

Appendix

Table of Contents

A Methods	1
A.1 Stimuli Statistics	1
A.2 Prompts and Completions	2
A.3 Narratives used in the study	7
A.4 List of Nouns	15
A.5 Story used for LLM reliability assessment	16
B Additional Results	17
B.1 Reliability of LLM scoring	17
B.2 Compressed Recall in Longer Stories	20
B.3 Lack of Output Interference in Recognition Experiments	21
B.4 Recall Probability Curves and Distributions	21
B.5 Semantic Similarity and Recall	25
B.6 Clauses with High Recall Probability	28

A Methods

Here we describe our experiments and data analysis methods in more detail.

A.1 Stimuli Statistics

Table S1: Coherent story stimuli: the first column lists the story ID in our dataset, and the following columns show: story length in clauses (L), approximate number of words (L_{words}), string length of full narrative (i.e. number of characters L_{char}), and an estimate of the total duration (in seconds) of learning phase of the experiment T . The final two columns show the number of participants we ran for the free recall (N_R) and the recognition (N_M) experiments with the corresponding story.

story	L	L_{words}	L_{char}	$T(\text{sec})$	N_R	N_M
schissel-v1	18	147	803	67	56	100
schissel-v2	18	149	775	65	100	100
boyscout	19	142	711	59	100	99
triplett-v1	32	254	1329	111	100	100
triplett-v2	32	256	1291	108	101	100
hester-v1	54	478	2414	201	98	99
hester-v2	54	617	3226	269	100	99
stein	130	1469	7323	610	99	99
panic	56	374	1890	158	100	

Table S2: Scrambled story stimuli: the first column lists the name of the story which was chosen for scrambling at the level of clauses. The next columns show: story length in clauses (L), approximate number of words (L_{words}), string length of full narrative (i.e. number of characters L_{char}), and an estimate for the total duration (in seconds) of learning phase of the experiment T . The final two columns show the number of participants we ran for the free recall (N_R) and the recognition (N_M) experiments with corresponding story.

story	L	L_{words}	L_{char}	$T(\text{sec})$	N_R	N_M
boyscout	19	142	715	60	100	100
triplett-v1	32	254	1333	111	100	100
hester-v2	54	618	3227	269	101	99
stein	130	1425	7176	598	100	99

Our main unit of measurement for the length of a narrative is the number of clauses. For completeness, we present here further statistics of the stories used in our experiments. We compute an estimate of the number of words by removing all punctuation and counting the number of text units separated by spaces. The total number of characters is computed using string length in Python, which includes all punctuation as well as spaces. The character count is used to estimate the total duration of the learning phase, the part of the experiments in which subjects read the story in a scrolling marquee which moves characters across the screen at a rate of approximately 12 characters per second.

Note that the the scrambled version of the stein story has a different number of words than the coherent version. This is due to a processing step we applied before scrambling, wherein we cleared up some typographical quirks of the original transcription of stein. For instance, transcriptions a stutter such as "I - I guess" were replaced with "I guess". This processing left the number of clauses invariant, and left the meaning of each clause unchanged. As such, it does not affect our analysis in the manuscript.

Finally, we included the details of the “panic” story which was only used for reliability scoring in our manuscript (see Fig. Figure 1 and Appendix A.5)

A.2 Prompts and Completions

A significant part of the successful use of LLMs is the effective design of inputs or “prompts”, which amount to instructions, written in standard English, for carrying out a particular task.

More precisely, the input to a LLM is a string, which we interchangeably refer to as the prompt or the context. The output of the LLM is also a string. This is also referred to as a “completion”, since LLMs are trained to complete text fragments given to them as inputs.

