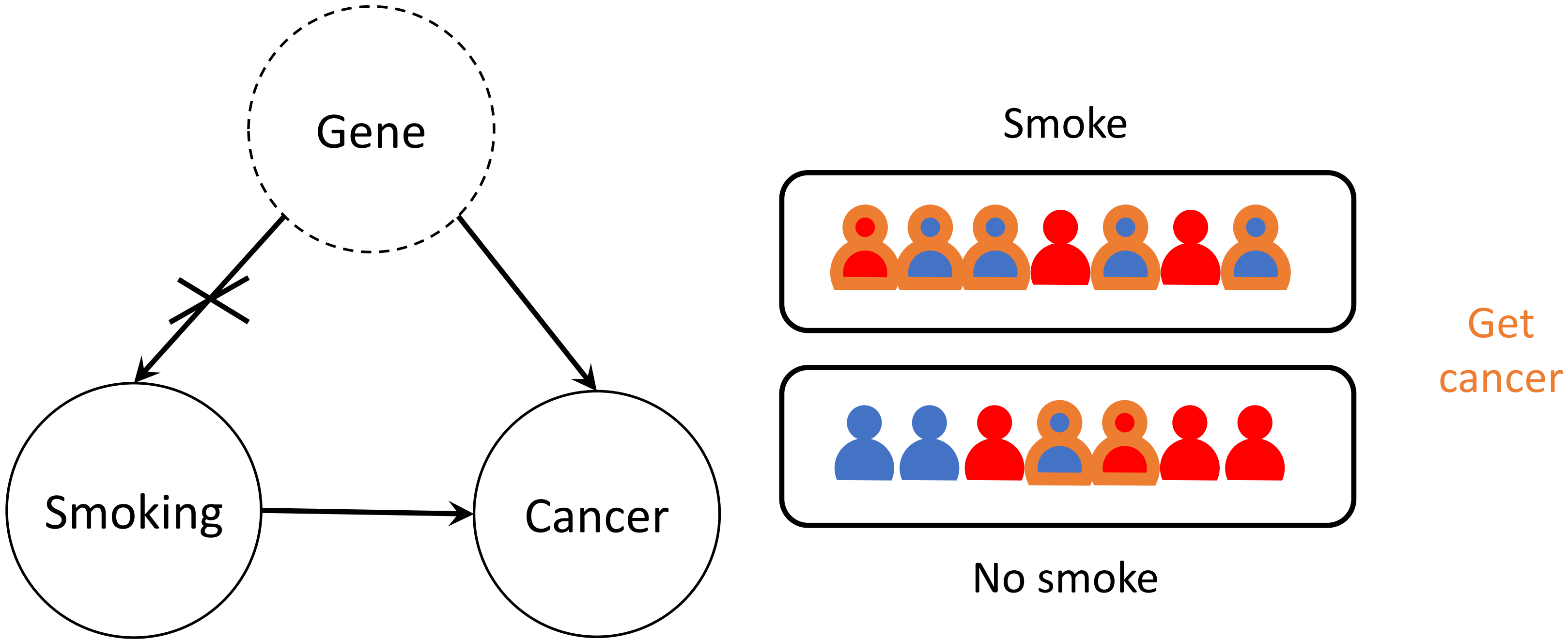
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



Randomized controlled trials

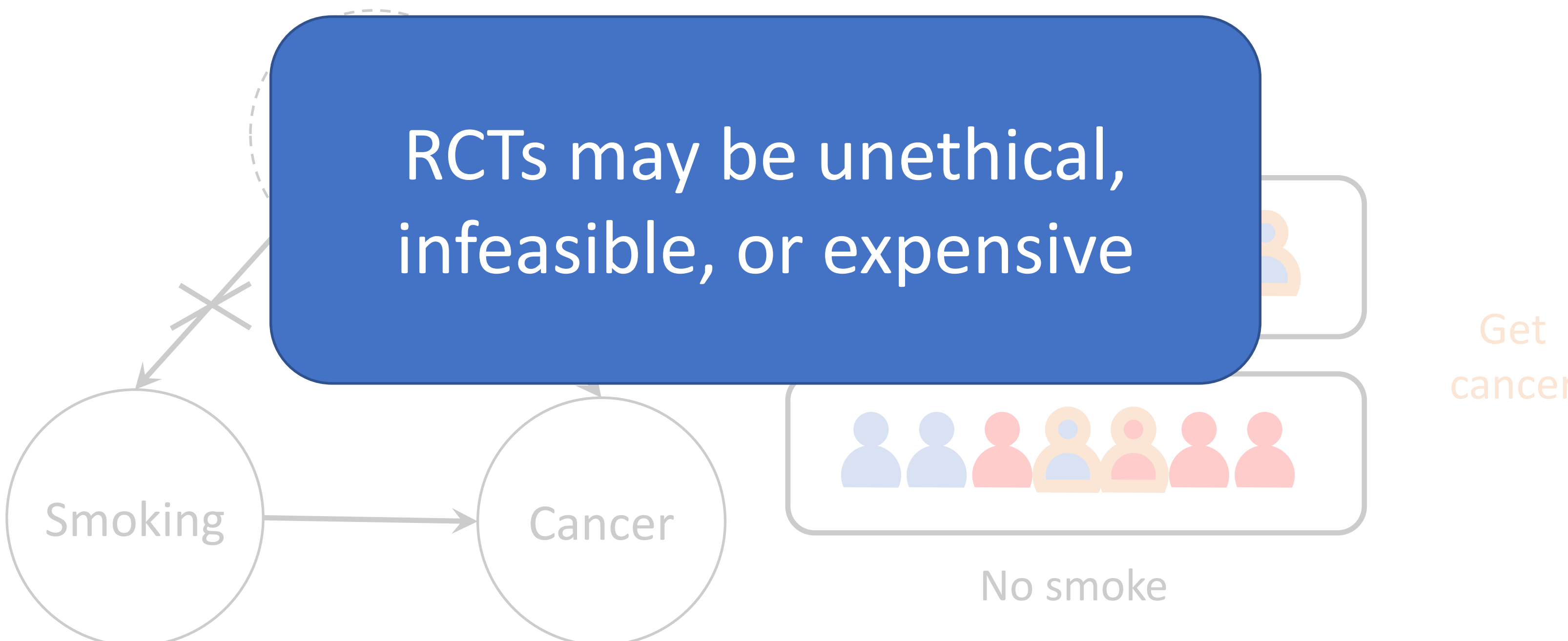
- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality



RCT removed causal link from “gene” to “smoking”
If smoking and cancer still highly correlated, then smoking causes cancer

