

Chain-of-Thought

- ▶ Most explanations we've seen are about interpreting models
- ▶ Chain-of-thought: prompting technique for **using explanations to improve model performance**, particularly for complex reasoning tasks
- ▶ Basic idea: the language model can “work through” different types of computation over multiple timesteps of inference, rather than needing to generate an answer immediately

Rationales as “Programs”

Problem 2:

Question: From a pack of 52 cards, two cards are drawn together at random. What is the probability of both the cards being kings?

Options: A) $2/1223$ B) $1/122$ C) $1/221$ D) $3/1253$ E) $2/153$

Rationale: Let s be the sample space.

Then $n(s) = 52C2 = 1326$

E = event of getting 2 kings out of 4

$n(E) = 4C2 = 6$

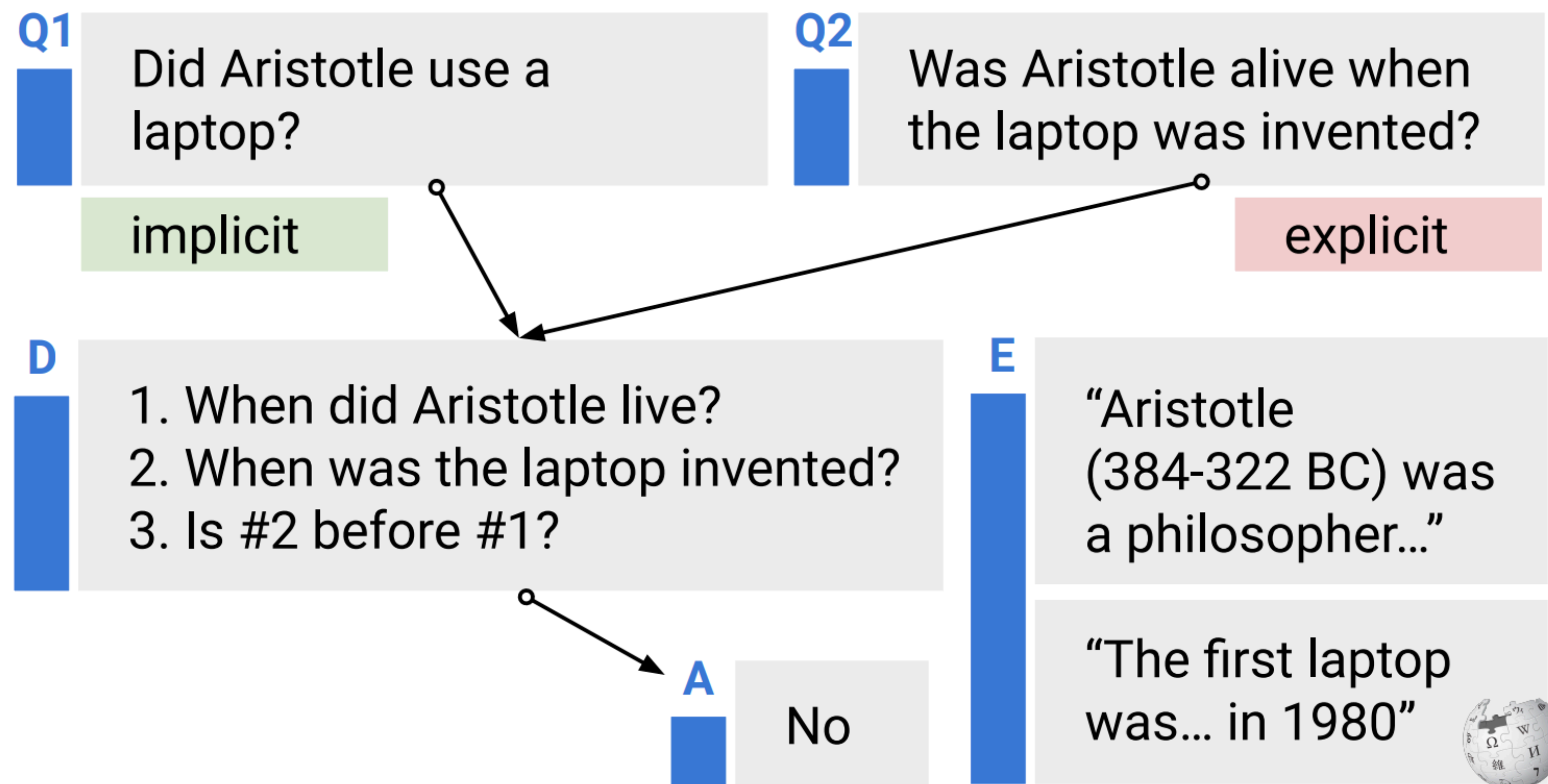
$P(E) = 6/1326 = 1/221$

Answer is C

Correct Option: C

- ▶ Rationales are most useful for problems where some computation is required. They can articulate the intermediate steps needed to solve it
- ▶ Some of the earliest work: math word problems

Rationales as “Programs”



- ▶ “StrategyQA”: dataset where different reasoning strategies are needed
- ▶ Related to multi-hop QA: “*What’s the capital of the country where Aristotle lived?*” (but these are easy with current models)

Chain-of-Thought

- ▶ For these kinds of problems, do “computation” entirely in natural language
- ▶ Unifies several ideas:
 - ▶ For math: relies on the fact that LLMs can at least do single steps of arithmetic okay
 - ▶ For QA: many problems involve reasoning decompositions
E.g., *What’s the capital of the country where Aristotle lived?* ->
ans = “*country where Aristotle lived*”
return *What’s the capital of [ans]*
 - ▶ For other tasks: capture the kinds of behavior written in rationales

Chain-of-Thought

- ▶ Chain-of-thought is usually a few-shot prompting technique where the in-context examples now contain explanations
- ▶ Answer is not generated in one go, but comes after an explanation that “talks through” the reasoning

