# ACRE: Abstract Causal REasoning Beyond Covariation

CVPR 2021
NASHVILLE, TENNESSEE

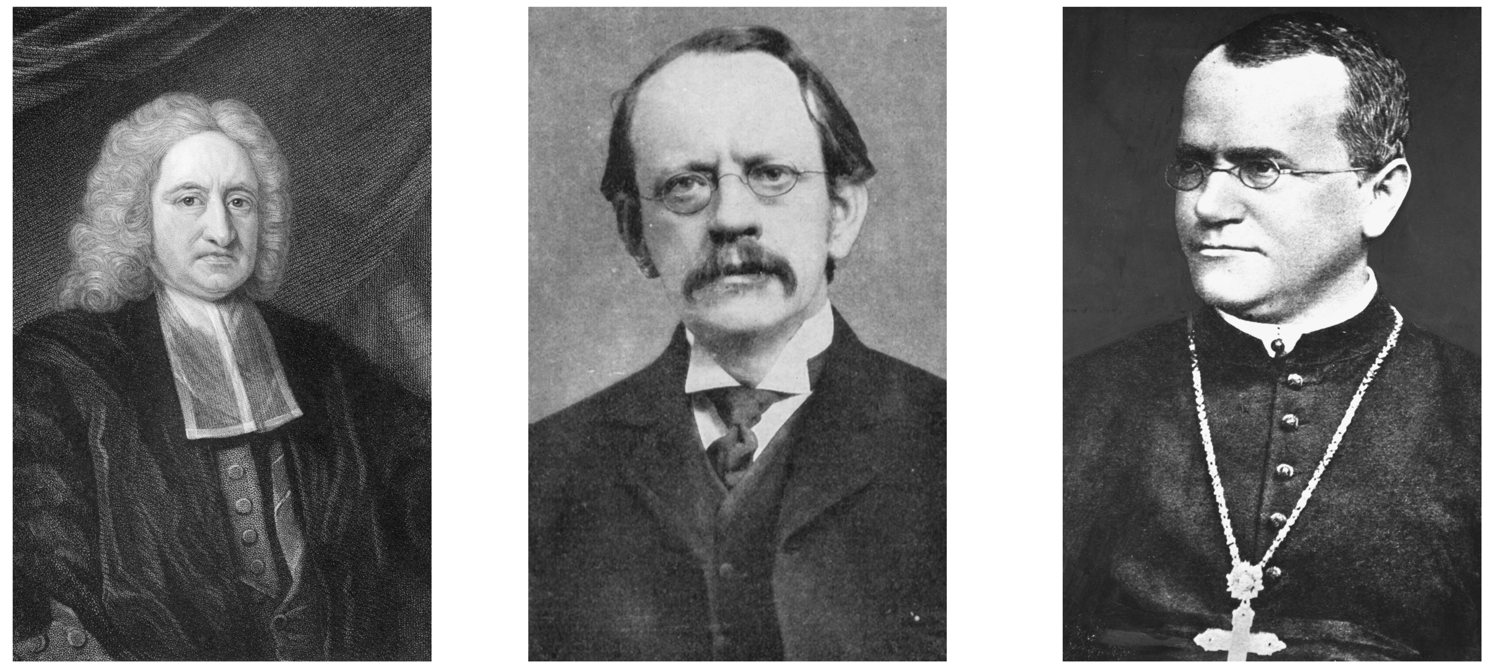
Chi Zhang Baoxiong Jia Mark Edmonds Song-Chun Zhu Yixin Zhu

UCLA Center for Vision, Cognition, Learning and Autonomy

{chi.zhang, baoxiongjia, markedmonds}@ucla.edu, sczhu@stat.ucla.edu, yixin.zhu@ucla.edu