Despite the accumulation of wisdom on prompt engineering, we still resorted to a good amount of trial-and-error and fine-tuning in designing the prompts used in our analysis. Since these prompts are tantamount to algorithms written for LLMs, we present them below in pseudocode boxes. In the following sections, we exhaustively detail the prompts used, along with examples of LLM outputs (i.e. completions). The specific model we utilize in most of the paper is OpenAI’s GPT-4. For the data analysis, in the interest of reproducibility, we opted to use the deprecated model `gpt-4-0613` which does not receive updates. In appendix B.1.1, we show how scoring using the latest model `gpt-4` can change with time as a result of OpenAI’s regular model updating.

A.2.1 Narrative Generation

We started with a template narrative, selected from the collection of oral narratives in (Labov, 2013), and instructed the LLM to produce variations of the story which changed the surface form, but kept the overall structure (e.g. number of clauses). For these stories, we used the segmentation given by Labov (2013), wherein the narratives are segmented word-for-word into linguistic clauses. Here is a sample of the prompt, with arguments shown presented in a box enclosed in brackets, like `{this}`.

Prompt 1 Narrative Generation

```
This is a true personal narrative about a single event in someone’s life. It has
exactly {N} clauses:
{ template narrative }
```

```
Generate a new personal narrative that is unique and about something completely
different. Try to keep the overall narrative structure of the personal narrative
above, but change as much of the subject matter and action as possible. Do
not just use the narrative and replace key persons, places and things. Make it
completely new. This new narrative must also contain exactly {N} clauses.
```

As an example, we used the story of Jacob Schissel from (Labov, 2013) as a template narrative which has $N = 18$:

Argument 1 Example `template narrative`

1. My brother put a knife in my head.
2. This was just a few days after my father had died
3. and we were sitting shiva.
4. And the reason the fight started,
5. he saw a rat out in the yard or
6. - this was out in Coney Island -
7. and he started talk about it.
8. And my mother had just sat down to have a cup of coffee
9. and I told him to cut it out.
10. 'Course kids, y'know, he don't hafta listen to me.
11. So that's when I grabbed his arm
12. and twisted it up behind him.
13. When I let go his arm, there was a knife on the table,
14. he just picked it up
15. and he let me have it.
16. And . . .; I started bleeding - like a pig.
17. And naturally first thing to do, run to the doctor,
18. and the doctor just says, "Just about this much more," he says, "and you'd a been dead.

The output of the LLM upon processing this prompt was a story which clearly bore a structural relation to the original, but contained different agents involved in a completely different scenario. We provide an example completion below obtained using gpt-4 with temperature $T = 0.6$. This example was used in experiments, and also appears in [Story 1](#) next to another generated narrative ([Story 2](#)).

Completion 1 from [Prompt 1, Argument 1](#)

1. My best friend pushed me into the pool.
2. This was right after I had broken up with my boyfriend
3. and I was feeling really down.
4. The reason she did it,
5. she thought I needed a distraction or
6. - this was at her house in the suburbs -
7. and she wanted to cheer me up.
8. I had just sat down with a plate of snacks
9. and I told her I was fine.
10. 'Course friends, y'know, they don't always believe you.
11. So that's when she sneaked up behind me
12. and gave me a little shove.
13. When I surfaced from the water, there was a float nearby,
14. she just grabbed it
15. and tossed it to me.
16. And . . .; I started laughing - uncontrollably.
17. And naturally the first thing to do, hug my friend,
18. and I just said, "Just a little more," I said, "and I would've been drowning in my sorrows.

The resulting narratives generated from this procedure vary in their coherence. For our experiments, we chose the most coherent generated narratives.

A.2.2 Lure Generation

We utilized `gpt-4` at $T = 0.3$ with the following prompt:

Prompt 2 Lure Generation

{segmentation}

The items above all fit together to tell a story. Add more items of roughly the same length, numbered 1.5, 2.5, and so on, interleaving the existing items, elaborating on the story, and without repetition. These new items should introduce completely new plot elements, but still make sense in the context of the rest of the story. Add as many items as possible.

