Good explanations fit prior knowledge

Joshua Hartshorne[1], Tobias Gerstenberg[1], Noah Goodman[2]
[1] Communication Sciences & Disorders, MGH Institute of Health Professions; [2] Department of Psychology, Stanford University
jkhartshorne@fas.harvard.edu

E because C is ambiguous, indicating either a causal explanation (*C caused E*) or an epistemic explanation (*C caused the speaker to believe that E*). Nonetheless, people seems to have more difficulty noticing the ambiguity than resolving it. For instance, the causal meaning m_{causal} is strong preferred in (1), and the epistemic meaning $m_{epistemic}$ in (2):

1. The grass is wet because it rained last night.
2. It rained last night because the grass is wet.

Plausibility provides an intuitive explanation for (2) — wet grass does not cause overnight rain — but less so for (1) — rain is strong evidence for wet grass. We test a Bayesian account of comprehension [1, 2], where the probability that utterance u conveys message m is given by:

3. P(m|u) $\propto$ P(u|m) * P(m)

Assuming a cooperative speaker — and ignoring, for the moment, Gricean reasoning — the prior $P(m)$ largely reduces to the prior probability that the message is true. As already noted, $P(m_{epistemic}) > P(m_{causal})$ is sufficient to explain (2). (1) could be explained if $P(u|m_{causal}) > P(u|m_{epistemic})$ for both sentences.

To test this account quantitatively, we created 32 event pairs (e.g., *the grass is wet* and *it rained*). For each pair, 71 participants rated both prior probabilities ($P(m_{causal})$, $P(m_{epistemic})$). From these pairs, we created 32 sentences. An additional 72 participants judged the relative probability of the causal and epistemic interpretation for each. All probability ratings were on a scale of 0-100. We assumed that the likelihoods $P(u|m_{causal})$ and $P(u|m_{epistemic})$ were the same for all sentences (i.e., they did not depend on the specific events) and fit them to the data. The resulting model fit the data extremely well (Fig. 1, left; $r = .84$, 95% CI $[.69, .92]$, $t(30) = 8.34$, $p < .001$). The best-fitting relative probabilities of $P(u|m_{causal})$ and $P(u|m_{epistemic})$ were 85% and 15%, respectively, matching the expectation that the former would be substantially higher than the latter.

We also created a second set of sentences by adding *I think* to the beginning of each:

4. I think the grass is wet because it rained last night.
5. I think it rained last night because the grass is wet.

Note that while the causal reading is still possible — (4) might indicate that the speaker believes it is the case that the rain last night caused the grass to be wet — the epistemic readings now seem more likely in both cases. This is likely because for these sentences, $P(u|m_{causal}) < P(u|m_{epistemic})$. Nonetheless the priors $P(m)$ are still the same, which would explain why the causal reading still seems more likely for (4) than (5). We obtained judgments for these new sentences from the same 72 participants. As expected, epistemic interpretations were much more likely (Fig. 1, right). We again fit likelihoods to the data. Overall, the model fits were again quite good ($r = .78$, 95% CI $[.59, .89]$, $t(30) = 6.73$, $p < .001$). The best-fitting relative probabilities of $P(u|m_{causal})$ and $P(u|m_{epistemic})$ were 13% and 87%, respectively, again matching expectations.

We discuss how adding Gricean reasoning to this model [e.g., using RSA; 3] would affect results. We also consider broader implications for theories of language comprehension.

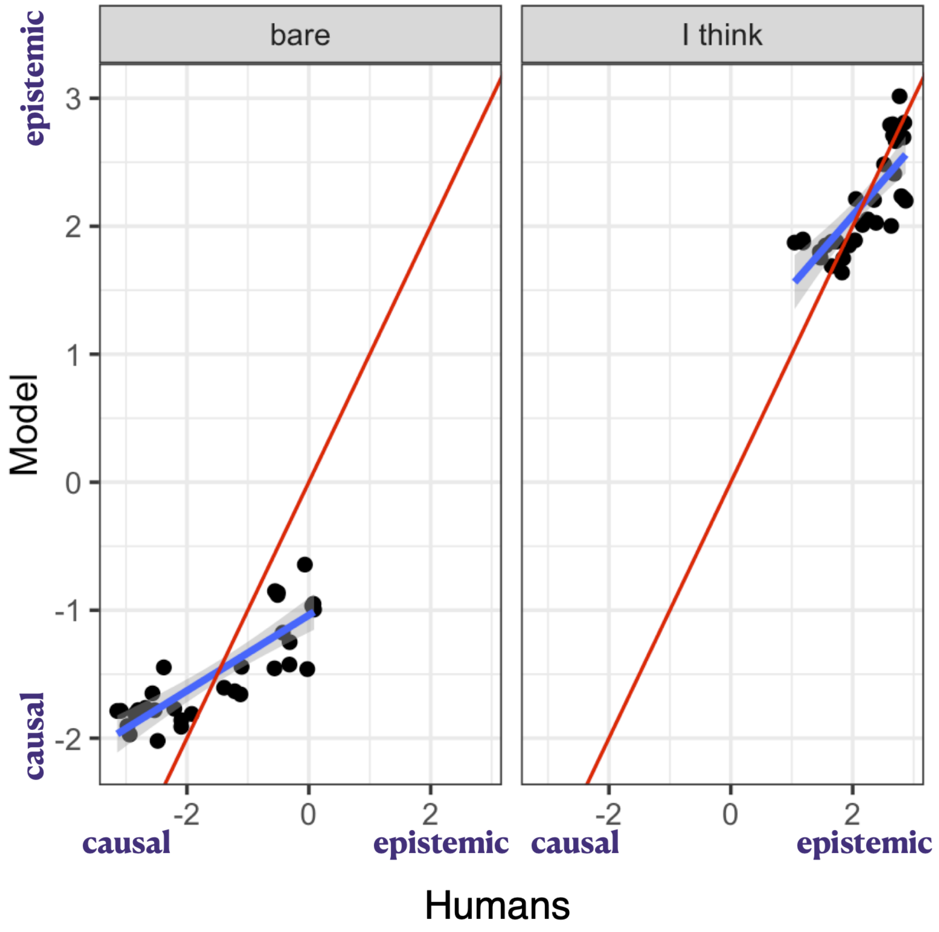

Figure 1: log-odds of epistemic vs. causal interpretation, with positive numbers indicating greater probability of the causal interpretation and negative numbers indicating greater probability of the epistemic interpretation. Each point represents a sentence, with bare sentences like (1) and (2) on the left and 'I think' sentences like (4) and (5) on the right. Separate linear fits are shown, with the 95% confidence interval indicated.

References

[1] Narayanan, S., & Jurafsky, D. (2022). Bayesian models of human sentence processing. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, 752–757.

[2] Hartshorne, J. K., Jennings, M. V., Gerstenberg, T., & Tenenbaum, J. (2019). When circumstances change, update your pronouns. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *41*.

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