## Motivation

"There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact." – Mark Twain

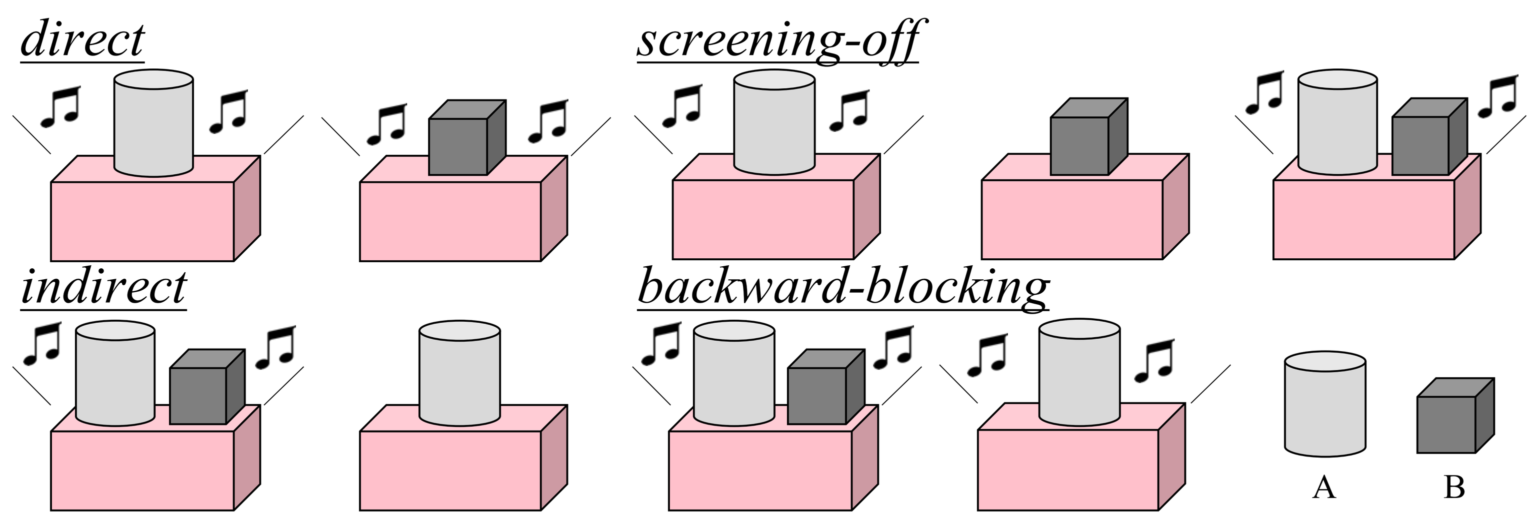


Edmond Halley
Halley's comet

Joseph Thomson
Electron

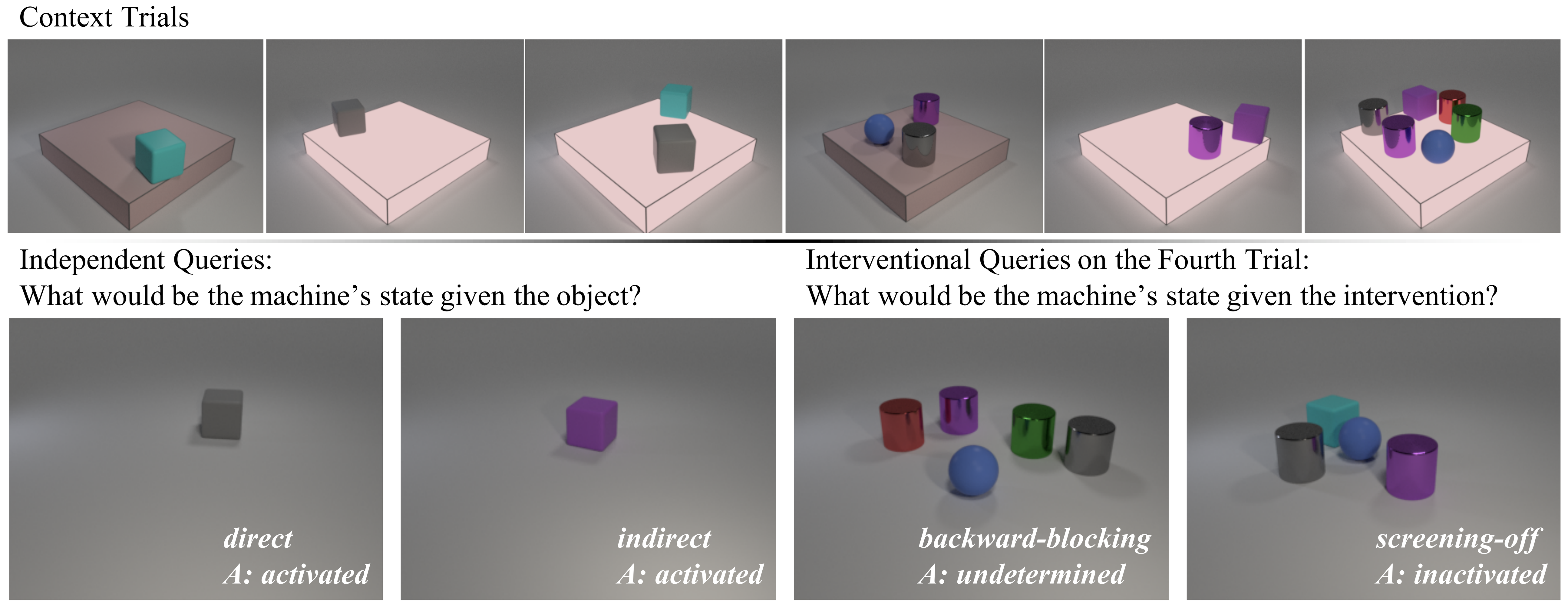
Gregor Mendel
Mendelian inheritance

*At what level do current visual reasoning systems induce causal relationships?*



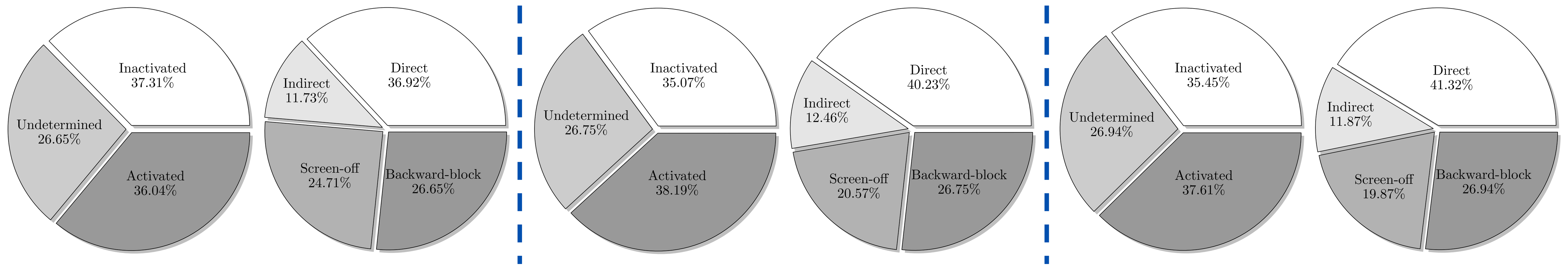

We ground the problem in the context of Blicket detection: Blicketness of an object can be solved via covariation in the direct and indirect case, but needs reasoning *beyond covariation* in the screening-off and backward-blocking case.

## ACRE




The first query tests causal reasoning from *direct* evidence, as the gray cube is independently tested and always associated with an activated machine. The second query requires comparing the fourth and fifth trial to realize that the Blicket machine is activated by the cube, not the cylinder, based on *indirect* evidence. As such, we infer that the red and green cylinders in the sixth trial may not activate the machine because the purple cube can already do so; despite their association with an activated machine only, their Blicketness is *backward-blocked* in the interventional trial. The cyan cube is *screened-off* by the gray cube's Blicketness from probabilistically activating it.