The following is an example completion using the template narrative segmentation ([Argument 1](#)). For purposes of presentation, we present the lures along with the original clauses from the narrative. The completion does *not* include the original clauses, which appear in gray, and only consists of the text shown in black.

Completion 2 from [Prompt 2](#)

1. My brother put a knife in my head.
- 1.5. It was a tense time for our family, with everyone on edge.
2. This was just a few days after my father had died
- 2.5. We were all grieving in our own ways, trying to cope with the loss.
3. and we were sitting shiva.
- 3.5. Our house was filled with relatives and friends, offering their condolences.
4. And the reason the fight started,
- 4.5. My brother had always been a bit of a troublemaker, but I never thought he'd go this far.
5. he saw a rat out in the yard or
- 5.5. He was always fascinated by the creatures that roamed the streets and alleys.
6. - this was out in Coney Island -
- 6.5. The sun was setting, casting an eerie glow on the scene.
7. and he started talk about it.
- 7.5. He started describing the rat in gruesome detail, trying to get a rise out of us.
8. And my mother had just sat down to have a cup of coffee
- 8.5. She looked so tired and worn out, I just wanted her to have a moment of peace.
9. and I told him to cut it out.
- 9.5. I tried to keep my voice calm, but I could feel my anger rising.
10. 'Course kids, y'know, he don't hafta listen to me.
- 10.5. He just smirked at me, clearly enjoying the attention.
11. So that's when I grabbed his arm
- 11.5. I didn't want to hurt him, just get him to stop his nonsense.
12. and twisted it up behind him.
- 12.5. He let out a yelp of pain, but I didn't think I'd done any real damage.
13. When I let go his arm, there was a knife on the table,
- 13.5. I didn't even see him reach for it, he was so quick.
14. he just picked it up
- 14.5. I barely had time to react before the blade was coming towards me.
15. and he let me have it.
- 15.5. The pain was sharp and intense, like nothing I'd ever felt before.
16. And . . .; I started bleeding - like a pig.
- 16.5. I could feel the warmth of my blood soaking through my clothes.
17. And naturally first thing to do, run to the doctor,
- 17.5. I stumbled out of the house, desperate for help.
18. and the doctor just says, "Just about this much more," he says, "and you'd a been dead.
- 18.5. I couldn't help but think how close I'd come to joining my father.

A.2.3 Recall Scoring

In order to score recall, we utilize zero-shot prompting with OpenAI’s deprecated model `gpt-4-0613`. We chose to use this particular model for the benefit of reproducibility, since we have found that results from the following analysis can drift in time as OpenAI updates its model (see Appendix B.1.1).

The recall scoring **Prompt 3**, shown below, was a block of text, represented as a python string, which took as arguments the original narrative, the numbered segmentation of the narrative, and a single participant’s recall of the narrative. The prompt was constructed so that it would identify which clauses from the narrative are present in the recall. A consequence of this scoring procedure is that a single clause in the recall can encode multiple clauses from the original narrative. `gpt-4` appeared to have no trouble identifying such compressed clauses in the recall, and frequently would identify multiple original clauses within a single clause in the participant’s recall. An example of this can be seen in **Completion 3**, where the same passage in the recall (“even the scoutmaster looked on”) is counted as recalling both clauses 15 and 16 from the original narrative (**Argument 2**).

Below, we provide the scoring prompt, as well as some example arguments, and the resulting LLM completion.