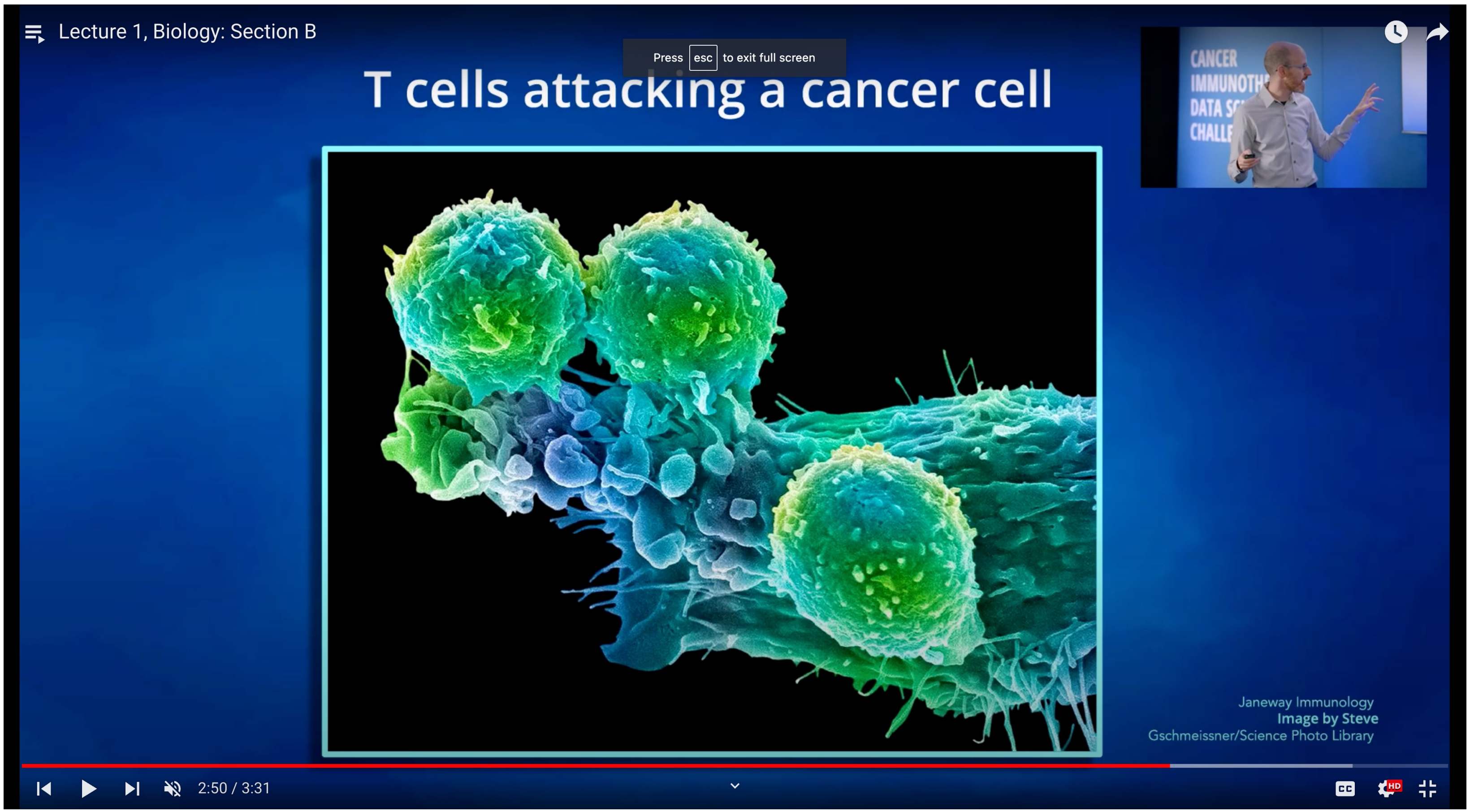
Randomized controlled trials

- Gold standard in scientific exploration
- RCTs ≡ Interventions in causality

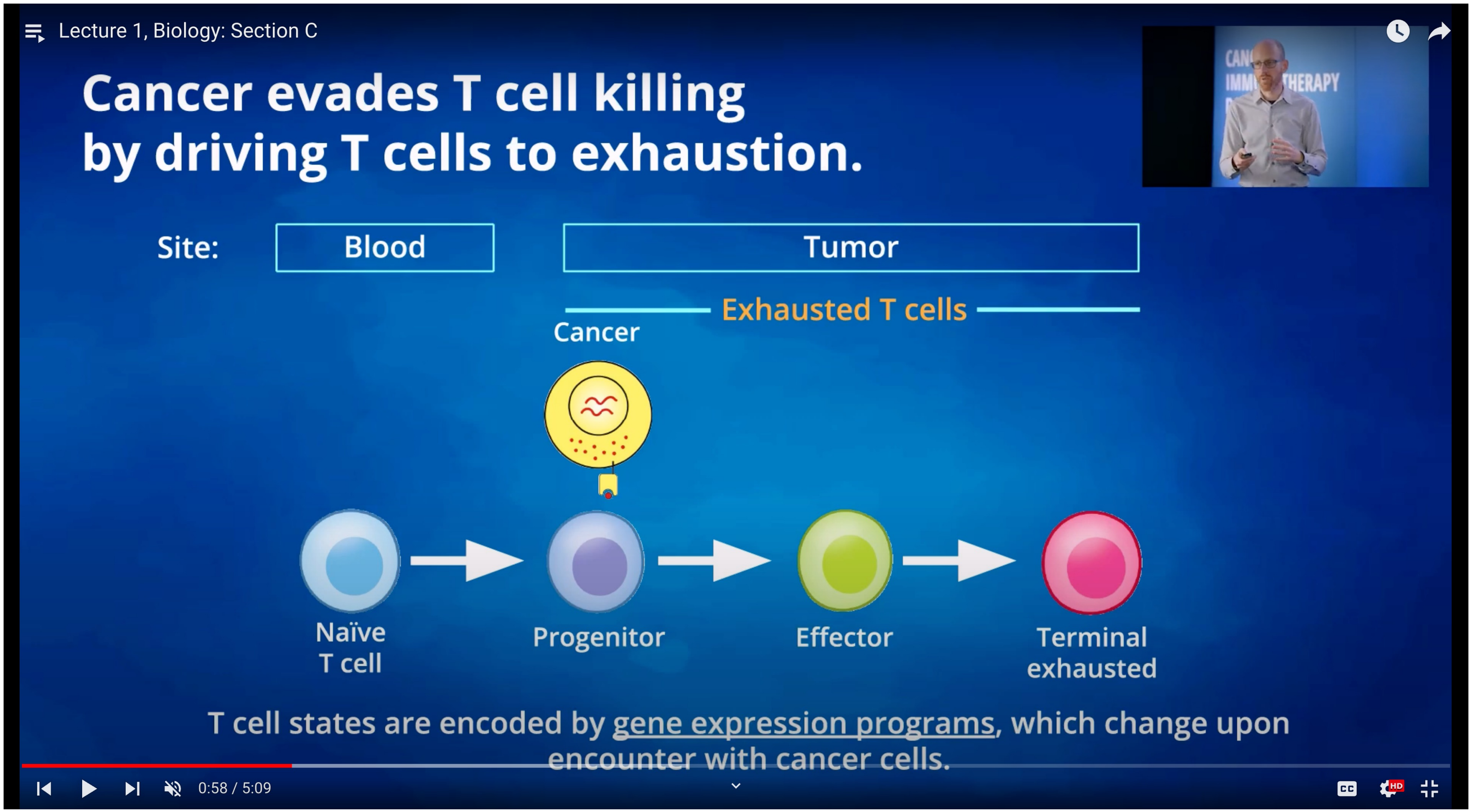


RCTs may be unethical, infeasible, or expensive

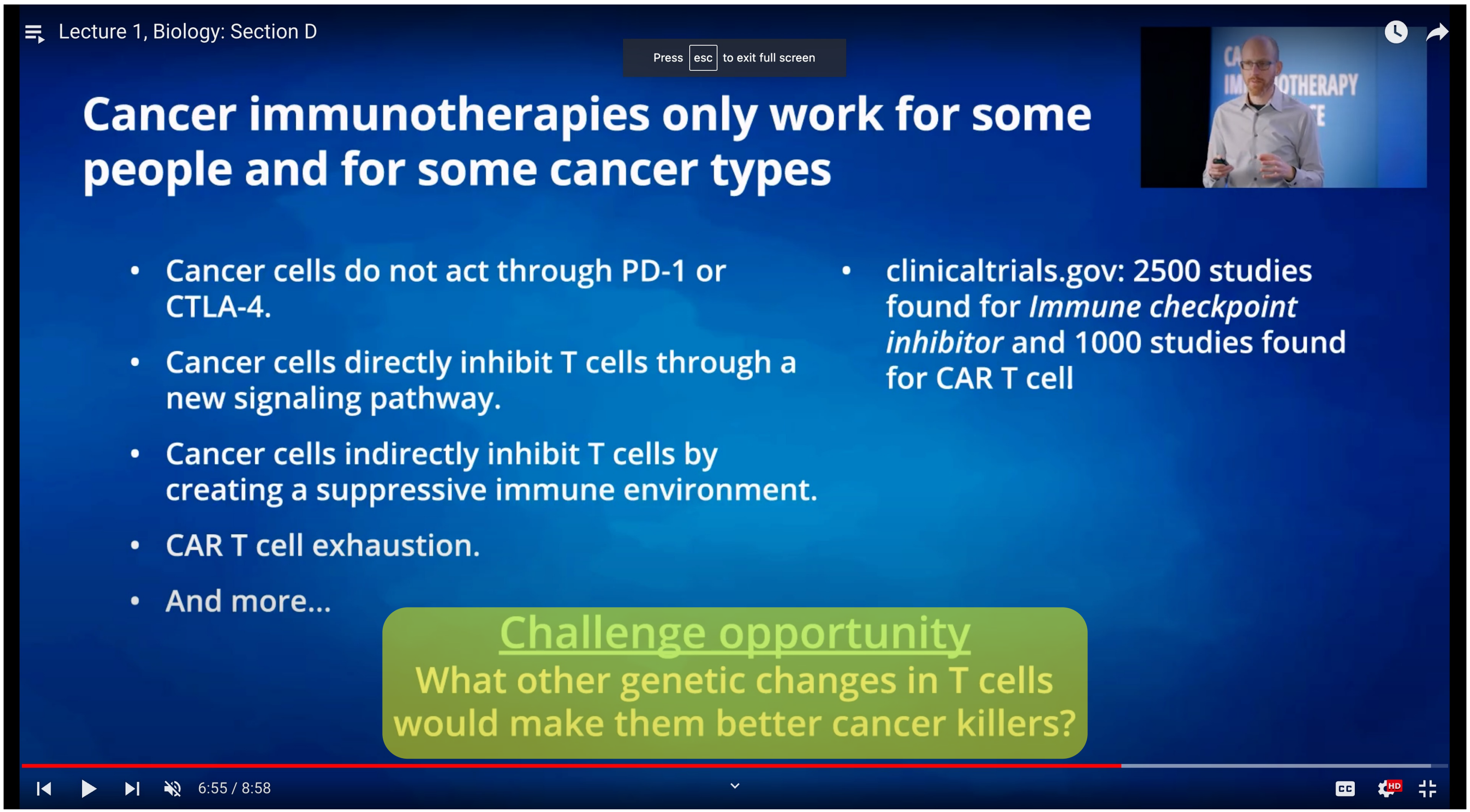
RCT removed causal link from “gene” to “smoking”
If smoking and cancer still highly correlated, then smoking causes cancer



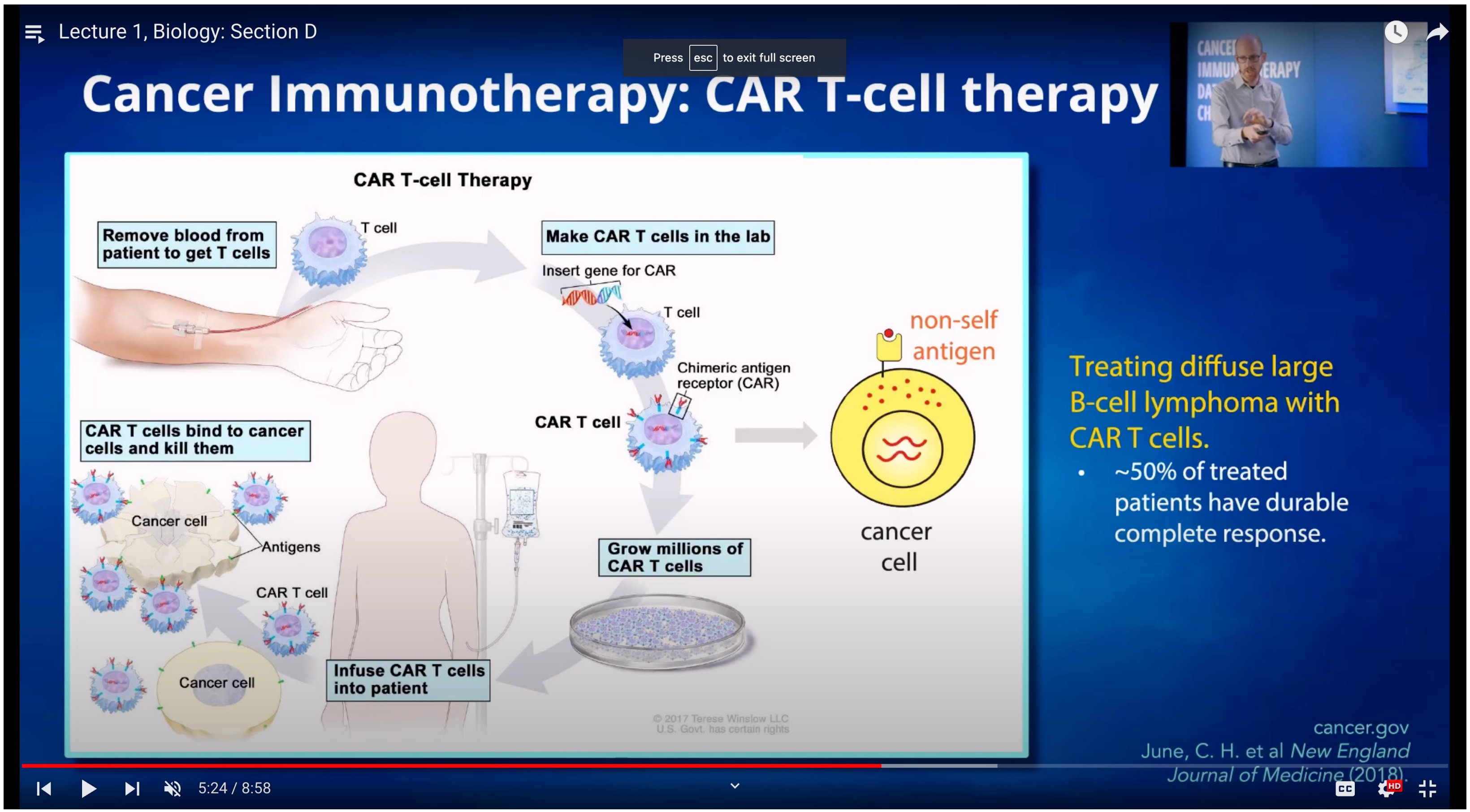
https://www.topcoder.com/challenges/25f60820-2e69-444b-bc03-490686af2c87
https://www.youtube.com/playlist?list=PL7loc09GvxoqmZOFJUc6ni5WS1FgpileN



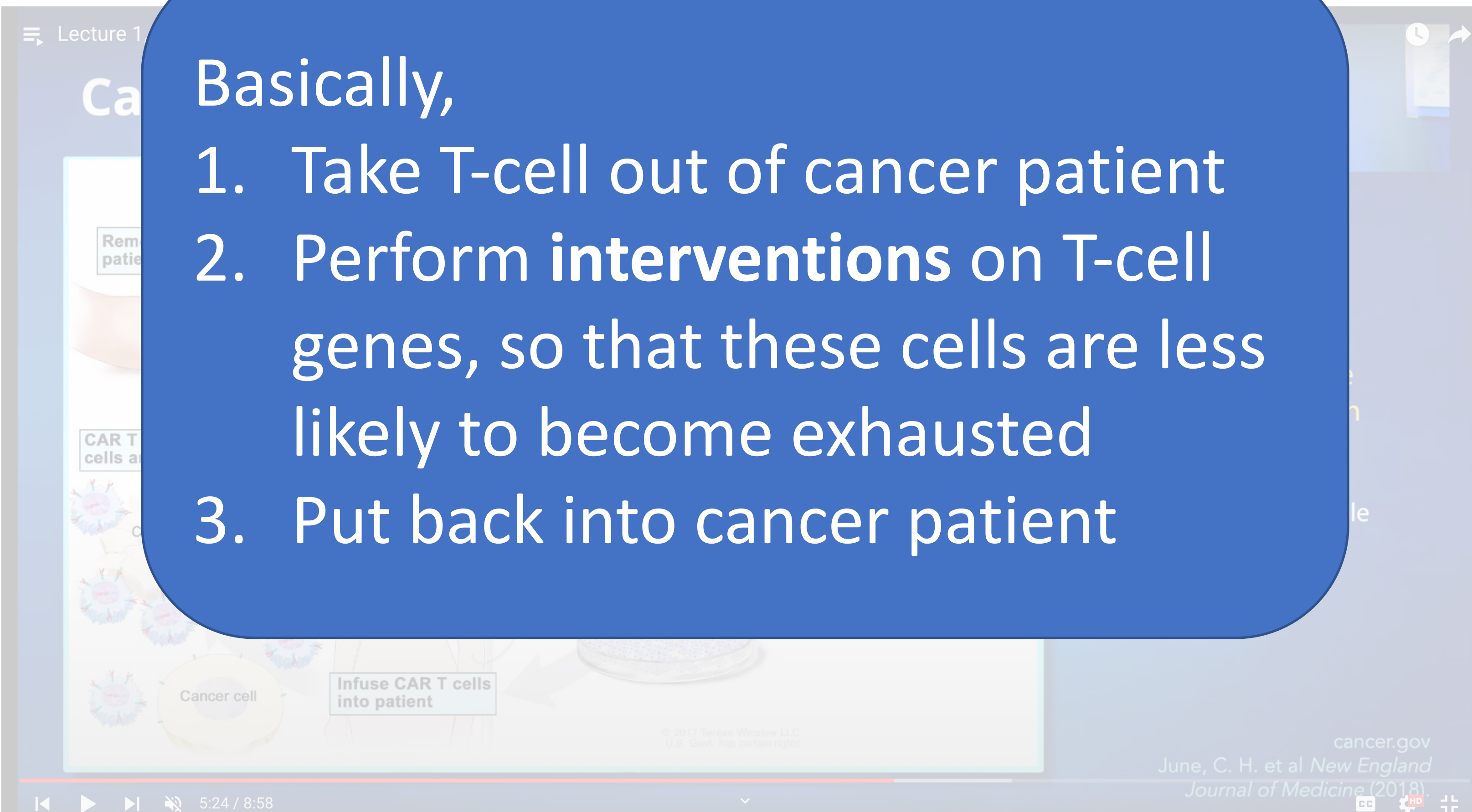
https://www.topcoder.com/challenges/25f60820-2e69-444b-bc03-490686af2c87
https://www.youtube.com/playlist?list=PL7loc09GvxoqmZOFJUc6ni5WS1FgpileN



