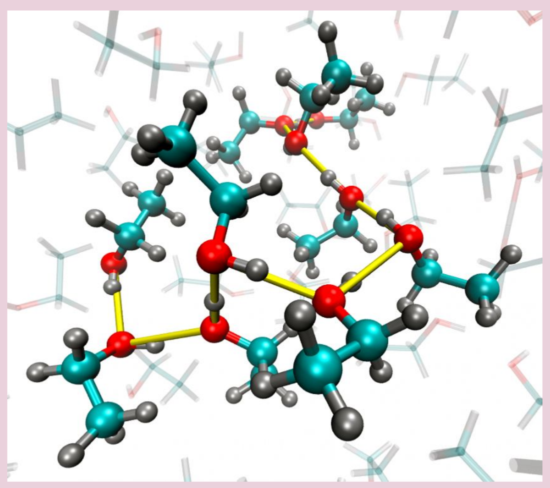
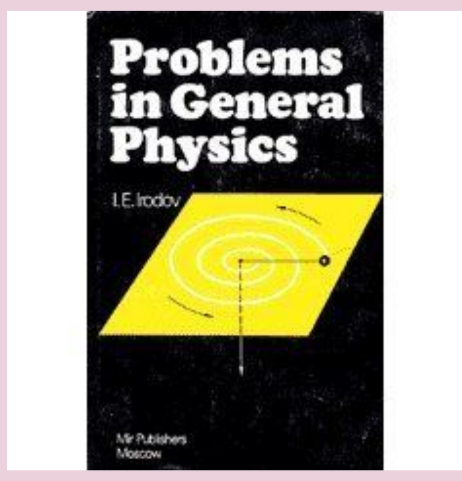
Multi-token prediction boosts creativity in algorithmic tasks

Vaishnavh Nagarajan* (Google Research NY), Chen Henry Wu* (CMU),
Charles Ding (CMU), Aditi Raghunathan (CMU)

We'd want LLMs to *creatively* solve *open-ended* tasks

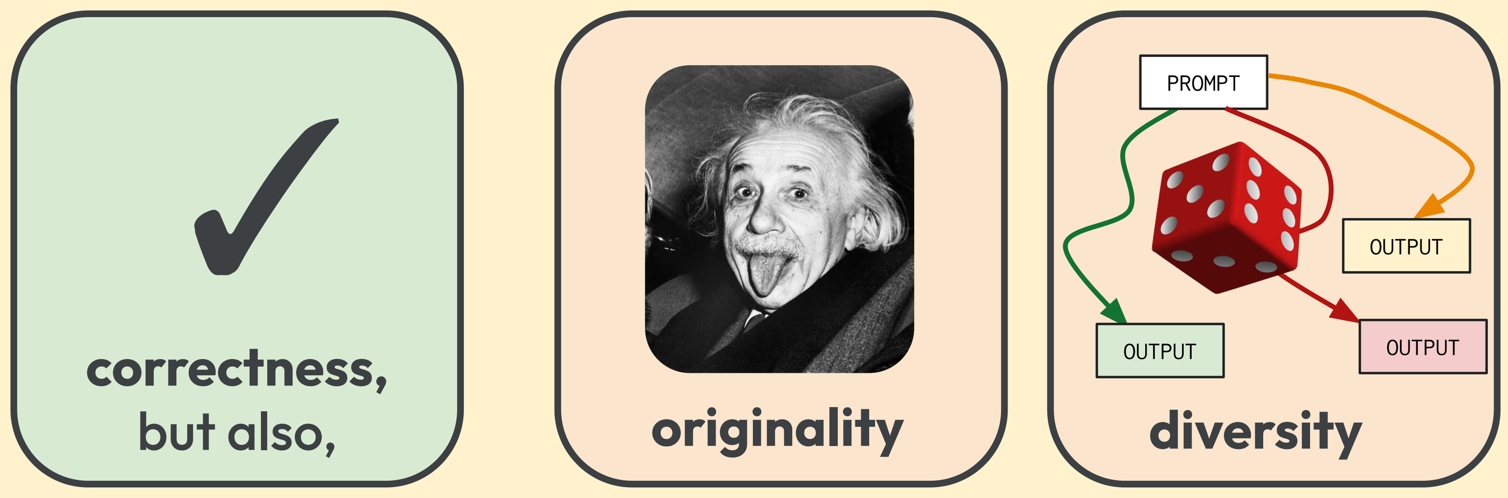




`Discovery:` "Are there any surprising connections between some of these molecules?"

`Synthetic data:` "Generate a dataset of original olympiad problems"

But this demands evaluating *subjective*, *unscalable* metrics including not only



Our questions: How do we precisely evaluate LLM creativity? Is the current LLM paradigm optimal for creativity? Can we improve it?