Prompt 3 Recall Scoring

```
This is the original text:
{narrative}
```

```
It can be broken down into the following independent pieces of information:
{segmentation}
```

```
Here is an alternative version of the original text where some of the above pieces
of information may be missing:
{recall}
```

```
For each of the numbered information pieces of the list above, evaluate whether
the information of each piece is given in the alternative version of the story,
stating the number and showing the corresponding passage from the alternative story
it is given in. After, write all the numbers of the pieces that are given in the
alternative version of the story in a set of brackets at the end of the response.
```

Argument 2 Example `narrative`

```
Yeah, I was in the boy scouts at the time. And we was doing the 50-yard dash,
racing, but we was at the pier, marked off and so we was doing the 50-yard dash.
There was about 8 or 9 of us, you know, going down, coming back. And going down
the third time, I caught cramps and I started yelling "Help!", but the fellows
didn't believe me, you know. They thought I was just trying to catch up, because I
was going on or slowing down. So all of them kept going. They leave me. And so I
started going down. Scoutmaster was up there. He was watching me. But he didn't
pay me no attention either. And for no reason at all there was another guy, who
had just walked up that minute... He just jumped over and grabbed me.
```

Argument 3 Example **segmentation**

1. Yeah, I was in the boy scouts at the time.
2. And we was doing the 50-yard dash,
3. racing,
4. but we was at the pier, marked off,
5. and so we was doing the 50-yard dash.
6. There was about 8 or 9 of us, you know, going down, coming back.
7. And, going down the third time, I caught cramps
8. and I started yelling "Help!",
9. but the fellows didn't believe me, you know.
10. They thought I was just trying to catch up, because I was going on or slowing down.
11. So all of them kept going.
12. They leave me.
13. And so I started going down.
14. Scoutmaster was up there.
15. He was watching me.
16. But he didn't pay me no attention either.
17. And for no reason at all there was another guy, who had just walked up that minute...
18. He just jumped over
19. and grabbed me.

Argument 4 Example **recall**

A boy was with the scoutmaster and he was going up and down and on the third time of going up the pier he had cramps and shouted 'help' but no-one believed him, even the scoutmaster looked on by until this guy came out of nowhere and grabbed the boy and saved him.

Completion 3 from **Prompt 3**

1. Not given
2. Not given
3. Not given
4. Not given
5. Not given
6. Not given
7. Given - "on the third time of going up the pier he had cramps"
8. Given - "and shouted 'help'"
9. Given - "but no-one believed him"
10. Not given
11. Not given
12. Not given
13. Not given
14. Given - "was with the scoutmaster"
15. Given - "even the scoutmaster looked on"
16. Given - "even the scoutmaster looked on"
17. Given - "until this guy came out of nowhere"
18. Not given
19. Given - "and grabbed the boy and saved him"

(7, 8, 9, 14, 15, 16, 17, 19)

We note here that the completion does not always appear as in this example. Some minor deviations might occur (e.g. parentheses replaced with brackets), or the format of the answers change (e.g. instead of "Not Given", some completions give "There is no corresponding passage for").

A.2.4 Ordered Recall Scoring

In order to extract the order in which the original clauses appeared in a participant’s recall, we added an extra step to the recall scoring discussed above. For a given narrative, segmentation, and recall, we concatenated **Prompt 3** with its **Completion 3**, and appended to this the following additional prompt:

Prompt 4 Ordered Recall Scoring

Now repeat the alternative version of the narrative with the number of the independent piece of information inserted next to the location in which it appears in the alternative version. Then, list the numbers separately in the order in which they appear in the alternative story immediately above. The final list of numbers should be enclosed in parentheses.

Therefore, the total input or prompt for ordered recall scoring would be **Prompt 3** + **Completion 3** + **Prompt 4**. The output for this example was the following:

Completion 4 from **Prompt 3** + **Completion 3** + **Prompt 4**

A boy (14) was with the scoutmaster and he was going up and down and on the third time of going up the pier he had cramps (7) and shouted 'help' (8) but no-one believed him (9), even the scoutmaster looked on (15, 16) by until this guy came out of nowhere (17) and grabbed the boy and saved him (19).

