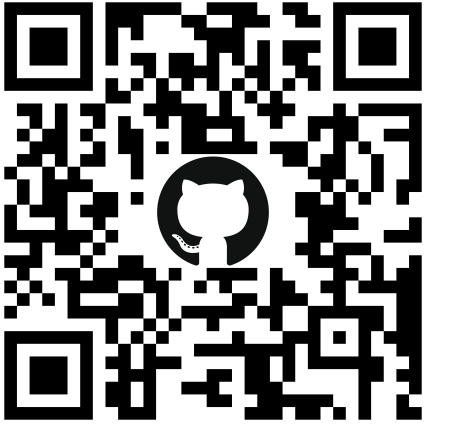


COPA-SSE: Semi-Structured Explanations for Commonsense Reasoning

Ana Brassard^{1,2} Benjamin Heinzerling^{1,2} Pride Kavumba^{2,1} Kentaro Inui^{2,1}

¹Riken AIP ²Tohoku NLP Lab



<https://github.com/a-brassard/copa-sse>

*Making explanation data is **hard**.*

Free-form crowdsourcing? → *Quality-control issues*
Extracting from KGs? → *Coverage/relevance issues*

Our solution: **semi-structured** crowdsourcing!

I cut my foot on the floor. What was the cause of this?

- a) The tile was cracked. (correct)
b) The tile was wet.

Cutting is capable of cutting foot.
Cracking is capable of cutting.
Tile is capable of cracking.

★★★★★ 3.9

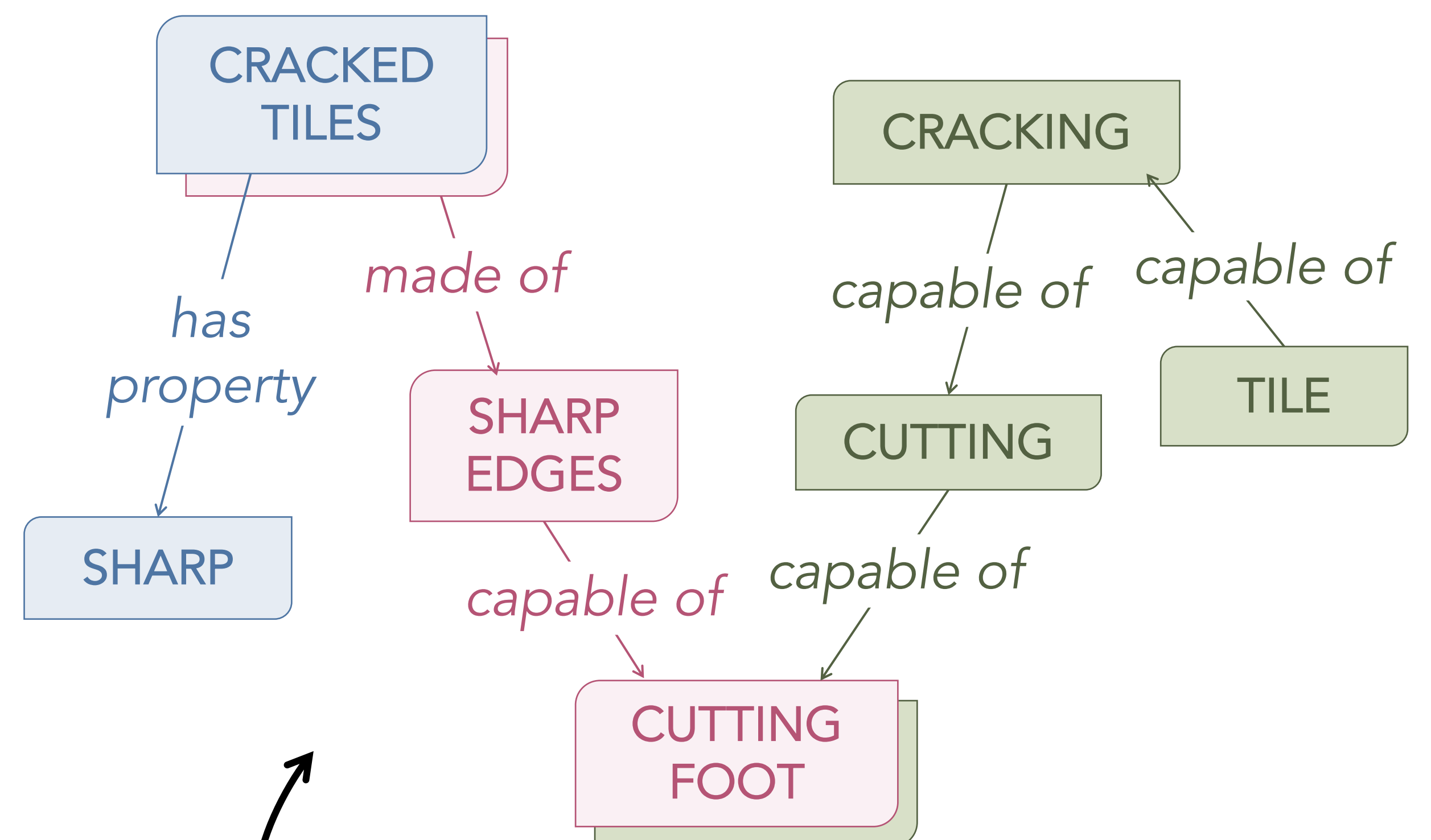
Cracked tile is sharp.