Input:

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

...

Q: John takes care of 10 dogs. Each dog takes .5 hours a day to walk and take care of their business. How many hours a week does he spend taking care of dogs?

A:

Model output:

John takes care of 10 dogs. Each dog takes .5 hours a day to walk and take care of their business. So that is $10 \times .5 = 5$ hours a day. $5 \text{ hours a day} \times 7 \text{ days a week} = 35 \text{ hours a week}$. The answer is 35 hours a week. ✓

Chain-of-Thought

From our work: a synthetic test of multi-hop reasoning with extractive explanations:

Context: *Christopher agrees with Kevin. Tiffany agrees with Matthew. Mary hangs out with Danielle. James hangs out with Thomas. Kevin is a student. Matthew is a plumber. Danielle is a student. Thomas is a plumber.*

Q: *Who hangs out with a student?*

A: *Mary.*

Explanation: *because Mary hangs out with Danielle and Danielle is a student.*

Chain-of-Thought

Context: *Christopher agrees with Kevin. [...]* **Q:** *Who hangs out with a student?*

Mary

Standard few-shot learning, no explanation

Context: *Christopher agrees with Kevin. [...]* **Q:** *Who hangs out with a student?*

Mary, *because Mary hangs out with Danielle and Danielle is a student.*

Predict-explain: answer **is not** conditioned on output explanation (original E-SNLI LSTM)

Context: *Christopher agrees with Kevin. [...]* **Q:** *Who hangs out with a student?*

Because Mary hangs out with Danielle and Danielle is a student, the answer is Mary.

Explain-predict: answer is conditioned on output explanation (Chain of Thought)

Chain-of-Thought

Prompt

Input
Label+
Explanation

Context: Christopher agrees with Kevin. [...] **Q:** Who hangs out with a student?
Mary, because Mary hangs out with Danielle and Danielle is a student.

Train Ex

Train Ex

Test Input

Context: Adam plays with Ellen. [...] **Q:** Who plays with a doctor?