(14, 7, 8, 9, 15, 16, 17, 19)

A.3 Narratives used in the study

The narratives that were generated by GPT-4 for this study used an existing segmentation into clauses. The original segmentation was taken from (Labov, 2013; Labov and Waletzky, 1966). These segmentations were provided to the LLM to generate new variants, as detailed in appendix A.2.1. For each narrative template, we generated about 10 variants, and selected two for use in our experiments, based on our judgment of their coherence. With the exception of the longest narrative ($L = 130$), we present below all narratives that were presented to participants in our experiments. In the examples, the narratives are labeled by a prefix reflecting the source story, with the exception of “panic attack” which was created by GPT-3. The scrambled narratives are based on intact narrative (without “-scrambled” in the narrative name) where all clauses were shuffled. The “panic” story was generated based on the “boyscout” story from (Labov and Waletzky, 1966), **Story 7** below. The story was generated using **Prompt 6**.

Story 1 and Story 2 below were generated using Prompt 1, with the Jacob Schissel story Argument 1 from (Labov, 2013) as a template.

Example Story 1 Narrative schissel-v1-pool	Example Story 2 Narrative schissel-v2-lake
1. My best friend pushed me into the pool. 2. This happened during my cousin's wedding reception, 3. where everyone was dressed to impress. 4. And the reason it happened, 5. she spotted a bee hovering near my face, 6. - this was at a fancy hotel garden - 7. and she tried to save me from being stung. 8. And my aunt had just handed me a glass of champagne, 9. and I warned her to be careful. 10. 'Course friends, y'know, they don't always think things through. 11. So that's when she gave me a little shove, 12. and I tumbled into the water. 13. When I resurfaced, gasping for air, 14. she just started laughing, 15. and she apologized profusely. 16. And... my beautiful dress was ruined, 17. and naturally, the first thing to do was to get out and dry off, 18. and my cousin just says, "Just about a few inches more," she says, "and you'd have landed on the cake."	1. My best friend pushed me into the lake. 2. This happened during our annual summer camping trip 3. and we were enjoying a peaceful afternoon. 4. The reason it all began, 5. she spotted a frog near the water's edge or 6. - this was at our favorite spot in the woods - 7. and she dared me to catch it. 8. I had just sat down to eat my sandwich 9. and I told her I wasn't interested. 10. 'Course friends, y'know, they don't always take no for an answer. 11. So that's when she snuck up behind me 12. and gave me a little nudge. 13. When I emerged from the water, there was a frog on the shore, 14. she just scooped it up 15. and handed it to me. 16. And . . .; I couldn't help but laugh – uncontrollably. 17. And naturally, the first thing to do was to get revenge, 18. and I said, "Just you wait," I said, "you'll get yours soon enough."

Story 3 and Story 4 below were generated using Prompt 1, with the Charles Triplet story from (Labov, 2013) as a template.