★★★★★ 3.4

Cracked tiles is made of sharp edges.
Sharp edges is capable of cutting foot.

★★★★★ 4.2

"Is this a good explanation?"



Finally, the explanations can be aggregated into **graphs**, serving as examples of **relevant generated and/or extracted subgraphs**.

We then asked (different) crowdworkers to **rate** the explanations on a scale of 1 to 5 stars.

Crowdworkers were shown COPA (Choice of Plausible Alternatives)³ questions & asked to support the answer with **triples**:

head concept <ConceptNet⁴ relation> tail concept

Question:
a) **Answer** ✓
b) **Alternative**

Why is this correct? What knowledge is the correct answer based on?
Please use as many fields as necessary to create a chain of two or more common sense statements that help explain the answer.

Write a concept... Select a relation Write a concept...
+add

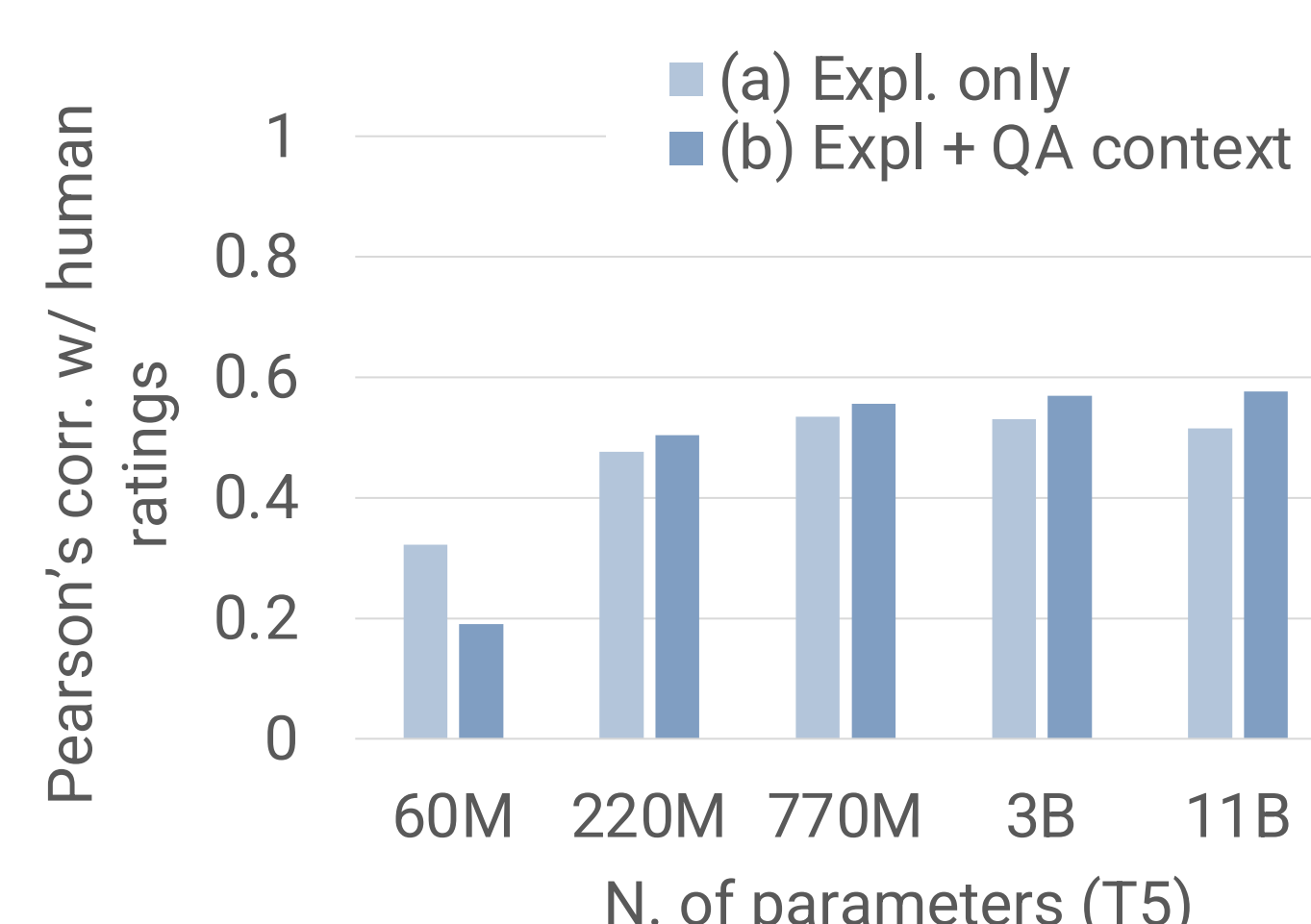
Question
a) **Answer** ✓
b) **Alternative**

The following cards contain candidate explanations for the above c
A good explanation **should provide implicit knowledge that con**
Please see the instructions for example ratings.

chain_0 ★ ★ ★ ★ ★
chain_1 ★ ★ ★ ★ ★

Usage example | **Automatic explanation scoring**

- COPA-SSE has explanations w/ various quality ratings; can be used to **create automatic scoring** systems.
