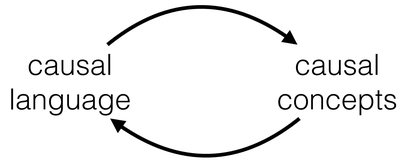
The Language of Causation

Ari Beller*, Erin Bennett*, Tobias Gerstenberg
Stanford University

Introduction

The words we use matter.

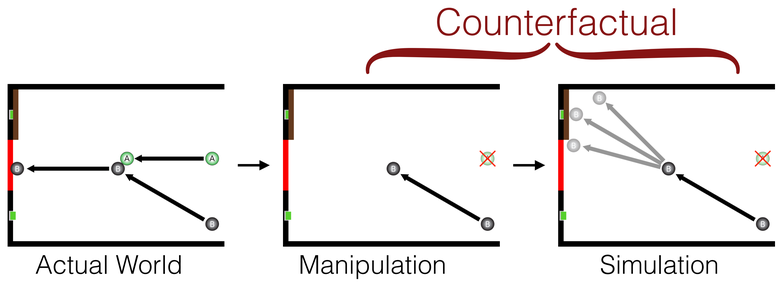
Dr. Jones killed the patient.
vs
Dr. Jones caused the patient's death.
vs
Dr. Jones enabled the patient's death.



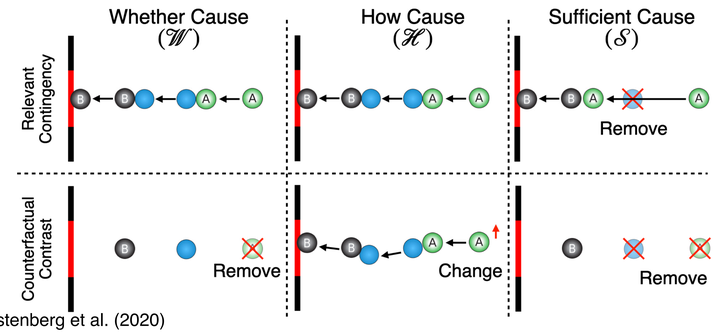
Rodriguez-Arias et al. (2020)

Model

1) Causal Inference



Different counterfactual manipulations under a physics generative model allow us to assess different ***aspects*** of causation.

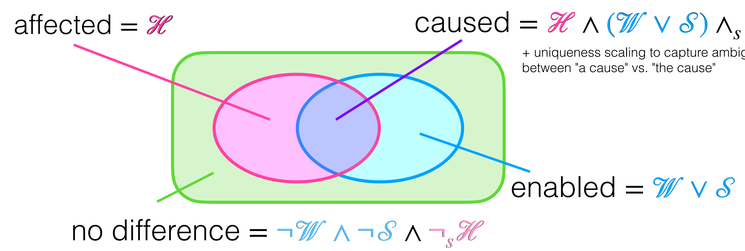


Gerstenberg et al. (2020)

Model (cont.)

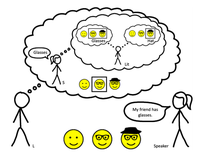
2) Semantics

Causal expressions overlap in the space of causal concepts they refer to.



3) Pragmatics

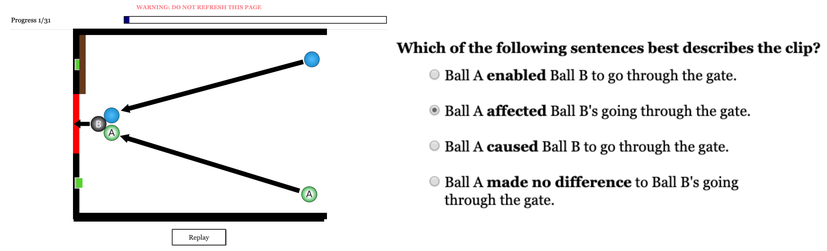
Rational Speech Act (RSA) models communication with recursive inference to capture rational principles of social communication.



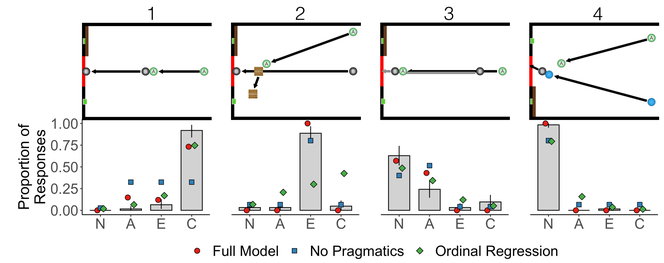
Frank & Goodman (2012)

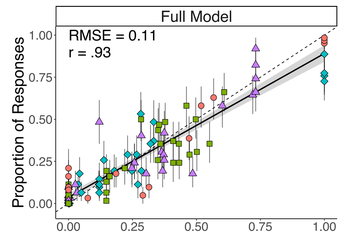
Experiment

64 participants recruited on Mturk; 30 video clips; Within-subject design



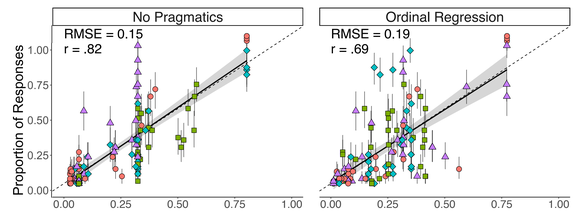
Results





- No Pragmatics fails in cases like clip 1 where multiple expressions are true.
- No Semantics and No Pragmatics (Ordinal Regression) fails in cases like clip 2 where it overrates "cause" and underrates "enabled".
- Full model improves on both these alternatives, suggesting both semantics and pragmatics are important.

Alternative Models



No Pragmatics
Uses semantic values and softmax to predict proportion for each expression.

No Semantics and No Pragmatics
Bayesian Ordinal Regression assumes ordering:
No difference < Affected < Enabled < Caused

Conclusions

- We investigated the words people use to describe causal scenarios.
- We find that causal knowledge, semantics, and pragmatic reasoning are all important.
- In future work, we plan to go beyond forced choice settings and account for more naturalistic descriptions.

References

- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336 (6084), 998–998.
- Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2020). A counterfactual simulation model of causal judgment. PsyArXiv. (https://psyarxiv.com/7zj94/)
- Rodríguez-Arias, D, Rodríguez López, B, Monasterio-Astobiza, A, Hannikainen, IR. How do people use 'killing', 'letting die' and related bioethical concepts? Contrasting descriptive and normative hypotheses. *Bioethics*. 2020; 34: 509– 518. https://doi.org/10.1111/bioe.12707