https://www.topcoder.com/challenges/25f60820-2e69-444b-bc03-490686af2c87
https://www.youtube.com/playlist?list=PL7Ioc09GvxoqmZOFJUc6ni5WS1FgpiIeN



https://www.topcoder.com/challenges/25f60820-2e69-444b-bc03-490686af2c87
https://www.youtube.com/playlist?list=PL7Ioc09GvxoqmZOFJUc6ni5WS1FgpiIeN



Basically,

1. Take T-cell out of cancer patient
2. Perform **interventions** on T-cell genes, so that these cells are less likely to become exhausted
3. Put back into cancer patient

https://www.topcoder.com/challenges/25f60820-2e69-444b-bc03-490686af2c87
https://www.youtube.com/playlist?list=PL7loc09GvxoqmZOFJUc6ni5WS1FgpileN

NEWS | 07 October 2020

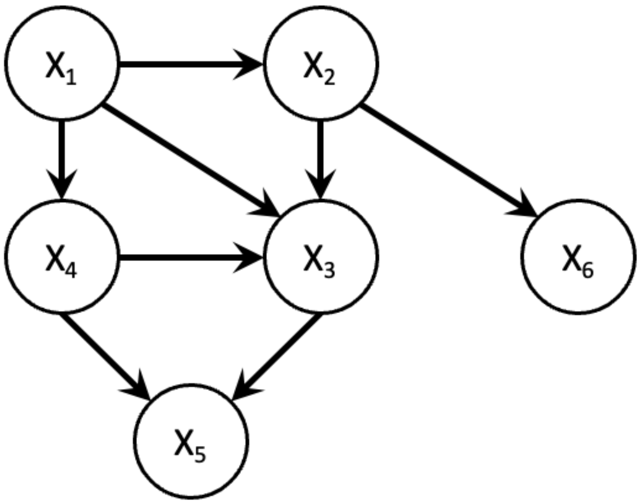
Pioneers of revolutionary CRISPR gene editing win chemistry Nobel

Emmanuelle Charpentier and Jennifer Doudna share the award for developing the precise genome-editing technology.

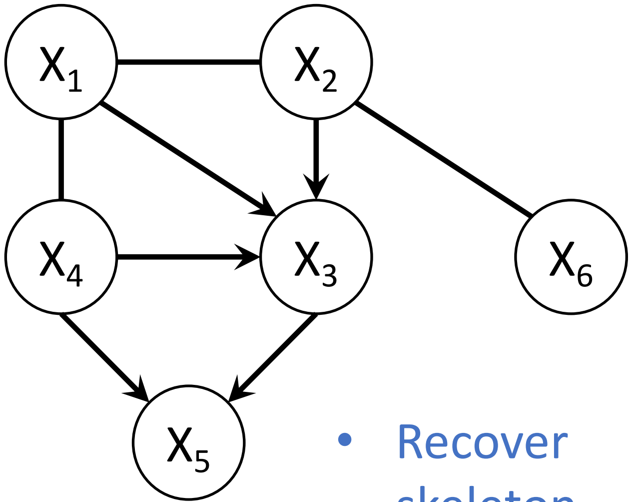
Heidi Ledford & Ewen Callaway

Jennifer Doudna and Emmanuelle Charpentier share the 2020 Nobel chemistry prize for their discovery of a game-changing gene-editing technique. Credit: Alexander Heinel/Picture Alliance/DPA

Structure learning (simplified)



This represents an equivalence class of graphs

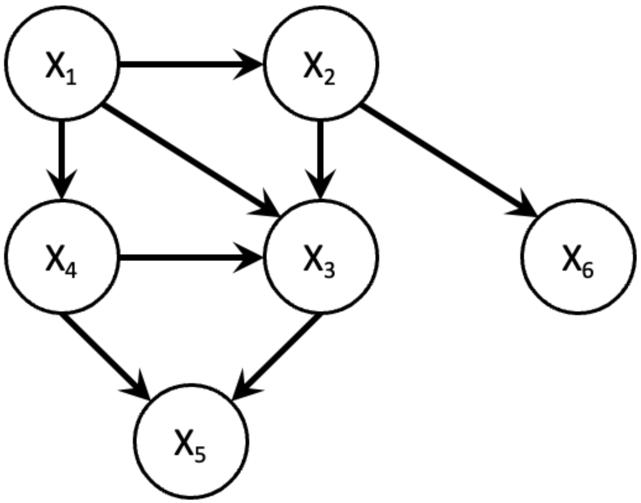


- Recover skeleton
- Orient *some* edges

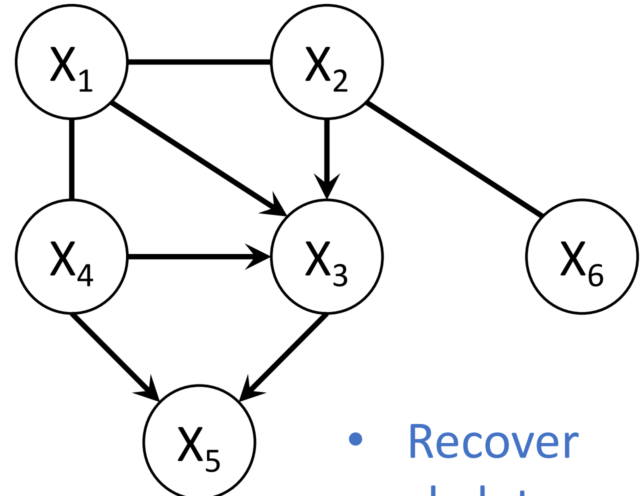
Get samples

	X_1	X_2	X_3	X_4	X_5	X_6
Sample 1	0.22	0.04	0.84	0.48	0.98	0.82
Sample 2	0.87	0.17	0.61	0.67	0.67	0.23
Sample 3	0.55	0.54	0.67	0.86	0.93	0.23
…	…	…	…	…	…	…
Sample M	0.12	0.95	0.79	0.47	0.05	0.92

Structure learning (simplified)



This represents an equivalence class of graphs



Get samples

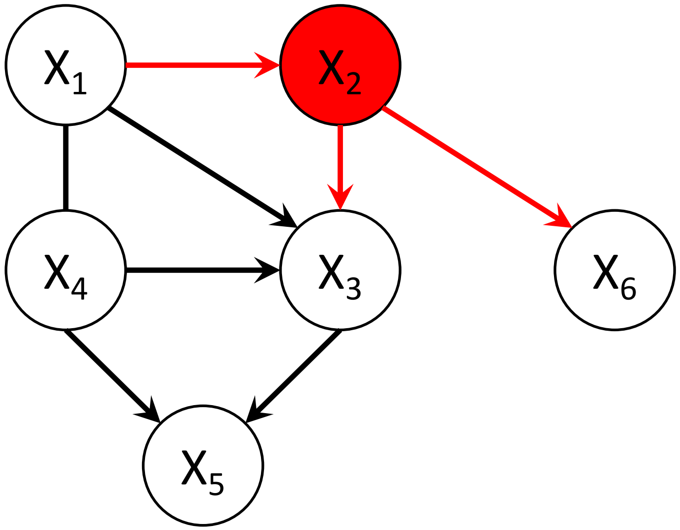
- Recover skeleton
- Orient *some* edges

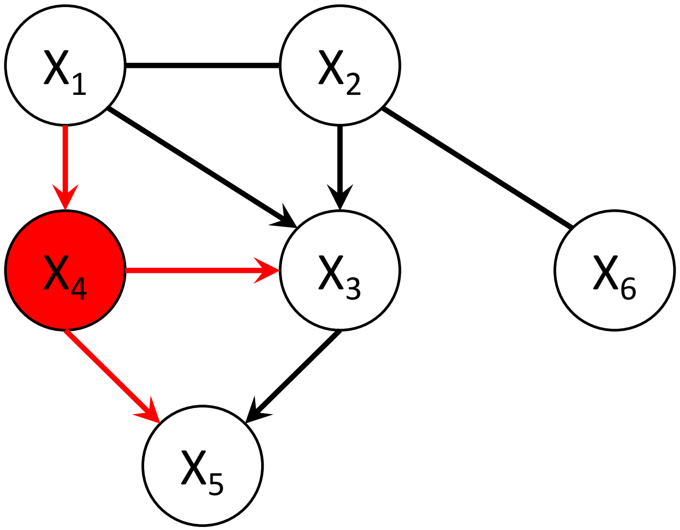
How to recover all the other arc orientations? Use interventions!

Sample M	0.12	0.95	0.79	0.47	0.05	0.92

What do interventions give us?
(*Under some causal assumptions)

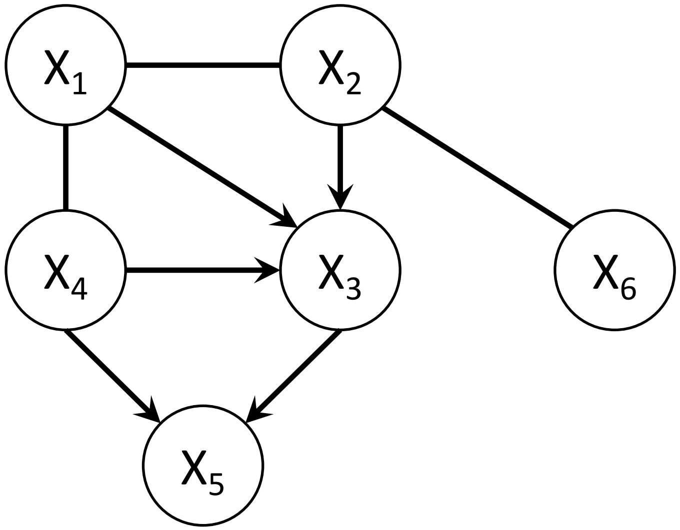
- When we intervene on a vertex, we recover the orientations of edges incident to the vertex

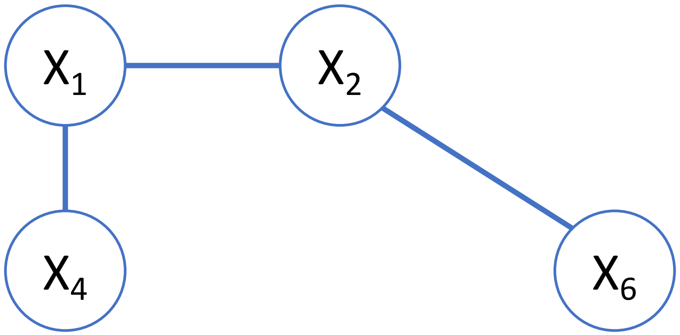




What do interventions give us?
(*Under some causal assumptions)

- When we intervene on a vertex, we recover the orientations of edges incident to the vertex