Example Story 3 Narrative triplet-v1-rookie	Example Story 4 Narrative triplet-v2-cat lady
1. Back in the days when I was a rookie in the police force, the Chief was a veteran. 2. He had caught a notorious criminal – 3. or had been awarded for bravery. 4. But he had – a young daughter. 5. And in those days I was a fitness enthusiast. 6. And it seemed she was trying to impress me. 7. I never observed it. 8. Truth is I wasn't very fond of her because she was – 9. She was a charming girl until she opened her mouth. 10. She was a chatterbox. 11. Goodness, she could talk. 12. Then she left a message one day saying she was going to run away because he was always scolding her about me. 13. He arrived at my apartment. 14. Big intimidating look on his face. 15. I managed to calm him down. 16. and proposed "Well we'll search for her 17. and if we can't locate her well you can – do whatever you think is right." 18. I was strategizing. 19. So he accepted my offer. 20. And we went to where they found her scarf – near a park. 21. and we traced down a little more 22. and we couldn't locate her. 23. And returned 24. – it was a police station – 25. she was sitting on a chair with a book in her hands. 26. She hadn't run away. 27. But – nevertheless – that resolved the issue for that day. 28. But that evening the deputy, Frank Mitchell, said "You better transfer and leave because that old man never forgets anything once he gets it into his mind." 29. And I did. 30. I transferred. 31. and left. 32. That was the end of my rookie year.	1. In the neighborhood I grew up in, Mrs. Baker was the resident cat lady. 2. She had adopted at least twenty – 3. or had given them homes. 4. But she had – a young grandson 5. and back then I was quite the tomboy. 6. And apparently, he was trying to befriend me. 7. I never caught on. 8. Truth is I didn't take to him much because he had – 9. He was a charming lad until you saw his manners. 10. He had poor manners. 11. Goodness gracious he had poor manners. 12. Then he left a note one day saying he was going to run away because she was always nagging about me. 13. She came to my treehouse. 14. Big old tabby cat in tow. 15. I talked her down 16. and said "Well we'll go find him 17. and if we can't locate him, you can – go ahead and call the police if you need to." 18. I was strategizing. 19. So she took my advice. 20. And we went to where they found his baseball cap – near the old mill 21. and we trailed a bit more 22. and we didn't find a trace. 23. And returned 24. – it was a treehouse club – 25. he was sitting on a branch with a comic book in his hand. 26. He hadn't run away. 27. But – however – that resolved it for that day. 28. But that evening the local sheriff, Officer Dawson said "You better lay low and stay out of sight because that woman never forgets anything once she gets it into her mind." 29. And I did. 30. I laid low 31. and stayed out of sight. 32. That was the second time.

Story 5 and Story 6 below were generated using Prompt 1, with the Adolphus Hester story from (Labov, 2013) as a template.

Example Story 5 Narrative hester-v1-park

1. Let me share a story from when I was twenty-one years old
2. I was living in a small apartment near the city center
3. I had just started my first job
4. And uh I was working as a graphic designer
5. My father, who was recovering from a surgery, hadn't been out in – in uh – well I'd say about eight months
6. He had to use a wheelchair to get around
7. The only person there with him during the day was a kind neighbor, a lady in her sixties
8. My younger sister was away at college –
9. so it was just my dad and the neighbor lady at home during the day
10. One day, I got this strong feeling while I was at work
11. It told me: "If you go to the nearby park during your lunch break and pray for your father's recovery, he'll start to improve
12. I looked around my office
13. I didn't see anyone nearby
14. So, I went to the park during my break
15. and found a quiet spot there
16. That feeling came back again
17. I looked around once more
18. I still didn't see anyone nearby
19. but I decided to take a chance
20. I wasn't too far from my office, maybe a ten-minute walk
21. I found a bench to sit on and closed my eyes – uh – uh –
22. that feeling came back once more
23. I just took a deep breath
24. I was sitting under a big oak tree now
25. I looked up at the branches above me
26. Nobody was around
27. And uh I said to myself, well now,
28. I had been taught that prayers can make a difference
29. And that it doesn't hurt to try
30. I knew I wasn't perfect, but I wanted to help my dad
31. I'd admit that to anyone
32. But I was willing to give it a try
33. I looked around one more time
34. and then I closed my eyes
35. and said a quiet prayer
36. I whispered, "God, I don't know if you're listening
37. But I ask you to please help my father recover
38. So he can regain his independence and enjoy life again
39. I opened my eyes
40. and went back to work
41. Suddenly, I received a text message from my father
42. It said, "I just took a few steps with my walker
43. I – I couldn't believe it
44. It felt like a sign that my prayer had been heard
45. I rushed home after work
46. and saw my father standing in the living room with the help of his walker

47. He hadn't been able to do that in months without assistance
48. And there he was, taking small steps
49. That was the turning point in his recovery
50. He continued to improve
51. and regained most of his mobility within a year
52. It's been, what, six or seven years since then?
53. My father is now able to walk without any assistance
54. He's living life to the fullest, even in his seventies