- Bigger T5 models⁶ had higher correlation with human ratings, but still much room for improvement!



(a) Expl. only input:

Rate this explanation: Cracked tile is sharp.

(b) Expl. + QA context input:

Rate this explanation: I cut my foot on the floor. because The tile was cracked. Explanation: Cracked tile is sharp.

Expected output: **3.4**

Quick stats

- 9,747** explanations for 1,500 Balanced COPA⁵ questions
- 61% one-statement, 39% multi-statement
- 44% with 3.5 stars or more
- 98% of questions have at least one 3.5+ star explanation

³Roemmele, M., Bejan, C., and Gordon, A. (2011) Choice of Plausible Alternatives: An Evaluation of Commonsense Causal Reasoning. AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning, Stanford University, March 21-23, 2011.

⁴Speer, R., Chin, J., and Havasi, C. (2017) ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In proceedings of AAAI 31.

⁵Kavumba, P., Inoue, N., Heinzerling, B., Singh, K., Reisert, P., & Inui, K. (2019). When Choosing Plausible Alternatives, Clever Hans can be Clever. EMNLP 2019, 33.

⁶Roberts, A., Raffel, C., Lee, K., Matena, M., Shazeer, N., Liu, P. J., Narang, S., Li, W., and Zhou, Y. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. Technical report, Google.