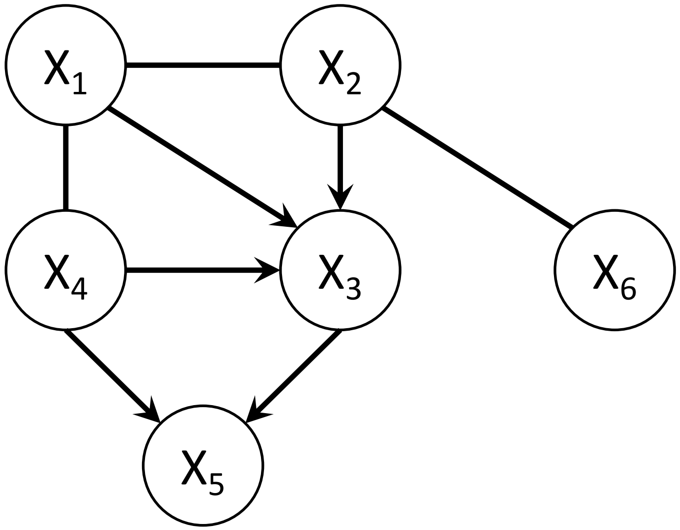


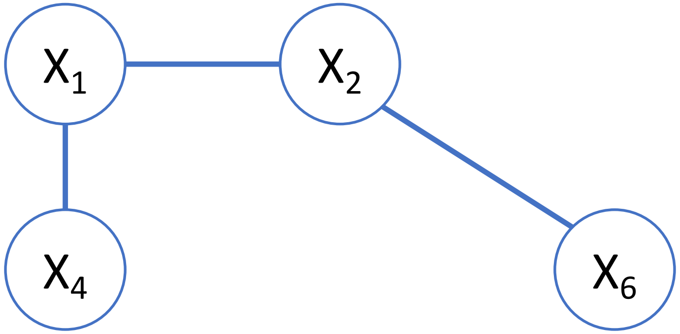
- Naïve: Compute **minimum vertex cover** on subgraph induced by unoriented arcs

What do interventions give us?

(*Under some causal assumptions)

- When we intervene on a vertex, we recover the orientations of edges incident to the vertex



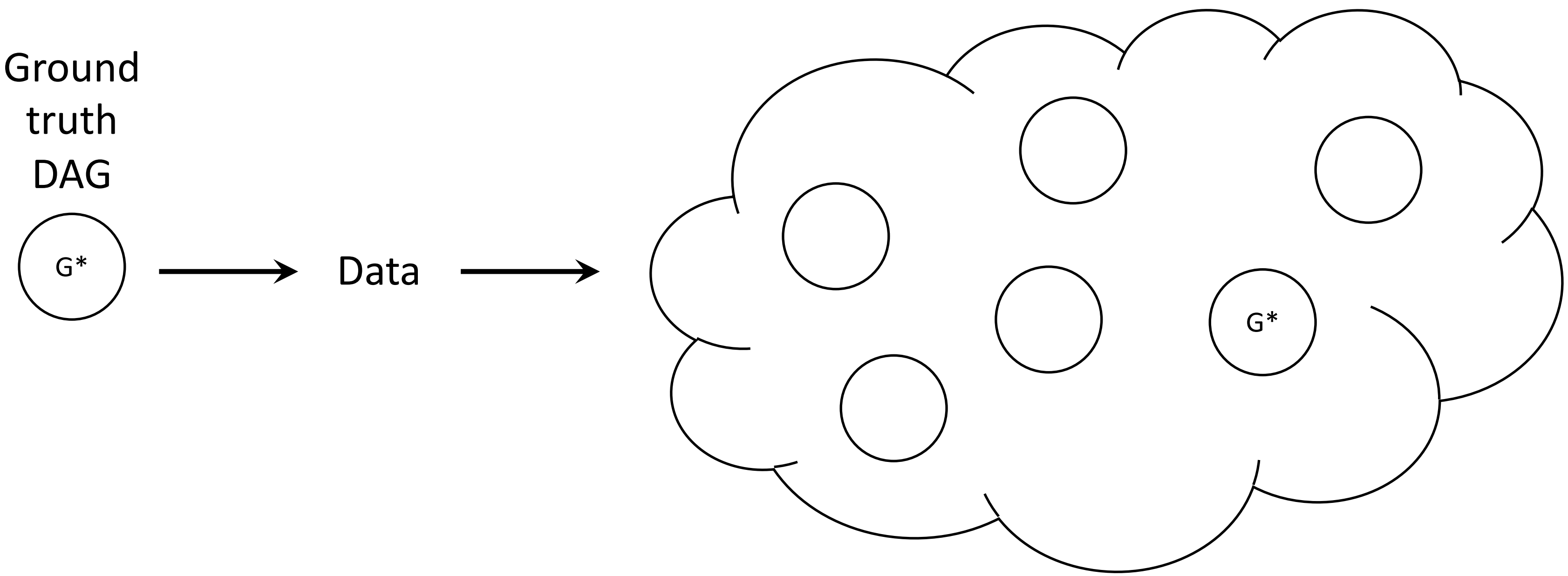


$\Omega(n)$ when the unoriented part is just a path on n nodes

- Naïve: Compute **minimum vertex cover** on subgraph induced by unoriented arcs

Searching using adaptive interventions

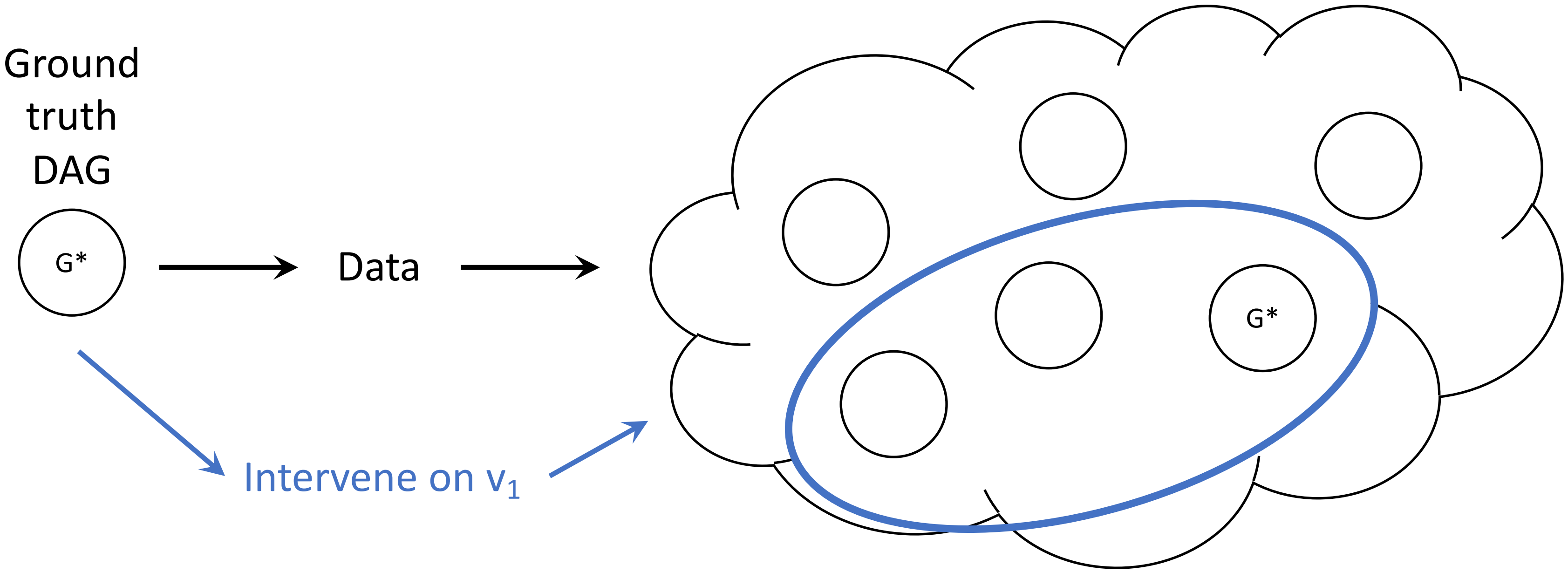
Identify G*



Equivalence class of causal graphs.
Can be represented by a partially
oriented causal graph

Searching using adaptive interventions

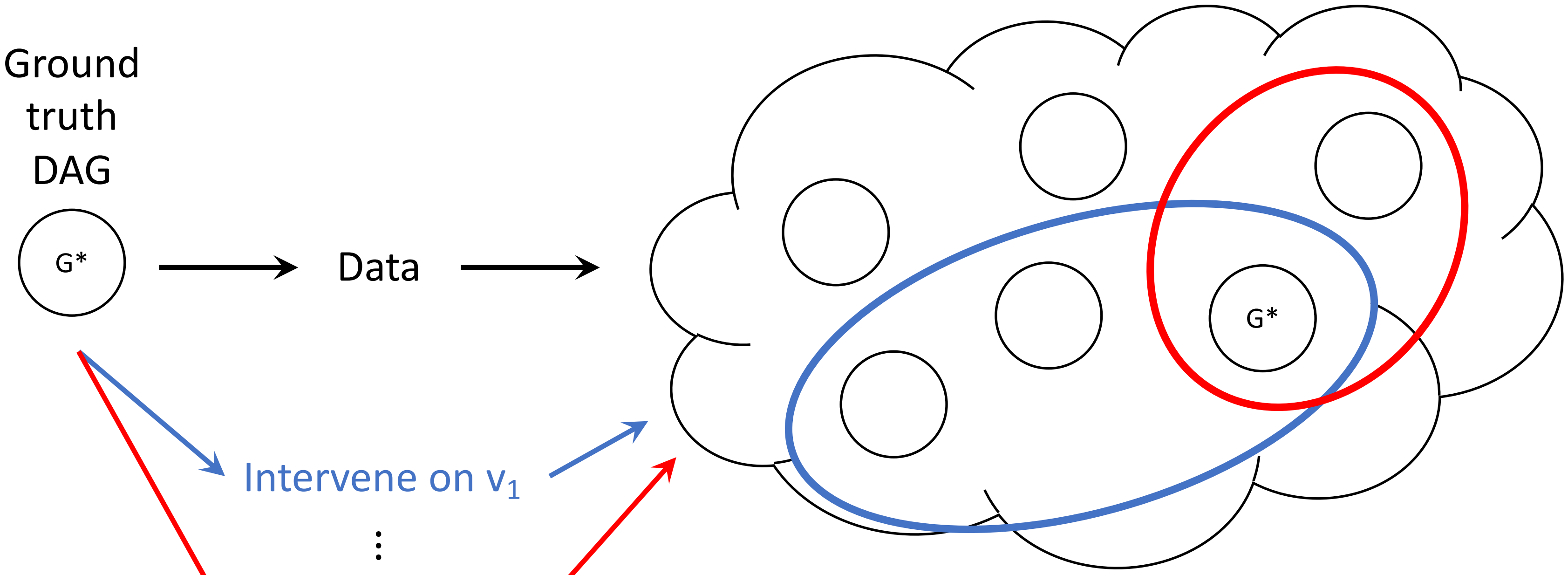
Identify G* using **interventions**



Equivalence class of causal graphs.
Can be represented by a partially
oriented causal graph