GPT-3

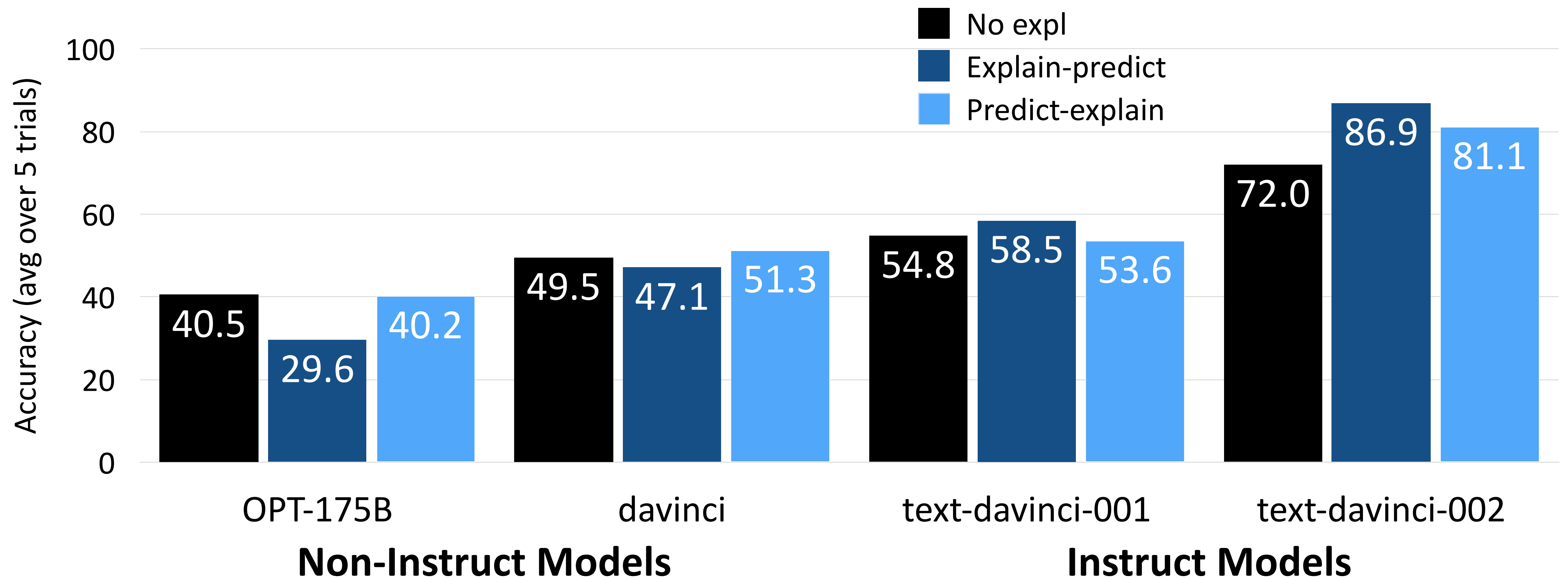
Output

Adam, **because Adam plays with Ellen and Ellen is a doctor.**

greedy decoding from GPT-3

Chain-of-Thought

Results on SYNTH data



- ▶ Instruct tuning / RLHF improves models' ability to use explanations
- ▶ Chain-of-thought helps on the biggest and best models, but isn't always effective on weaker models

Chain-of-Thought: Results

| | MultiArith | GSM8K |
|--|-------------|-------------|
| Zero-Shot | 17.7 | 10.4 |
| Few-Shot (2 samples) | 33.7 | 15.6 |
| Few-Shot (8 samples) | 33.8 | 15.6 |
| Zero-Shot-CoT | 78.7 | 40.7 |
| Few-Shot-CoT (2 samples) | 84.8 | 41.3 |
| Few-Shot-CoT (4 samples : First) (*1) | 89.2 | - |
| Few-Shot-CoT (4 samples : Second) (*1) | 90.5 | - |
| Few-Shot-CoT (8 samples) | 93.0 | 48.7 |
| Zero-Plus-Few-Shot-CoT (8 samples) (*2) | 92.8 | 51.5 |
| Finetuned GPT-3 175B [Wei et al., 2022] | - | 33 |
| Finetuned GPT-3 175B + verifier [Wei et al., 2022] | - | 55 |
| PaLM 540B: Zero-Shot | 25.5 | 12.5 |
| PaLM 540B: Zero-Shot-CoT | 66.1 | 43.0 |
| PaLM 540B: Zero-Shot-CoT + self consistency | 89.0 | 70.1 |
| PaLM 540B: Few-Shot [Wei et al., 2022] | - | 17.9 |
| PaLM 540B: Few-Shot-CoT [Wei et al., 2022] | - | 56.9 |
| PaLM 540B: Few-Shot-CoT + self consistency [Wang et al., 2022] | - | 74.4 |

- ▶ “Let’s think step by step” paper introduced a new zero-shot prompt. CoT works much better than non-CoT, and few-shot is better

Kojima et al. (2022)