Open Access

Milestone 012 — Natural-source task distinctions

Tyler J.H.G.

Abstract

Can compact states distinguish the same context leading to different futures?

Research period: September 2026
Historical anchors: #178–#180, #316, #318, #323, #325

The unit models occupied a compact region of their available state space. Effective rank fell, vectors became more similar, and the geometric diagnostics looked like compression. I still needed to know what had been compressed away.

Consider two representations. One discards most variation while retaining every distinction needed for a task. The other retains many dimensions but merges two cases that require different predictions. Rank and variance can describe both states, yet they cannot tell us which one supports the task.

I needed data that exposed a predictive distinction directly: **the same context leading to different futures**. Then I could test whether the compact state retained information about that distinction.

1. Visual intuition

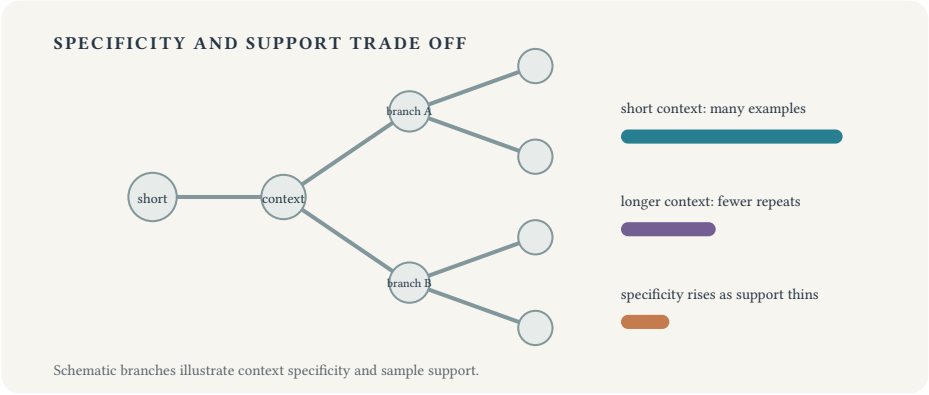


Figure 1: A schematic branch shows the tradeoff between context specificity and repeated observations.

2. Why Wikipedia revision history?

Published Sep 29, 2026

Two versions of a Wikipedia article can share a prefix and diverge where an editor changes the continuation. That gives us natural text with closely controlled context and different observed futures.

The revision history also preserves the source of each case. I can identify the page and revision, detect duplicates and reverts, and group related revisions when splitting or evaluating the data.

This supports a specific question: given the shared or closely controlled context, can the representation distinguish the continuations that branch in the source? We can test that prediction instead of using geometric spread as a proxy for usefulness.

3. Building a benchmark that can fail

The first check is whether a constant representation fails. If every input can map to the same state and still satisfy the criterion, the benchmark has no basis for claiming that input distinctions were preserved.

I required constant and mean-state controls to fail on held-out branch cases. The protocol also compared against established sparse-context baselines, froze context lengths before qualifying evaluation, collapsed duplicates and reverts, and grouped related cases by page and revision lineage.

The grouping follows the earlier source-unit correction. Thousands of tokens from related revisions provide many measurements, but their independence still depends on the sources that produced them.

> **Sticky note — experimental unit:** the independent unit contributing evidence to a comparison. Several measurements from one source can be useful while remaining correlated.

The bounded benchmark qualified at a context length of **32 bytes**. I froze that task-distinction dataset for the next stage.

4. Research context

The dynamic-sufficiency question in Milestone 010 defines a useful representation by the future distinctions it preserves. Revision branches made one such distinction observable in natural data, giving the representation an empirical target.

There is a second requirement: a prediction concerns possible outcomes, while a single recorded continuation is only one outcome. The evaluation needs to respect that difference.

> **Sticky note — proper scoring rule:** a rule for evaluating predicted probabilities whose expected score is best when we report the probabilities we actually believe. Log loss is one example.

Repeated or related branch evidence can help estimate the predictive distinction. One next token, on its own, cannot define the conditional distribution of plausible continuations.

5. Try a small example

Five invented continuations show how counts become an empirical probability distribution. Change the continuations and compare the resulting probabilities.

```
from collections import Counter
continuations = [b'a',b'a',b'b',b'a',b'b']
counts=Counter(continuations); n=sum(counts.values())
print({k.decode():v/n for k,v in counts.items()})
# A branch needs multiple supported continuations; one realized token is
not a conditional distribution.
```

Saved output:

```
{'a': 0.6, 'b': 0.4}
```

6. Interpretation

The qualified benchmark provides a specific predictive distinction under a frozen natural-source protocol. I can now ask three separate questions about it: does the state retain the signal, can a small consumer recover it, and does the model's own consumer use it?

A compact state may preserve that distinction, and a larger state may lose it. Testing the distinction directly makes either outcome observable.

7. Sources and chronology

- Milestone notebook 012 — preserved chronological source
- Research chronology
- Publication dispositions and Atlas-to-article map

The matching Atlas theme is **Natural-source context support**. The earlier Visual Intuition Atlas notebook preserves the source visualization.

The next chronological article is **Milestone 013 — Accessible does not imply used**.