Searching using adaptive interventions

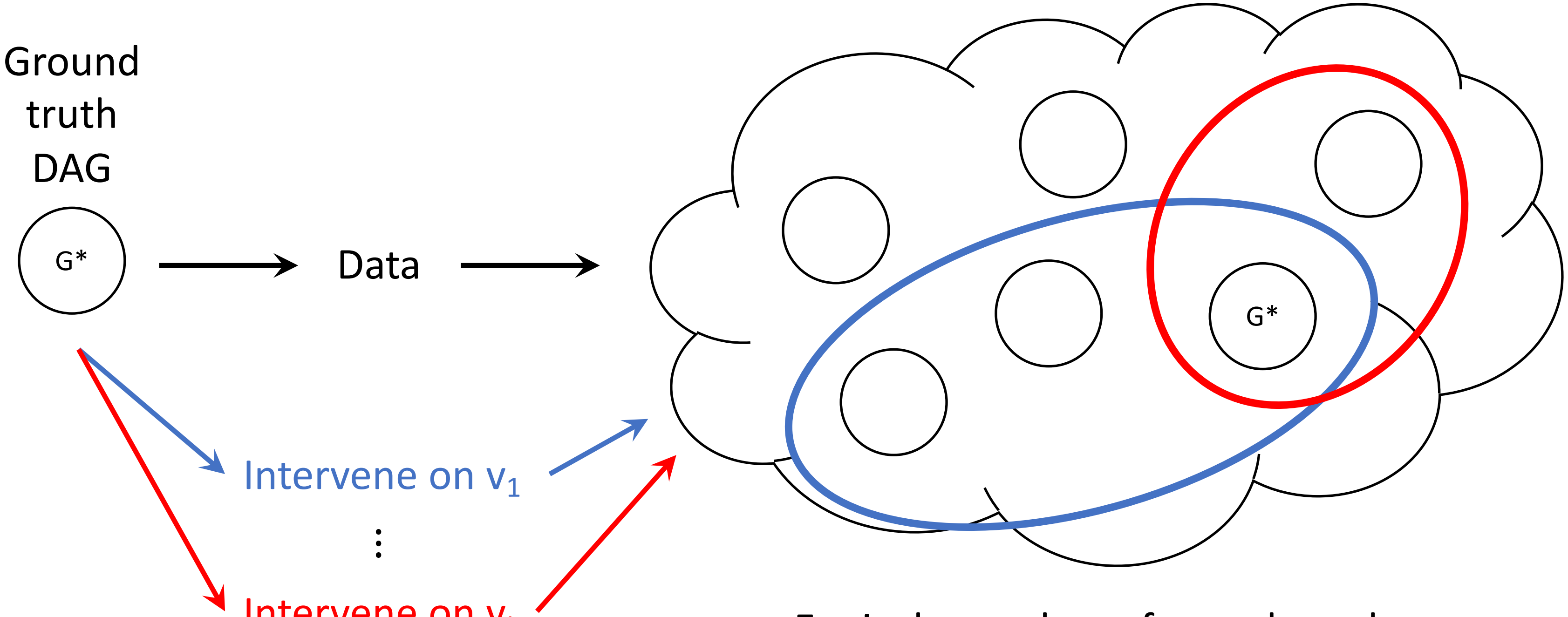
Identify G* using **interventions**



Equivalence class of causal graphs. Can be represented by a partially oriented causal graph

Searching using adaptive interventions

Identify G* using **as few interventions as possible** (minimize t)

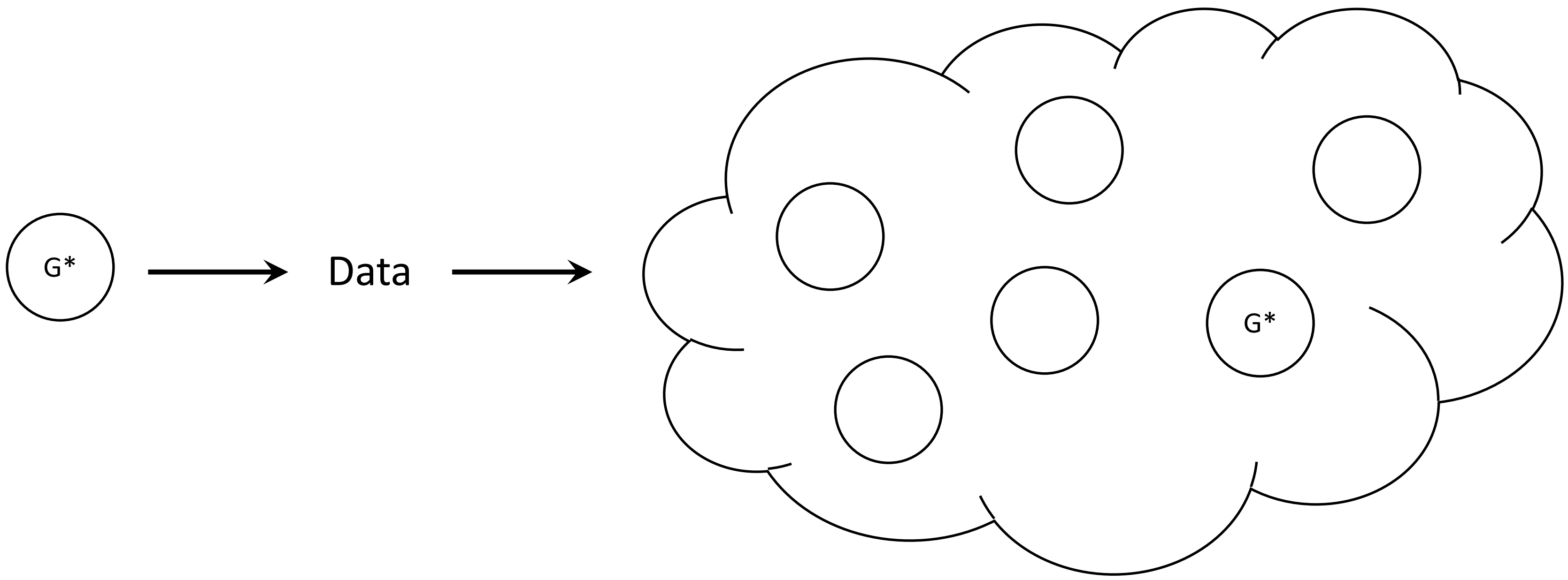


Equivalence class of causal graphs. Can be represented by a partially oriented causal graph

What is known

- Let $\nu(G^*)$ be the **minimum number of interventions** that an oracle (that *knows* G^*) needs to fully orient G^* from the partially oriented graph
- Known [Choo, Shiragur, Bhattacharyya 2022]
  - $\mathcal{O}\big(\log n \cdot \nu(G^*)\big)$ interventions suffice
  - $\Omega\big(\log n \cdot \nu(G^*)\big)$ interventions worst case necessary
- If we only care about arc orientations within a node-induced subgraph $H \subseteq G^*$ [Choo, Shiragur 2023]
  - $\mathcal{O}\big(\log |V(H)| \cdot \nu(G^*)\big)$ interventions suffice

In many problem domains…



There are domain experts!

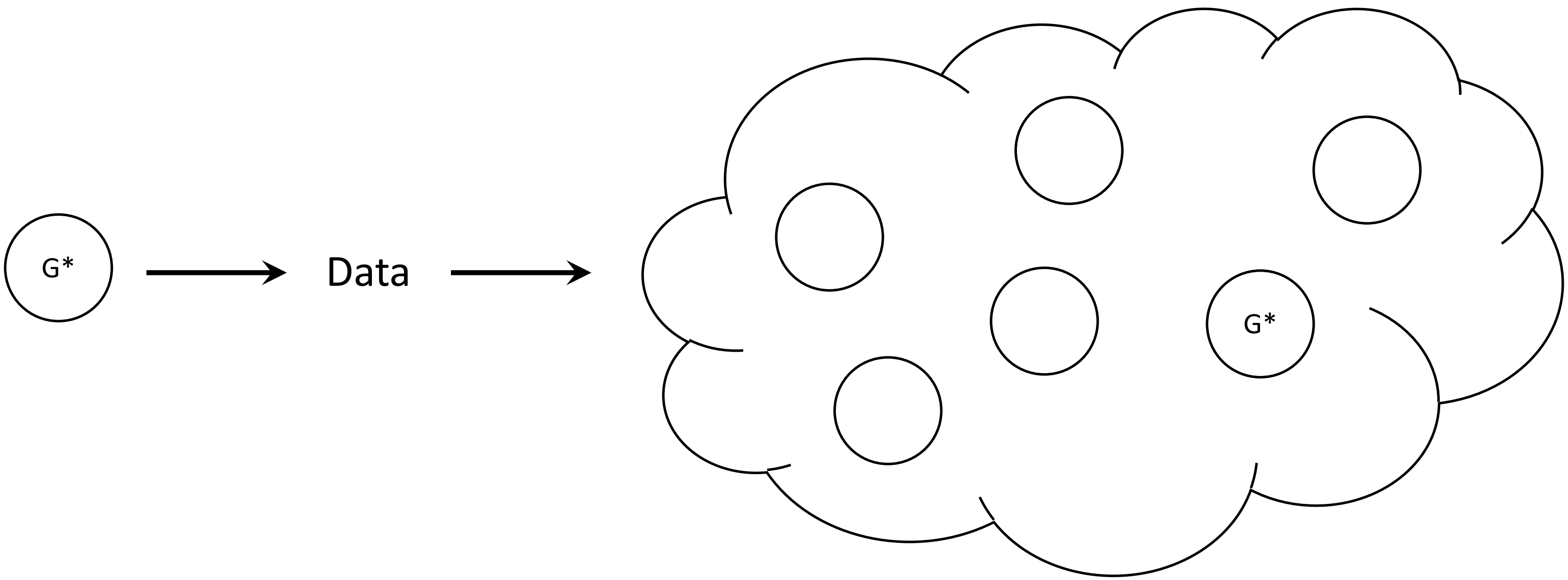


Image credit: https://thenounproject.com/icon/doctor-1285618/

There are domain experts!

