Inferences from Causal Selection Explanation

Nicolas Navarre, Can Konuk, Neil Bramley, Salvador Mascarenhas

July 26, 2024

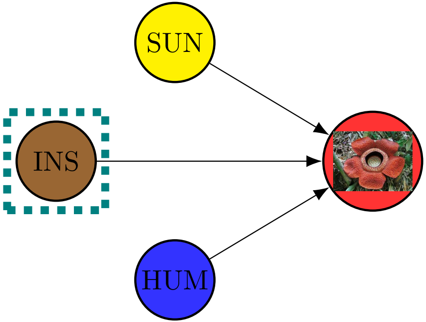
The problem of causal abduction

- Say your new exotic plant starts to wither as the summer season comes. Is this because it dislikes
  - The sun ?
  - The humidity?
  - The insects coming out with the new season?
  - The other plants newly growing around it?
  - Any combination of the above?

Figure: A plant

Causal selection judgments

- An expert explanation: 'The exotic plant withered because of the insects'
- Singles out one salient causes amongst all of the factors that might have influenced the outcome: *causal selection*.



;

Figure: An uncle's advice

Counterfactual models

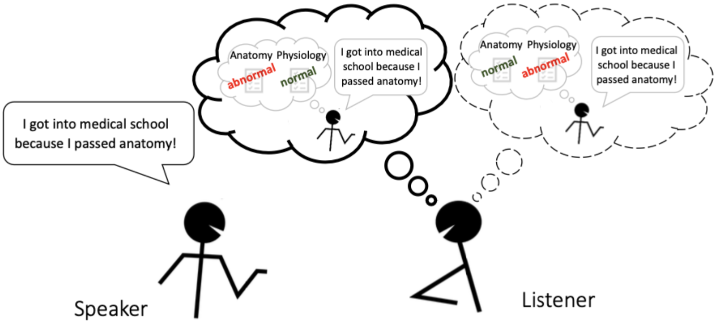
Quillien and Lucas (2023); Icard et al. (2017)

- Our best theories of such judgments (Icard et al. (2017); Quillien and Lucas (2023)) rely on counterfactual dependence:

  1. Sample a number of counterfactual worlds to the present situation, where the probability for each world to feature a certain event depends on its normality, and the outcome on one's causal knowledge.
  2. Look at the co-variation between events of interest and the outcome across counterfactuals.

Causal Inferences from Explanation

Previous Study: Kirfel et al. (2022)



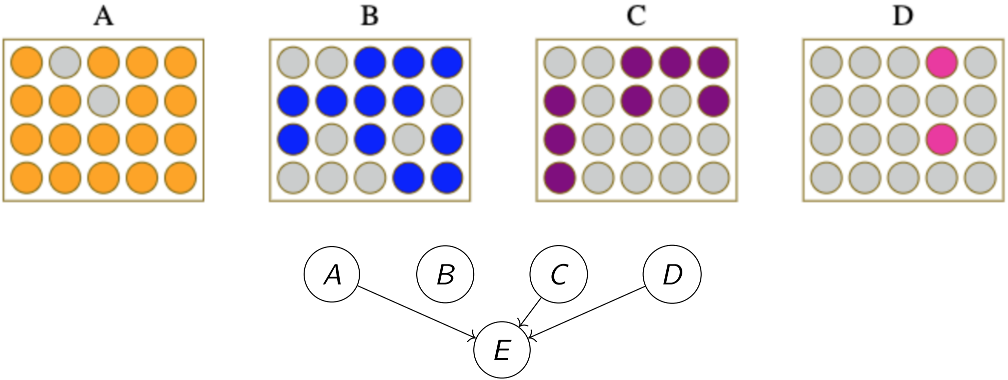
Limitations of Kirfel et al. (2022)

Suppose that A is an abnormal variable, B is a normal variable.

$A \wedge B \longleftarrow$ "Because of A"

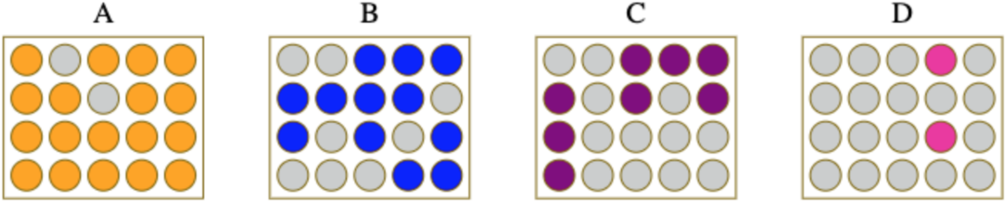
$A \vee B \longleftarrow$ "Because of B"

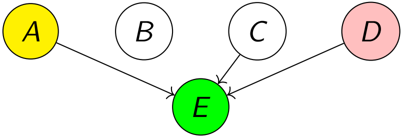
The present study



- Subjects wins and losses are determined as a function of the rule $WIN \longleftarrow (A \wedge D) \vee C$
- Draws from each urn can have **coloured** or **uncoloured** balls, allowing for a total of $2^{16} = 65,536$ different possible rules.