Example Story 6 Narrative hester-v2-church

1. Let me share a story from when I was twenty-two years old, just fresh out of college.
2. I was staying in a small town about five miles from here.
3. I had just started my first job at a local bookstore.
4. It was a quiet afternoon, and I was alone in the store.
5. My grandpa, who had been bedridden for almost a year, was at home with my sister.
6. She had to care for him daily, as he couldn't move around on his own.
7. There were only two of us siblings, my sister and me.
8. My parents had gone out of town for the weekend.
9. As I was shelving books, I suddenly heard a voice inside my head.
10. It said, "If you go to the old church down the street and light a candle for your grandpa, he will get better."
11. I looked around the bookstore, but no one was there.
12. I continued with my work, trying to ignore the voice.
13. It spoke to me again, repeating the same message.
14. I looked around once more, but still found no one.
15. I decided to take a break and walk to the church.
16. I was only a short distance away, maybe a quarter-mile.
17. As I approached the church, I hesitated for a moment.
18. I was not a particularly religious person.
19. I remembered my grandma always saying that prayers from a sincere heart could bring miracles.
20. On the other hand, she also said that prayers without faith were useless.
21. I admitted to myself that I was far from being a saint.
22. I had my share of mistakes and misdeeds.
23. I thought about my grandpa and how much he meant to our family.
24. Finally, I went inside the church and found a candle.
25. I knelt down in front of the altar.
26. I said, "God, I don't know how to pray properly.
27. But I am asking you with all my heart to heal my grandpa.
28. Please help him regain his strength and guide us in taking care of him."
29. After saying my prayer, I returned to the bookstore.
30. I resumed my work, shelving books.
31. Suddenly, I felt a strong urge to go home.
32. It was as if something was pulling me towards the house.
33. I didn't hear any voice this time.
34. It was just an overwhelming feeling that I needed to go.
35. I locked up the store and rushed home.
36. To my surprise, I found my grandpa sitting in the living room, talking to my sister.
37. He hadn't been able to leave his bed for months without assistance.

38. My sister told me that he had suddenly felt a surge of energy and was able to stand up and walk.
39. That day marked the beginning of his recovery.
40. He gradually regained his strength and independence.
41. It was a turning point in our lives.
42. We became more grateful for the time we had together.
43. I often thought back to that day in the church.
44. I wondered if my prayer had anything to do with my grandpa's recovery.
45. I couldn't be sure, but it definitely changed my perspective on faith and the power of prayer.
46. My grandpa lived for many more years after that incident.
47. He celebrated his ninetieth birthday surrounded by family and friends.
48. We cherished every moment we had with him.
49. That experience happened almost twenty years ago.
50. It remains a vivid memory in my mind.
51. My grandpa passed away a few years later, but his spirit lives on in our hearts.
52. Sometimes, when I'm feeling lost or uncertain, I think back to that day.
53. I remember the power of faith and the miracles it can bring.
54. And I am reminded that we are never truly alone, even when it seems like we are.

Below, **Story 7** is taken directly from (Labov and Waletzky, 1966). **Story 8** is the scrambled version used in experiments.

Example Story 7 Narrative boyscout	Example Story 8 Narrative boyscout-scrambled
1. Yeah, I was in the boy scouts at the time. 2. And we was doing the 50-yard dash, 3. racing, 4. but we was at the pier, marked off, 5. and so we was doing the 50-yard dash. 6. There was about 8 or 9 of us, you know, going down, coming back. 7. And, going down the third time, I caught cramps 8. and I started yelling "Help!", 9. but the fellows didn't believe me, you know. 10. They thought I was just trying to catch up, because I was going on or slowing down. 11. So all of them kept going. 12. They leave me. 13. And so I started going down. 14. Scoutmaster was up there. 15. He was watching me. 16. But he didn't pay me no attention either. 17. And for no reason at all there was another guy, who had just walked up that minute... 18. He just jumped over 19. and grabbed me.	1. They leave me. 2. racing, 3. Yeah, I was in the boy scouts at the time. 4. but we was at the pier, marked off, 5. and so we was doing the 50-yard dash. 6. And we was doing the 50-yard dash, 7. but the fellows didn't believe me, you know. 8. and I started yelling "Help!", 9. and grabbed me. 10. Scoutmaster was up there. 11. So all of them kept going. 12. And so I started going down. 13. There was about 8 or 9 of us, you know, going down, coming back. 14. He just jumped over 15. They thought I was just trying to catch up, because I was going on or slowing down. 16. And, going down the third time, I caught cramps 17. And for no reason at all there was another guy, who had just walked up that minute... 18. He was watching me. 19. But he didn't pay me no attention either.