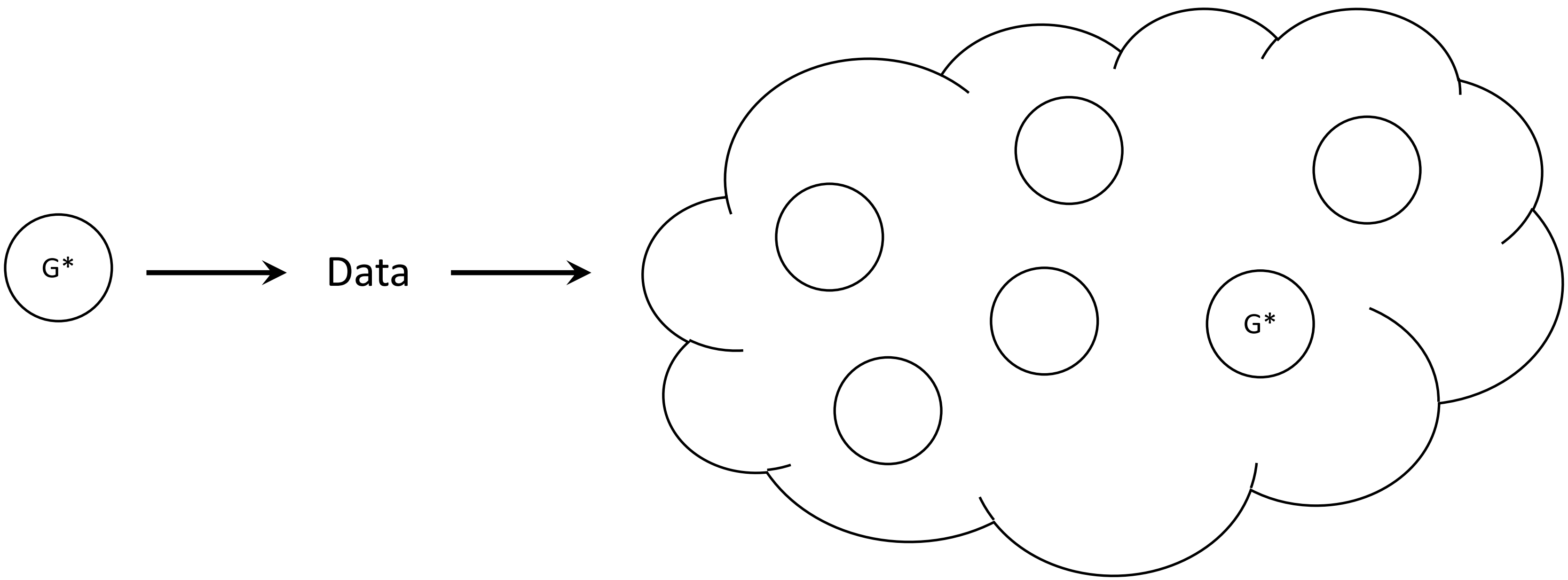
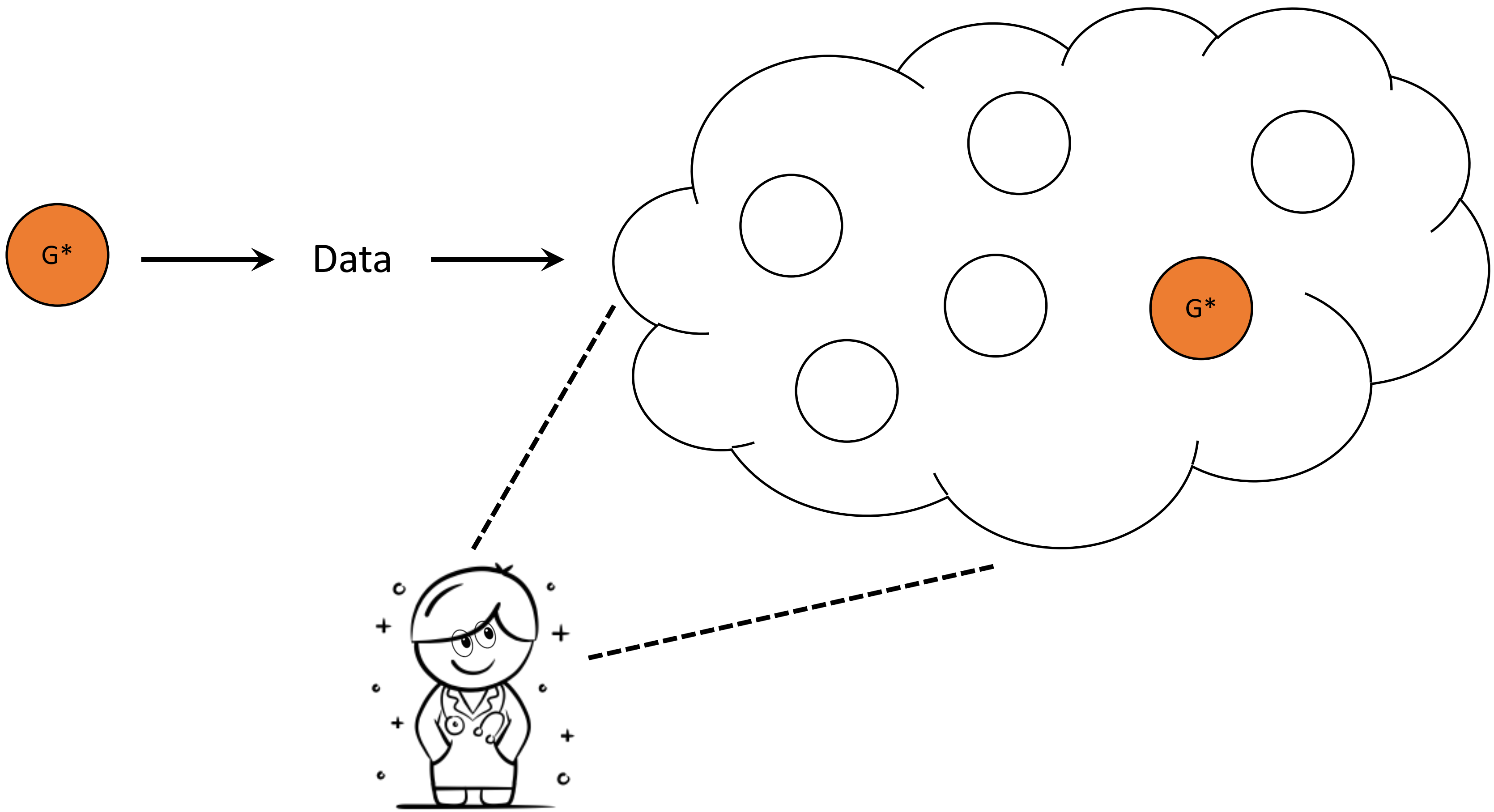
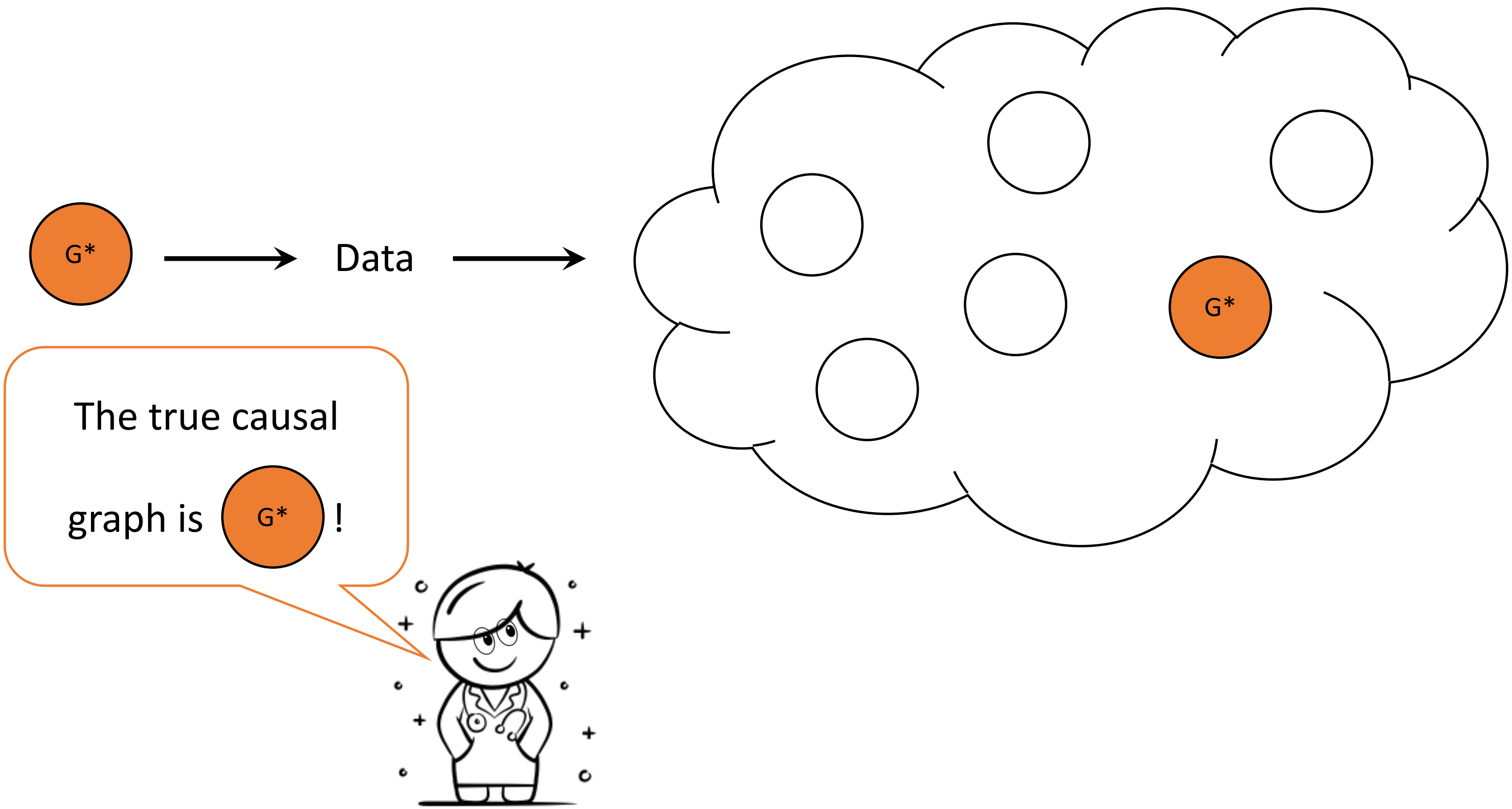


Image credit: https://thenounproject.com/icon/look-4968809/

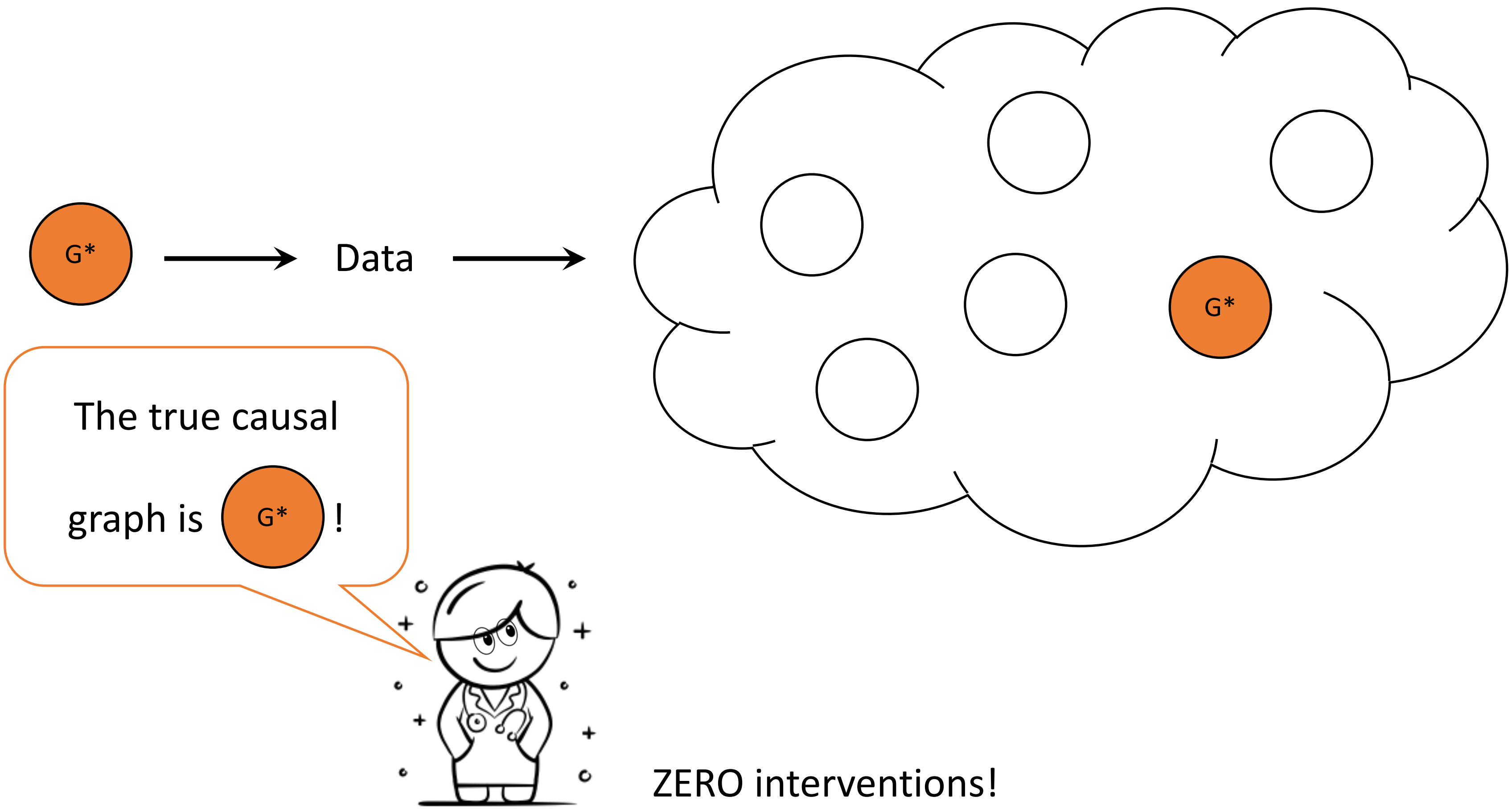
There are domain experts!



There are domain experts!