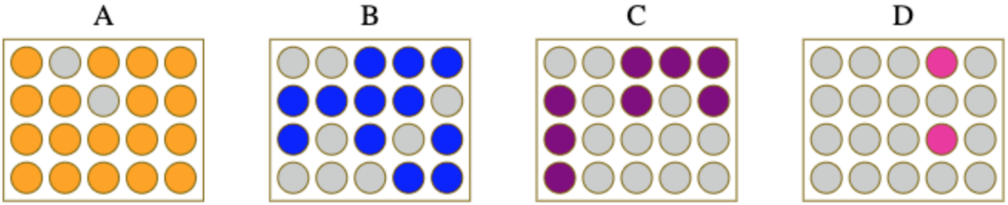
The present study

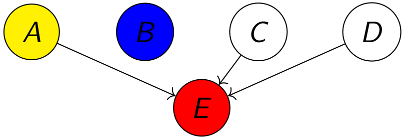




- Subjects wins and losses are determined as a function of the rule $WIN \longleftarrow (A \wedge D) \vee C$
- Draws from each urn can have **coloured** or **uncoloured** balls, allowing for a total of $2^{16} = 65,536$ different possible rules.

The present study

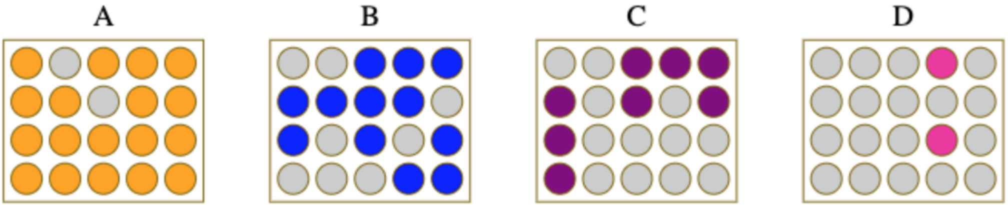




- Subjects wins and losses are determined as a function of the rule $WIN \longleftarrow (A \wedge D) \vee C$
- Draws from each urn can have **coloured** or **uncoloured** balls, allowing for a total of $2^{16} = 65,536$ different possible rules.

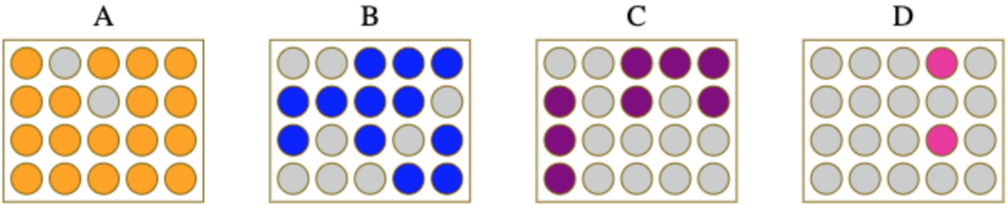
Design

Three conditions:

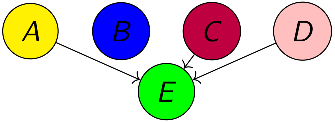


Design

Three conditions:

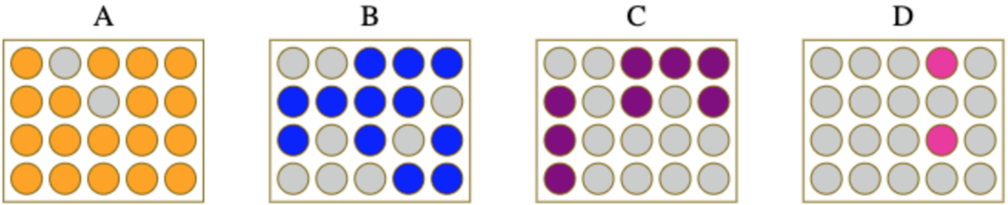


Observations only

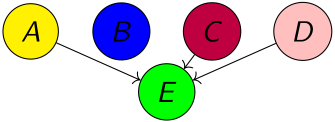


Design

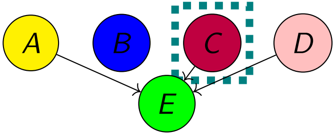
Three conditions:



Observations only

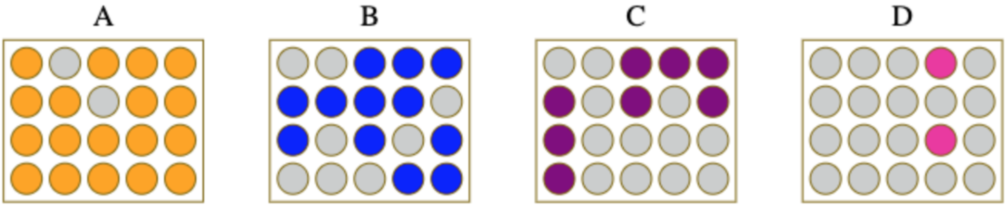


Causal selection

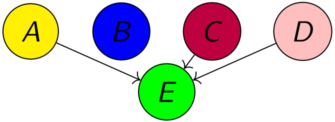


Design

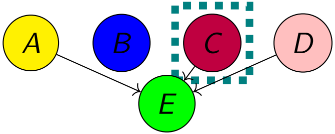
Three conditions:



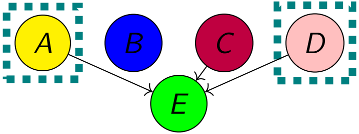
Observations only



Causal selection

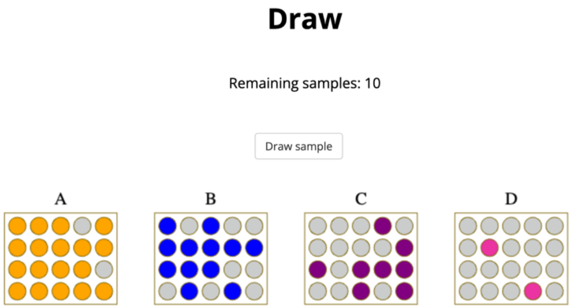


Other causal explanations



Experiment

Causal selection condition

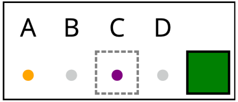


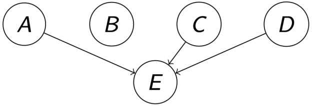
Experiment

Causal selection condition

$$\boxed{Win \leftarrow (A \wedge D) \vee C}$$

Draw

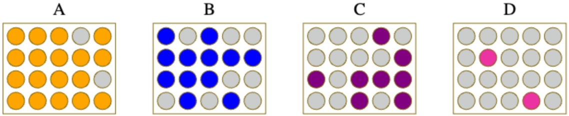




Here, you **won** **because** you drew a colored ball from urn **C!**

Remaining samples: 9

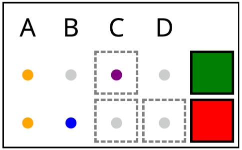
Draw sample



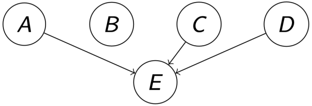
Experiment

Causal selection condition

Draw



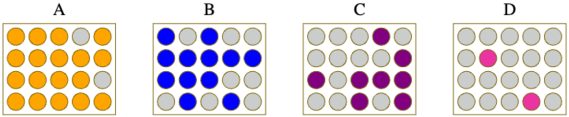
$$\boxed{Win \leftarrow (A \wedge D) \vee C}$$



Here, you **lost** **because** you drew gray balls from urns **C and D!**

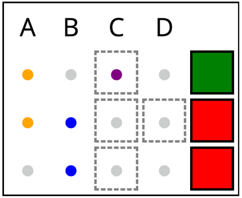
Remaining samples: 8

Draw sample

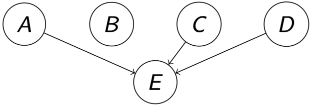


Experiment

Causal selection condition



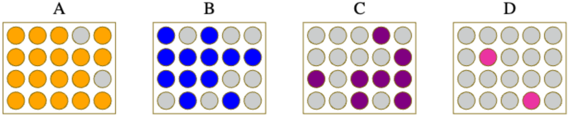
$$Win \leftarrow (A \wedge D) \vee C$$



Here, you **lost because** you drew a gray ball from urn **C!**

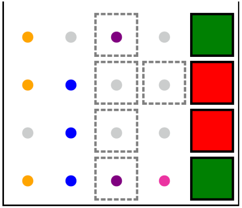
Remaining samples: 7

Draw sample

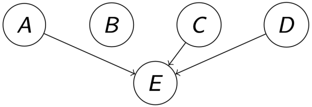


Experiment

Causal selection condition



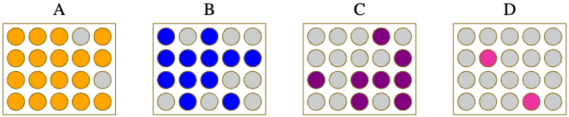
$$\boxed{Win \leftarrow (A \land D) \lor C}$$