### Dataset Statistics in Generalization Splits

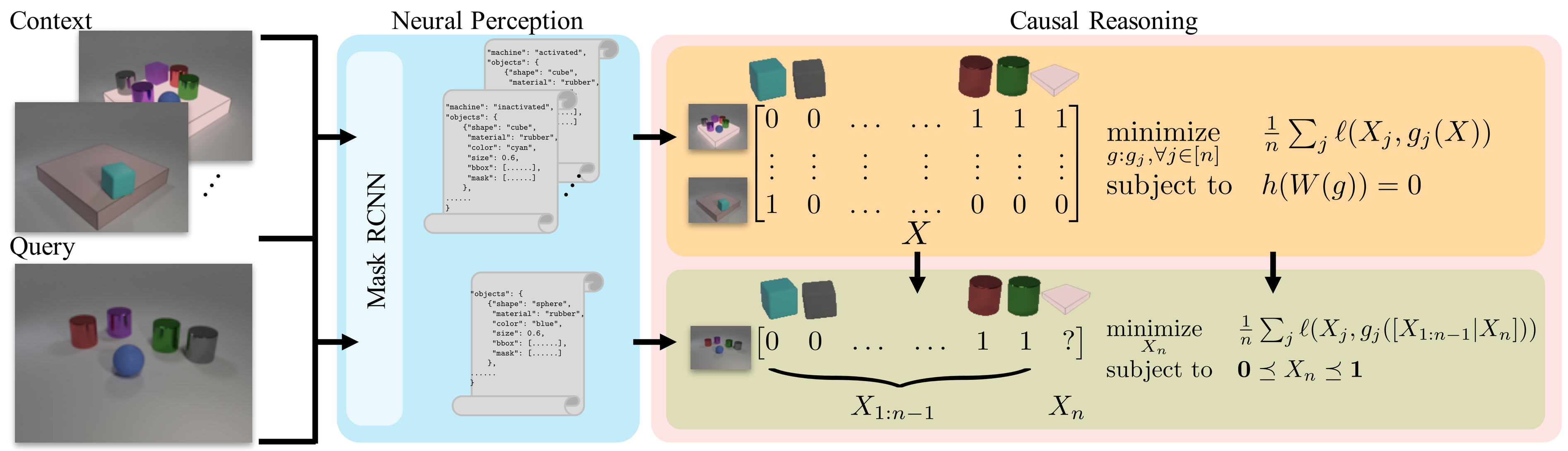



I.I.D.
Same distributions

Compositionality
Different compositions of attributes

Systematicity
Different patterns of activations

## A Neuro-Symbolic Method




## Benchmarking

### Performance of models on generalization splits

| Method | | MXGNet | LEN | CNN-MLP | WReN | CNN-LSTM | ResNet-MLP | CNN-BERT | NS-RW | NS-PC | NS-Opt |
|---|---|---|---|---|---|---|---|---|---|---|---|
| I.I.D. | Qry. | 33.01% | 38.08% | 40.86% | 40.39% | 41.91% | 42.00% | 43.56% | 46.61% | 59.26% | **66.29%** |
| | Pro. | 1.00% | 2.05% | 3.25% | 2.30% | 3.60% | 3.35% | 3.50% | 6.45% | 21.15% | **27.00%** |
| Comp. | Qry. | 35.56% | 38.45% | 41.97% | 41.90% | 42.80% | 42.80% | 43.79% | 50.69% | 61.83% | **69.04%** |
| | Pro. | 1.55% | 2.10% | 2.90% | 2.65% | 2.80% | 2.60% | 2.40% | 8.10% | 22.00% | **31.20%** |
| Sys. | Qry. | 33.43% | 36.11% | 37.45% | 39.60% | 37.19% | 37.71% | 39.93% | 42.18% | 62.63% | **67.44%** |
| | Pro. | 0.60% | 1.90% | 2.55% | 1.90% | 1.85% | 1.75% | 1.90% | 4.00% | 29.20% | **29.55%** |

### A detailed look on each query type

| Method | | MXGNet | LEN | CNN-MLP | WReN | CNN-LSTM | ResNet-MLP | CNN-BERT | NS-RW | NS-PC | NS-Opt |
|---|---|---|---|---|---|---|---|---|---|---|---|
| I.I.D. | D.R. | 27.73% | 49.07% | 55.56% | 51.04% | 48.20% | 54.87% | 52.24% | 88.88% | 84.46% | **91.64%** |
| | I.D. | 29.63% | 45.11% | 56.31% | 41.04% | 36.76% | 48.37% | 44.50% | **99.29%** | 29.33% | 69.25% |
| | S.O. | 14.88% | 33.68% | 44.88% | 29.75% | 53.23% | 42.29% | 42.59% | 7.21% | 78.31% | **85.37%** |
| | B.B. | **59.09%** | 23.91% | 9.71% | 35.61% | 24.91% | 21.12% | 32.15% | 1.66% | 20.50% | 11.98% |
| Comp. | D.R. | 36.93% | 47.58% | 57.59% | 55.29% | 56.58% | 62.79% | 54.07% | 91.74% | 89.50% | **92.50%** |
| | I.D. | 55.99% | 52.51% | 64.38% | 66.94% | 65.10% | 70.01% | 46.88% | **99.80%** | 28.66% | 76.05% |
| | S.O. | 0.00% | 18.01% | 31.66% | 8.44% | 19.69% | 30.52% | 40.57% | 4.07% | 85.28% | **88.33%** |
| | B.B. | **52.35%** | 33.63% | 15.26% | 35.99% | 29.27% | 8.54% | 28.79% | 0.67% | 15.21% | 13.48% |
| Sys. | D.R. | 15.24% | 46.22% | 70.79% | 53.56% | 42.57% | 65.19% | 55.97% | 92.44% | 89.76% | **94.73%** |
| | I.D. | 5.42% | 47.90% | 87.61% | 71.35% | 37.61% | 85.07% | 68.25% | **99.89%** | 57.08% | 88.38% |
| | S.O. | 42.58% | 30.91% | 11.57% | 16.80% | 63.28% | 9.57% | 0.00% | 0.20% | 73.93% | **82.76%** |
| | B.B. | **56.38%** | 24.89% | 3.60% | 31.62% | 8.70% | 13.38% | 45.59% | 0.46% | 24.88% | 16.06% |

## Future Work

- How to learn the hidden causal relations beyond just capturing the covariation?
- What is the symbolic mechanism that leads to the reasoning in backward-blocking?
- How to combine learning and reasoning in a unified framework to address the problem?