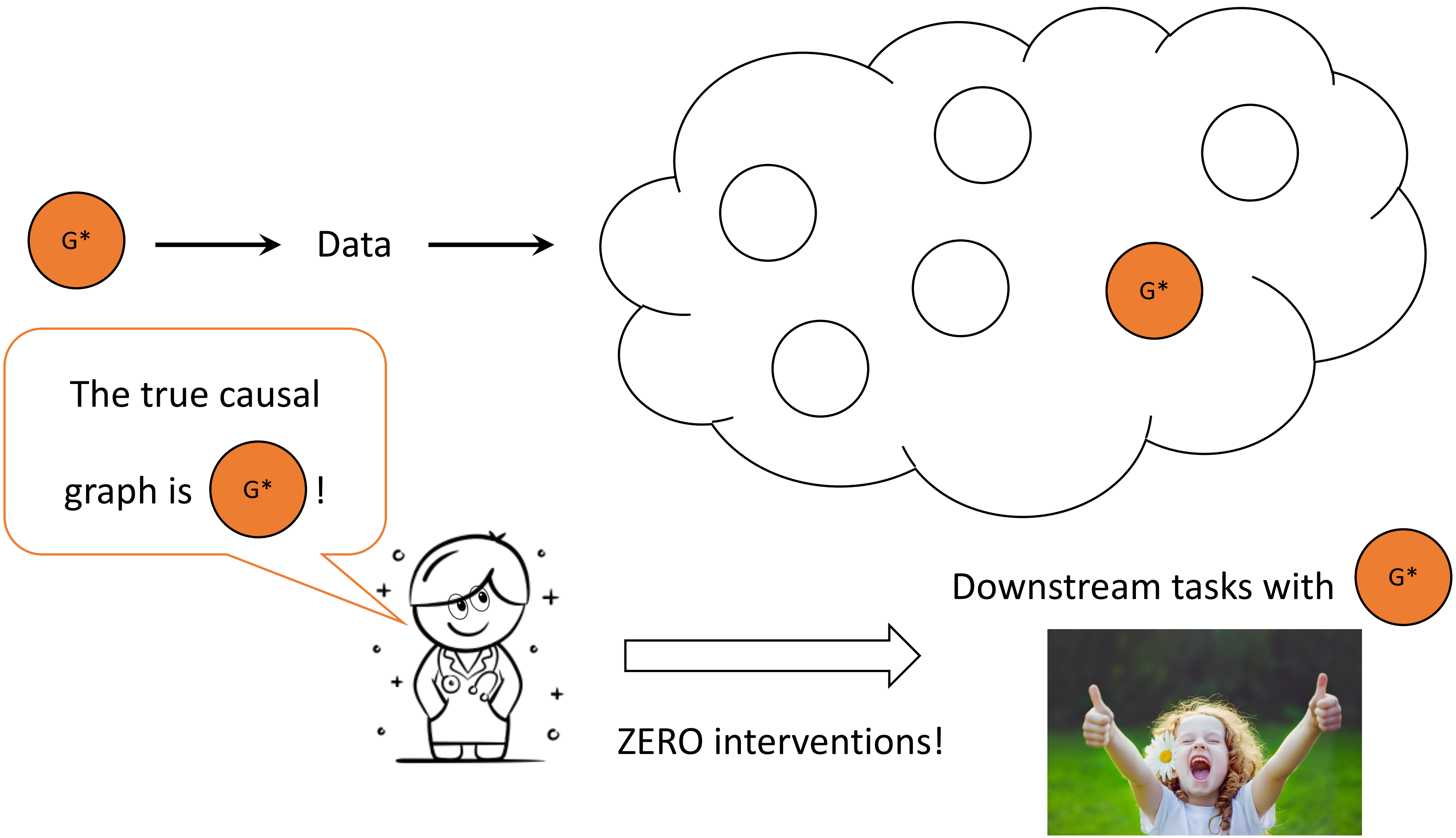


There are domain experts!

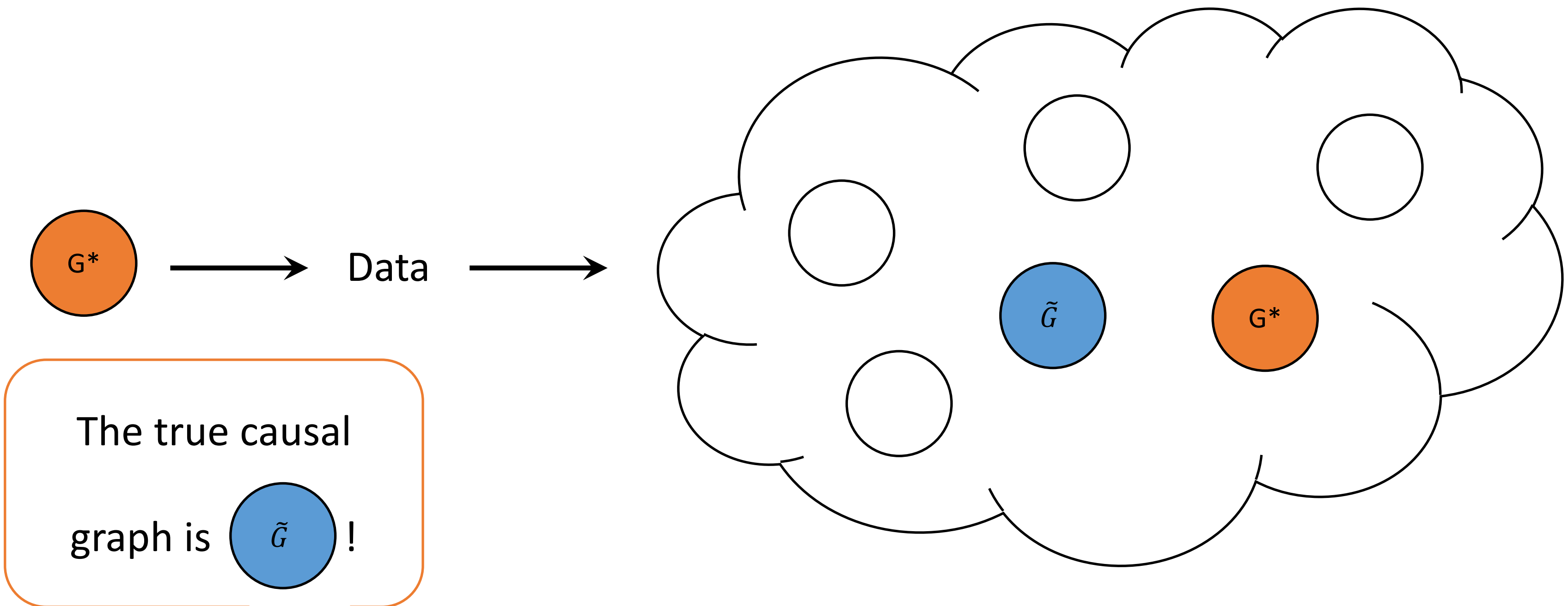


ZERO interventions!

There are domain experts!



But… experts can be wrong



ZERO interventions!

Image credit: https://dribbble.com/shots/14489872-Devil

But... experts can be wrong

G* ⟶ Data ⟶

$\tilde{G}$ G*

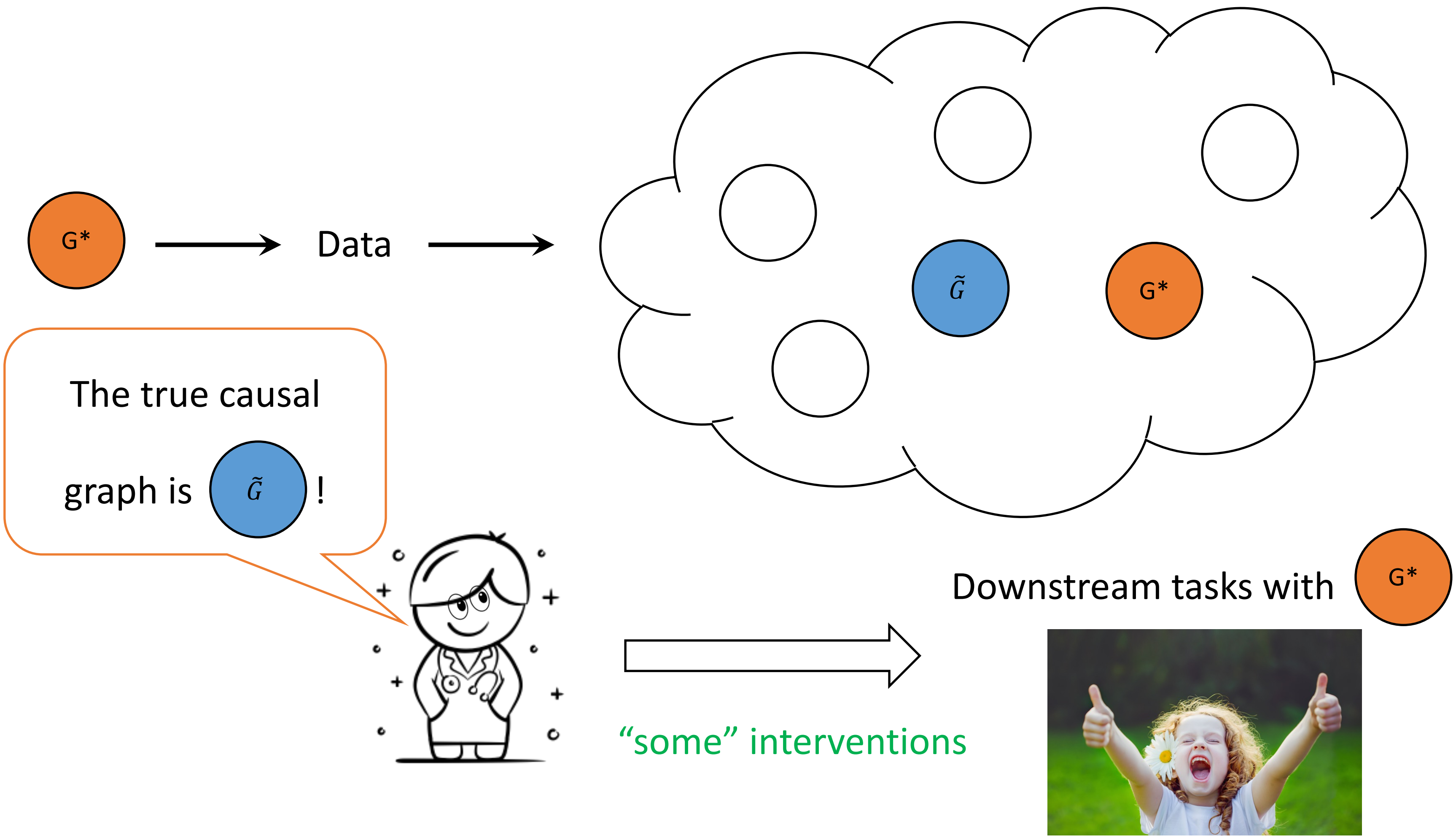
The true causal graph is $\tilde{G}$!

ZERO interventions!

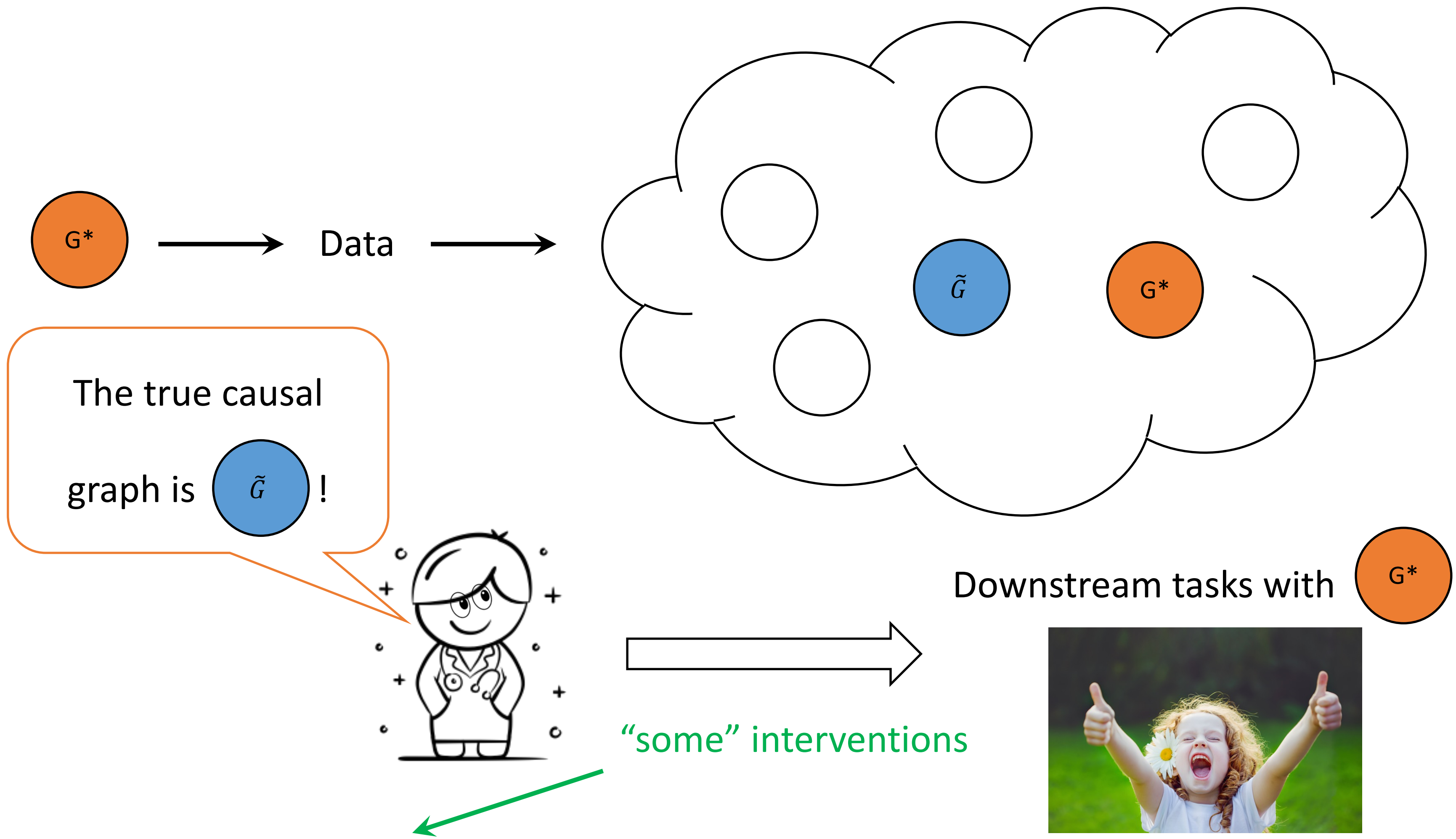
Downstream tasks with $\tilde{G}$

Image credit: https://dribbble.com/shots/3759014-Atomic-Illustrations/attachments/3759014-Atomic-Illustrations?mode=media

How to use imperfect advice?