Here, you **won** **because** you drew a colored ball from urn **C!**

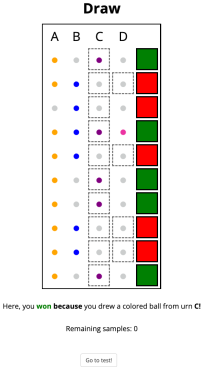
Remaining samples: 6

Draw sample

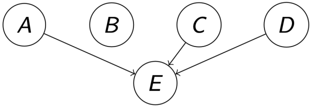


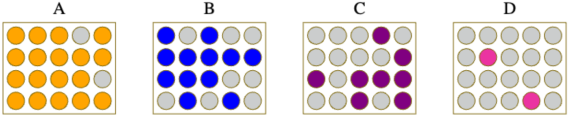
Experiment

Causal selection condition



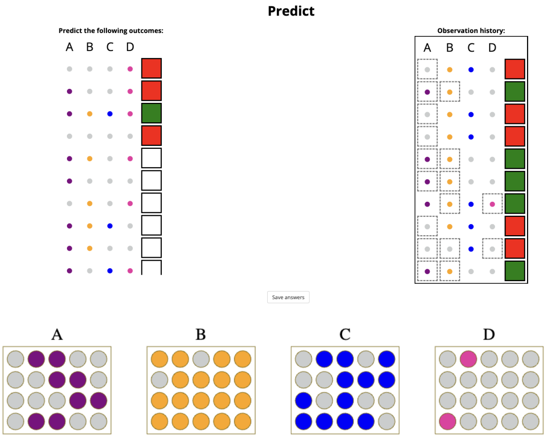
$$\boxed{Win \leftarrow (A \wedge D) \vee C}$$





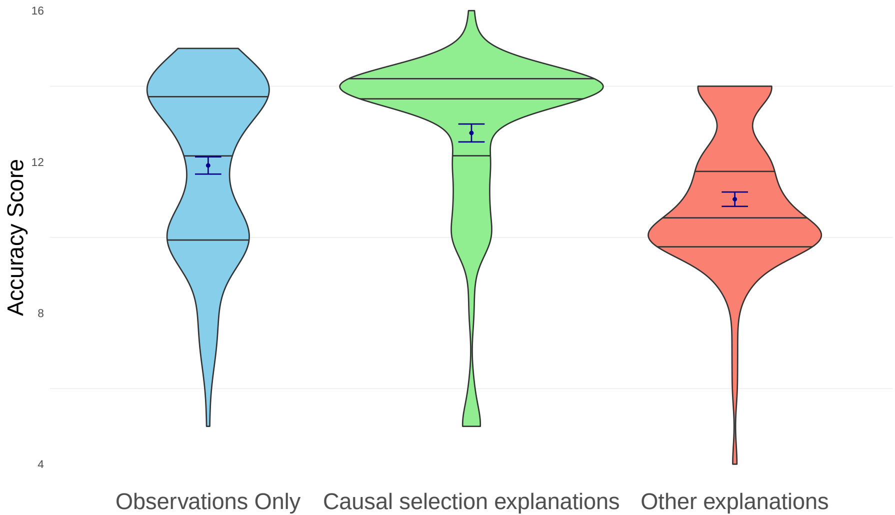
Experiment

Evaluation



Results

(N=298)