Our approach: Design *minimal open-ended algorithmic tasks* modeled after creative tasks.

General setup: In all our tasks, the model must learn an underlying distribution D through a training set S of m independent samples. *Algorithmic creativity* is defined as fraction of generations that are *unique*, *original (not in S)*, and *coherent (in D)*.

Through this, we isolate two types of creativity in CogSci [1]:

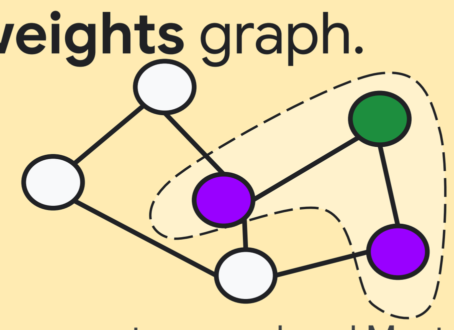
1. Combinational creativity

Discovering novel connections from **known things** (e.g., wordplay, humor, analogies)

`What shoes do spies wear? Sneakers!`

`What genre do balloons enjoy? Pop music!`

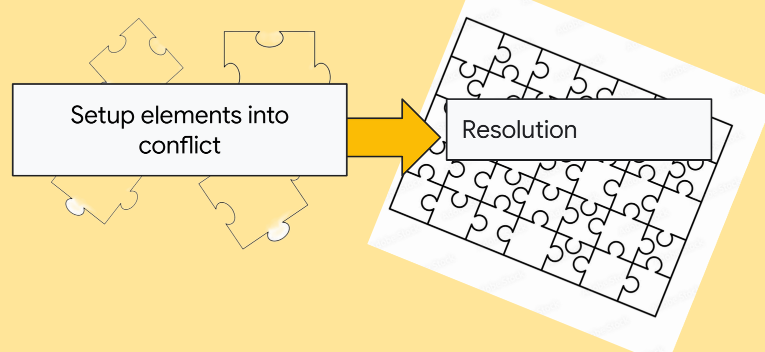
Our algorithmic tasks: Generate novel multi-hop connections (siblings / triangles / ..) from **in-weights** graph.



No one correct answer here! Must produce diverse, unseen siblings/triangles

2. Exploratory creativity

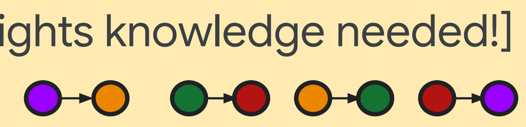
Constructing patterns resolvable in novel ways per some rules (e.g., stories, puzzles, word problems)



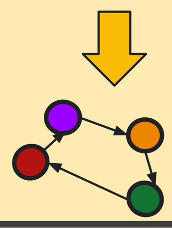
Our algorithmic tasks: Generate adjacency lists that resolve to a circle / line / ... through novel steps.

[**No** in-weights knowledge needed!]

Generate:

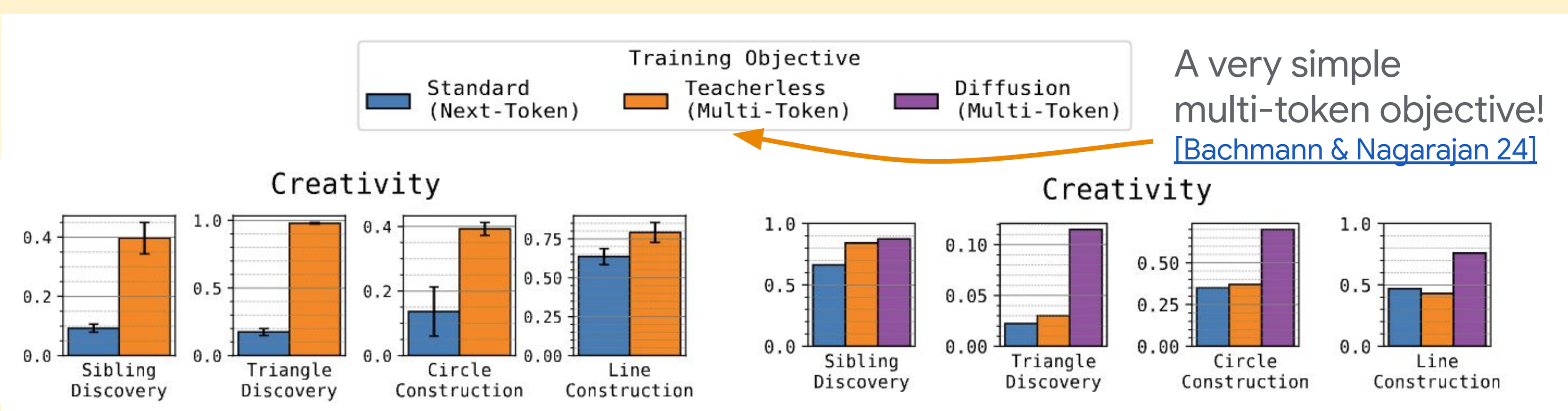


Such that:



[1] Margaret Boden 2003, *The creative mind: Myths and mechanisms*

Result 1: Multi-token learning boosts creativity over NTP



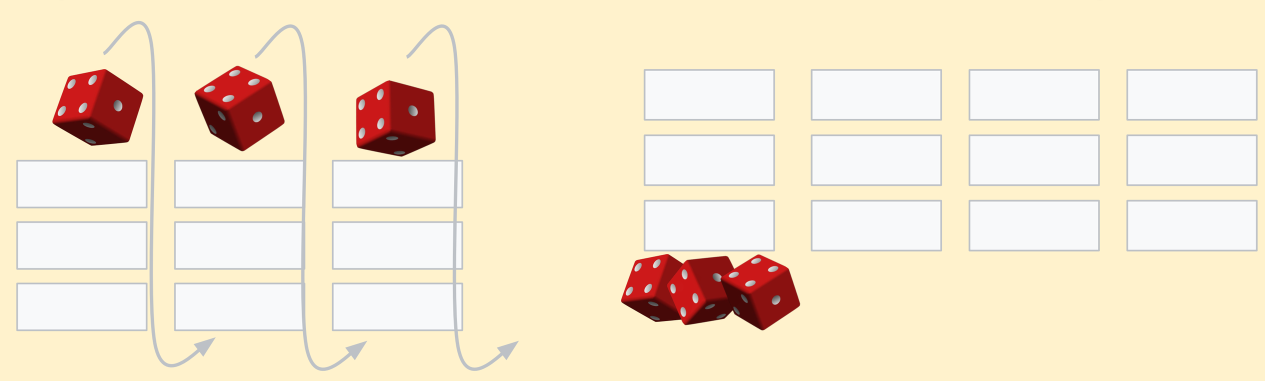
Gemma 2B   GPT-2 (86M)

Intuition : Creative tasks require planning for **global constraints** ; NTP succumbs to "Clever Hans" ***shortcuts*** .

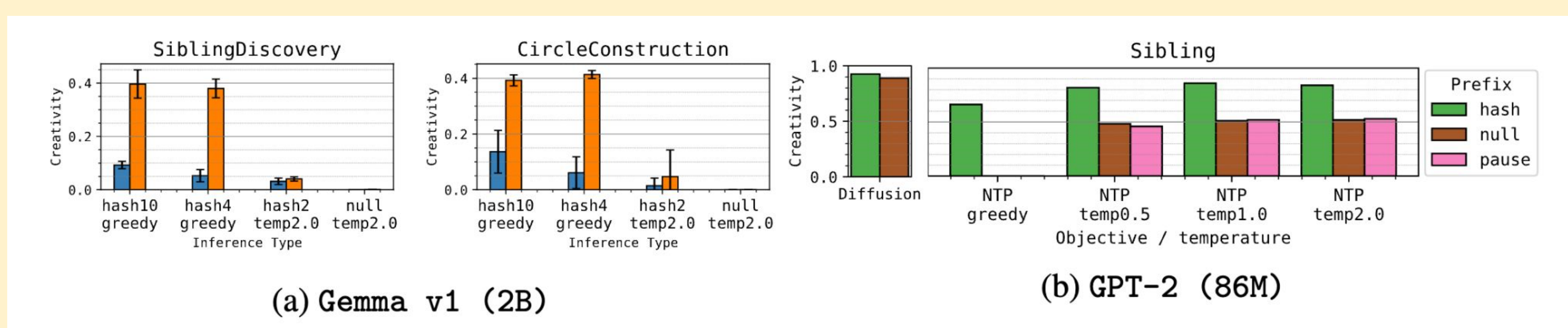
Result 2: "Hash-conditioning" improves creativity over temperature sampling

Hash-conditioning:

1. During training: prefix unique hash string per point
2. During inference: prompt with novel hash strings



Temperature sampling   **Hash-conditioning**



(a) Gemma v1 (2B)   (b) GPT-2 (86M)

Intuition : Hash-conditioning fixes dice upfront; then, easier to plan/coordinate multiple random choices.

What's next?

- Generalize results to non-algorithmic, linguistic results
- Consider other forms of creativity
- Multi-token prediction / hash-conditioning for
  - Test-time compute
  - Synthetic data generation