How to use imperfect advice?

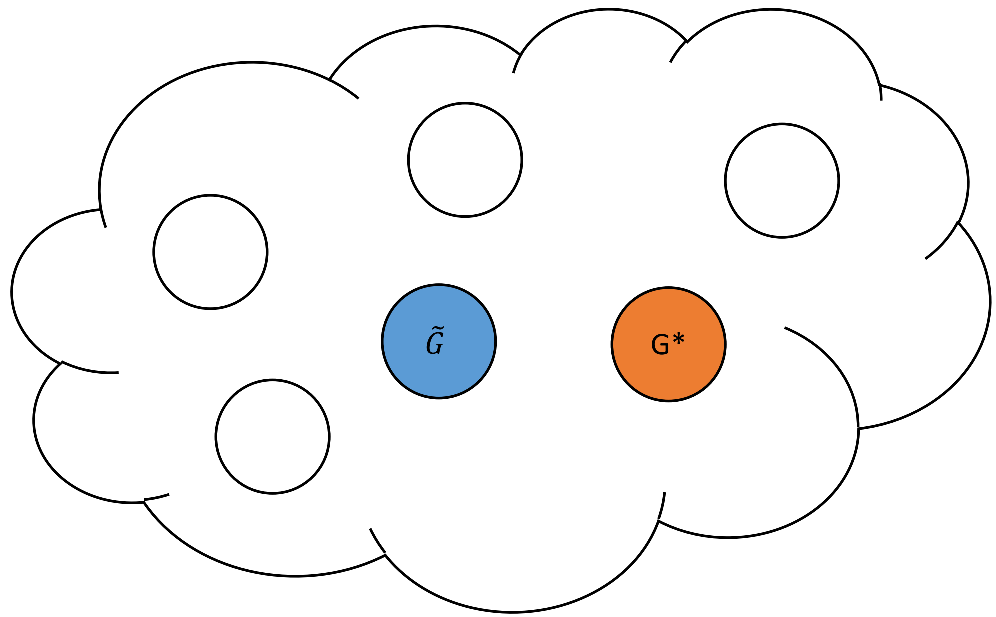


Advice search: $\mathcal{O}\left(\max\{1, \log \psi(G^*, \tilde{G})\} \cdot \nu(G^*)\right)$ interventions

[Choo, Gouleakis, Bhattacharyya 2023]

How to use imperfect advice?

G* ⟶ Data ⟶



The true causal graph is $\tilde{G}$!

Quality of advice

$0 \leq \psi(G^*, \tilde{G}) \leq n$

(Good) (Bad)

Downstream tasks with G*

"some" interventions

Advice search: $\mathcal{O}\left(\max\{1, \log \boxed{\psi(G^*, \tilde{G})}\} \cdot \nu(G^*)\right)$ interventions

[Choo, Gouleakis, Bhattacharyya 2023]

Other problem settings
(beyond causality)

Estimating Gaussians

- Given: i.i.d. Gaussian s samples from $N(0, \Sigma)$
- Output: $\hat{\Sigma}$ such that $\hat{\Sigma} \approx \Sigma$
- Goal: Minimize s required for "up to ϵ" closeness
- Known: $\tilde{\Theta}\left(\frac{n^2}{\epsilon}\right)$ samples
- Known: What if Σ is sparse?
- Known: What if Σ^{-1} is sparse?
  - Requires dependency on condition number

Estimating Gaussians


- Given: i.i.d. Gaussian s samples from $N(0, \Sigma)$
- Output: $\hat{\Sigma}$ such that $\hat{\Sigma} \approx \Sigma$
- Goal: Minimize s required for "up to ϵ" closeness
- Known: $\widetilde{\Theta}\left(\frac{n^2}{\epsilon}\right)$ samples
- Known: What if Σ is sparse?
- Known: What if Σ^{-1} is sparse?
  - Requires dependency on condition number
- Can we remove/weaken this dependency if we know a good guess of the sparsity pattern of Σ^{-1}?

https://www.comp.nus.edu.sg/cs/people/arnab/
https://www.linkedin.com/in/dimitriosimyrisiotis

Online bipartite matching

- Given: Bipartite graph G = (U, V, E)
  - Offline nodes U fixed in advance
  - Online nodes V appear one by one, with edges incident to it
- When $v \in V$ arrives, decide irrevocably whether to match v to some unmatched neighbor
- Goal: Maximize size of matching at the end
- Maximal matching: $\left(\frac{1}{2}\right)$-approximation to OPT offline
- RANKING algorithm: $\left(1 - \frac{1}{e}\right)$-approximation

Online bipartite matching

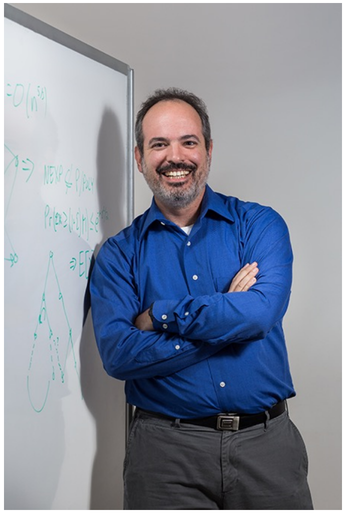
- Given: Bipartite graph G = (U, V, E)
  - Offline nodes U fixed in advance
  - Online nodes V appear one by one, with edges incident to it
- When $v \in V$ arrives, decide irrevocably whether to match v to some unmatched neighbor
- Goal: Maximize size of matching at the end
- Maximal matching: $\left(\frac{1}{2}\right)$-approximation to OPT offline
- RANKING algorithm: $\left(1 - \frac{1}{e}\right)$-approximation
- NeurIPS'22: "Final degree of $u \in U$" as advice
  - Doesn't achieve the desiderata that I mentioned earlier
  - I believe more can be done

Distributed computing

- Setting: Broadcast networks (*Simplified variant)
- Problem: When > 1 device speak, all hear garbage
- Standard solution: Exponential backoff
  - Discrete time steps
  - Each device does the following:
    - Initialize $i = 1$. Send message.
    - If detect collision
      - Pick a random delay time t between $[0,2^i - 1]$
      - Wait t steps before trying to send again.
  - When $i = \log n$, everyone starts succeeding
- Performance measure: "Throughput"

Distributed computing

- Setting: Broadcast networks (*Simplified variant)
- Problem: When >1 device speak, all hear garbage
- Standard solution: Exponential backoff



What if we have a good guess of how many devices are in the network?

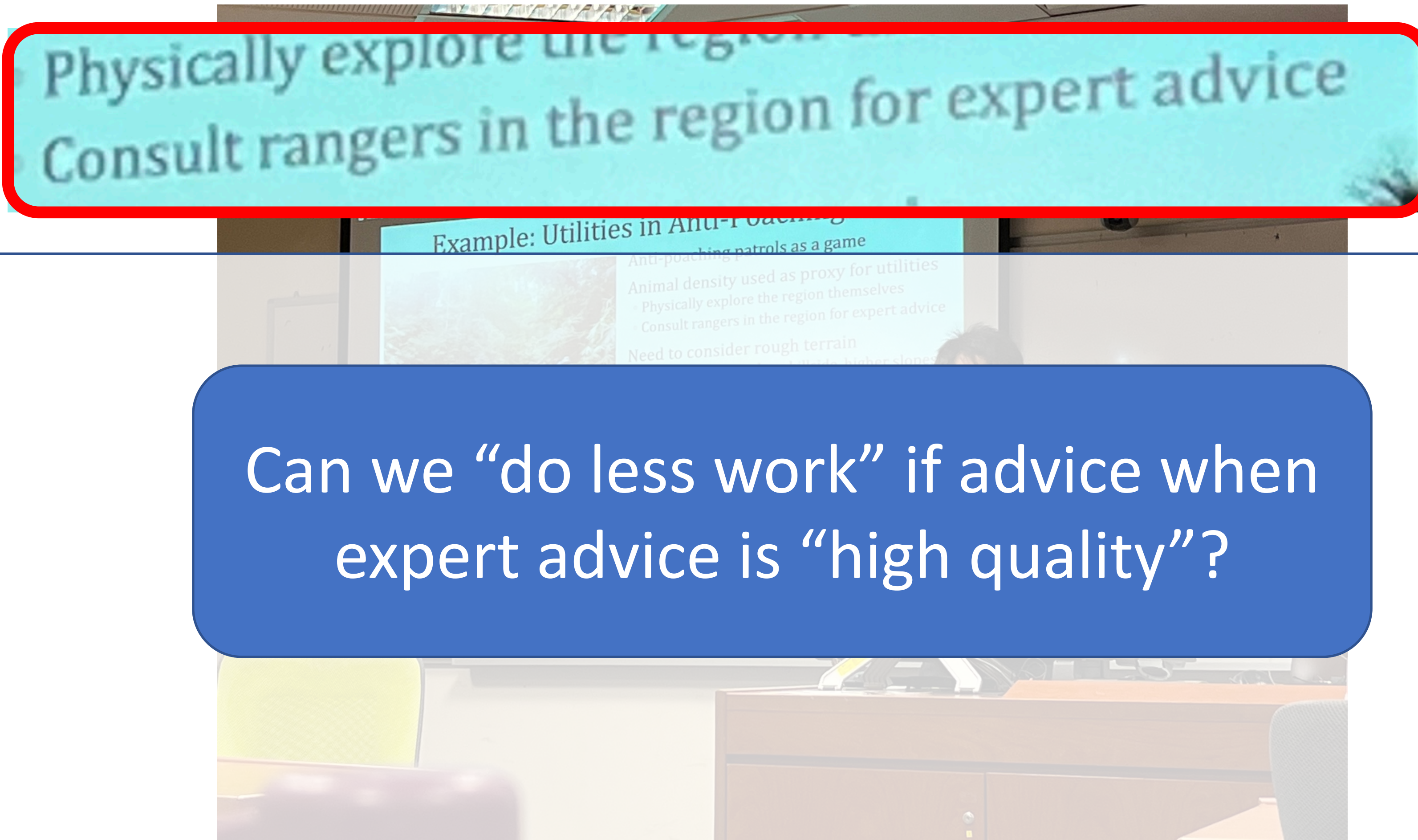
We thought about it back in 2021 (along with some other distributed computing problems where imperfect advice may make sense), had some partial ideas, but did not manage to work out all the details… ☹

(Let us know if you have ideas and are keen on working out the details together)

https://www.comp.nus.edu.sg/cs/people/gilbert/

Inverse game theory

Inverse game theory



Can we “do less work” if advice when expert advice is “high quality”?

The results mentioned in this talk are based on joint work with

Arnab Bhattacharyya

Themis Gouleakis

Kirankumar Shiragur

Let’s talk!

What other problems are amenable to imperfect advice in your problem domain?