Explanations in context

Erin Bennett (erindb@stanford.edu), **Noah D. Goodman** (ngoodman@stanford.edu)
Department of Psychology, Stanford University.

Abstract

Proposal: 250-500 words

Different factors seem to affect the selection of the best explanation from a set of possible causes. Some of these factors seem to be the result of pragmatic inference. We implement an existing probabilistic model of counterfactual reasoning to identify possible causes (and gradable judgements of how causally relevant each might be) (Lucas & Kemp, 2015), along with a model of rational communication (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013) to investigate how changes to the background causal model, changes to what information is common knowledge, and changes to the question under discussion (QUD) can affect people's judgments of how good a potential cause is at explaining an effect. In particular, we test such a model in situations where interlocuters have different knowledge states to see whether it tracks people's judgments of the goodness of explanations, and also look at whether the effect of moral judments and/or exportability on causal selection can be explained by QUD effects.

Keywords: explanations; counterfactuals; pragmatics

Background

Causal Selection

Hesslow (1988) discusses the issue of *causal selection* as a seperate task from *causal attribution*. The latter determines whether a condition can truthfully be considered a contributing cause. The former determines, of possible contributing causes, which is the best *explanation*.

Hesslow claims that an explanation must always explain why an effect occurred in the actual situation but did not occur (or would not have occurred) in some contextually-specified comparison class of situations. He uses this account to unify many previously proposed criteria for causal selection. The relevant comparison class must depend on the interlocuters' knowledge of what alternative situations are possible, and on the topic of their conversation. Under some topics, the utility, moral acceptability, temporal proximity, typicality or stability of the different alternatives might matter to the generation of the comparison class.

Hilton (1996) desribes an informal model of Gricean pragmatics in explanations. He explains between casual selction processes of causal "discounting" and causal "backgrounding" in terms of different Gricean maxims. Causal discounting, where one cause becomes a less good explanation as a result of an alternative cause gaining salience, seems to be an effect of changes in the underlying generative model (*quality*, or truthfulness) or in the question under discussion (*relevance*). Causal backgrounding, where a cause becomes a less good explanation as a result of becoming especially predictable, seems to happen as a result of changes what common knowledge is assumed between the interlocuters (*informativity*).

Reuter, Kirfel, van Riel, and Barlassina (2014) describe factors that seem to guide causal selection that they consider to be independent of pragmatic effects or the underlying causal structure: temporal proximity and morality. They show through a series of experiments that the most recent potential causes A and B is regarded as the best explanation of an effect E when the causal structure is "A and B implies E", and that this effect is overridden by a tendency to hold as responsible any cause that violates a norm (e.g. if A broke a rule, and B did not, then A will be held responsible for the effect jointly caused by both A and B).

Woodward (2011), responding to some arguments that people's notion of "cause" is purely subjective and contextual, argues that some factors that may seem pragmatic and context-dependent are actually systematic and invariably matter to determining whether something is considered a "cause" or just an "enabling condition". The factors that Woodward cites as probably systematic and independent of context are "stability" and "specificity". By "stability", he means the sufficiency of the cause to bring about the effect in other counterfactual scenarios.[1] By "specificity", he means the extent to which minor changes in the manner in which the cause occurred would have brought about corresponding changes in the way the effect occurred. better reference for this, e.g. with experiments...?

Counterfactuals

Lucas and Kemp (2015) ...

Jimnez-Leal (2013) show through a series of experiments that reasoning about cause and reasonging about counterfactuals might be the same process. Previous results had shown that people respond differently to questions about cause than to questions about counterfactuals, but ... show that this is probably due to task demands in the language used to elicit causal versus counterfactual judgments.

Rational Speech Act Models

RSA models (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013) and related models (Franke, 2011; Russell, 2012).

References

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

[1] Other authors have used "sensitivity" to refer to how much variations in other features of a counterfactual world would negate the potential cause's ability to bring about the effect.

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics*, *4*, 1–82.

Goodman, N., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*.

Hesslow, G. (1988). *The problem of causal selection.* Harvester Press, Brighton.

Hilton, D. J. (1996). Mental models and causal explanation: Judgements of probable cause and explanatory relevance. *Thinking & Reasoning*, *2*(4), 273–308.

Jimnez-Leal, W. (2013). Causal selection and counterfactual reasoning. *Revista Colombiana de Psicologia*, *22*, 179–197.

Lucas, C. G., & Kemp, C. (2015). An improved probabilistic account of counterfactual reasoning. *Psychological Review*.

Reuter, K., Kirfel, L., van Riel, R., & Barlassina, L. (2014). The good, the bad, and the timely: how temporal order and moral judgment influence causal selection. *Fronteirs in Psychology*.

Russell, B. (2012). *Probabilistic reasoning and the computation of scalar implicatures*. Unpublished doctoral dissertation, Brown University.

Woodward, J. (2011). *Causes, conditions, and the pragmatics of causal explanation.*