Experiment Model

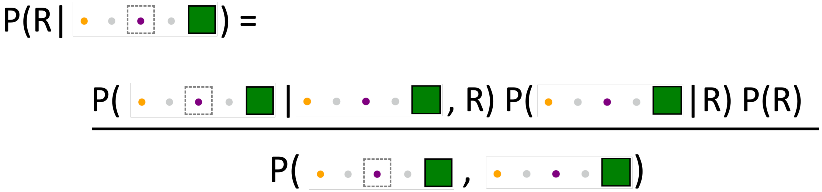
Inference from explanation over complex rules

$$P(R|O, E) = \frac{P(E|O, R)P(O|R)P(R)}{P(O, E)}$$

Experiment Model

Inference from explanation over complex rules

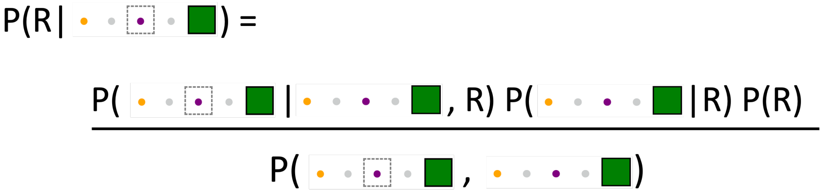
$$P(R|O,E) = \frac{P(E|O,R)P(O|R)P(R)}{P(O,E)}$$



Experiment Model

Inference from explanation over complex rules

$$P(R|O,E) = \frac{P(E|O,R)P(O|R)P(R)}{P(O,E)}$$

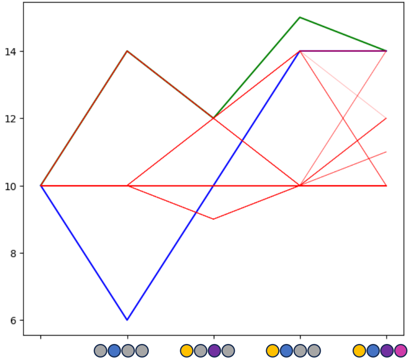


Explanation likelihood:

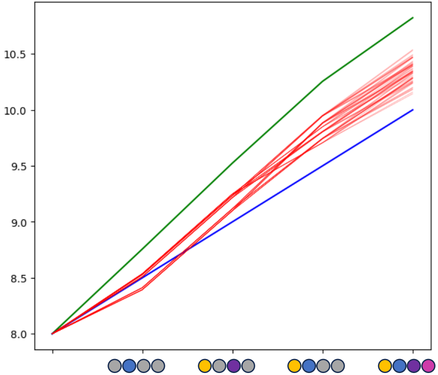
$$P(E \mid O, R) = \frac{\exp\left(\kappa(E,O,R)/\tau\right)}{\exp\left(\sum_{E_i \in \mathbf{E}} \kappa(E_i,O,R)/\tau\right)}$$

Model Results

- Observations only (blue)
- Causal selection explanations (green)
- Other explanations (red)



(a) MAP score



(b) **E**[*score*]

Discussion

Conclusions

- Causal selection judgements seem to be more informative forms of explanation to infer complex causal rules.

Discussion

Conclusions

- Causal selection judgements seem to be more informative forms of explanation to infer complex causal rules.

Next steps

- Extend inference task to other rules of varying complexity
- Collect explanation judgements from participants to provide as explanations to other participants.

Thank you for your attention.

Thanks to Salvador Mascarenhas and Neil Bramley for their guidandce on the project.
Thanks to Tom Icard and Tadeg Quillien for fruitful discussions and advice.

(a) Salvador Mascarenhas

(b) Neil Bramley

Bibliography

Icard, T. F., Kominsky, J. F., and Knobe, J. (2017). Normality and actual causal strength. *Cognition*, 161:80–93.

Kirfel, L., Icard, T., and Gerstenberg, T. (2022). Inference from explanation. *Journal of Experimental Psychology: General*, 151(7):1481–1501.

Quillien, T. and Lucas, C. G. (2023). Counterfactuals and the logic of causal selection. *Psychological review*.

Inference from Causal Explanation

Possible Worlds

Table

A	B	C	D	Times Seen in Observation	Probability ($*10^{-4}$)
0	0	0	0	0	216
0	0	0	1	0	24
0	0	1	0	0	144
0	0	1	1	0	16
0	1	0	0	1	324
0	1	0	1	0	36
0	1	1	0	0	216
0	1	1	1	0	24
1	0	0	0	0	1944
1	0	0	1	0	216
1	0	1	0	4	1296
1	0	1	1	0	144
1	1	0	0	4	2916

Causal Inference from Explanation

Causal Selections

A	B	C	D	Actual causes	Causal Selection
0	1	0	0	[A], [D], [A,D], [A,C], [A,C,D]	[C]
1	0	1	0	[A,C]	[C]
1	1	0	0	[C], [D]	[C,D]
1	1	1	1	[A], [D], [A,C], [A,D], [D,C] [A,C,D]	[C]

Table: List of actual causes for each sample, as well as the intuitive causal selection given the normality of variables A,B,C,and D.