Story 9 and Story 10 are the scrambled versions of Story 3 and Story 6, respectively. We do not reproduce here the scrambled version of the longest story ($L = 130$). These stimuli can be found in our code and data repository.

The scrambling is done by taking a random permutation of the clauses. The numbering in the scrambled versions reflects the presentation order.

Example Story 9 Narrative triplett-v1-rookie-scrambled

1. and we traced down a little more
2. and left.
3. And I did.
4. I never observed it.
5. I was strategizing.
6. Big intimidating look on his face.
7. and proposed “Well we’ll search for her
8. Back in the days when I was a rookie in the police force, the Chief was a veteran.
9. and we couldn’t locate her.
10. And in those days I was a fitness enthusiast.
11. And we went to where they found her scarf – near a park.
12. But – nevertheless – that resolved the issue for that day.
13. That was the end of my rookie year.
14. And it seemed she was trying to impress me.
15. or had been awarded for bravery.
16. He had caught a notorious criminal –
17. She hadn’t run away.
18. She was a chatterbox.
19. I managed to calm him down.
20. – it was a police station –
21. And returned
22. Truth is I wasn’t very fond of her because she was –
23. Then she left a message one day saying she was going to run away because he was always scolding her about me.
24. But that evening the deputy, Frank Mitchell, said “You better transfer and leave because that old man never forgets anything once he gets it into his mind.”
25. I transferred.
26. But he had – a young daughter.
27. Goodness, she could talk.
28. So he accepted my offer.
29. She was a charming girl until she opened her mouth.
30. she was sitting on a chair with a book in her hands.
31. He arrived at my apartment.
32. and if we can’t locate her well you can – do whatever you think is right.”

Example Story 10 Narrative hester-v2-church-scrambled

1. I remembered my grandma always saying that prayers from a sincere heart could bring miracles
2. Finally, I went inside the church and found a candle
3. As I was shelving books, I suddenly heard a voice inside my head
4. I remember the power of faith and the miracles it can bring
5. It was a turning point in our lives
6. That experience happened almost twenty years ago

7. I was only a short distance away, maybe a quarter-mile
8. My grandpa lived for many more years after that incident
9. I said, "God, I don't know how to pray properly
10. And I am reminded that we are never truly alone, even when it seems like we are
11. It spoke to me again, repeating the same message
12. I was not a particularly religious person
13. Suddenly, I felt a strong urge to go home
14. I looked around the bookstore, but no one was there
15. I was staying in a small town about five miles from here
16. My sister told me that he had suddenly felt a surge of energy and was able to stand up and walk
17. Sometimes, when I'm feeling lost or uncertain, I think back to that day
18. I wondered if my prayer had anything to do with my grandpa's recovery
19. Let me share a story from when I was twenty-two years old, just fresh out of college
20. My parents had gone out of town for the weekend
21. I locked up the store and rushed home
22. It was as if something was pulling me towards the house
23. I admitted to myself that I was far from being a saint
24. I had just started my first job at a local bookstore
25. As I approached the church, I hesitated for a moment
26. She had to care for him daily, as he couldn't move around on his own
27. It was a quiet afternoon, and I was alone in the store
28. We cherished every moment we had with him
29. My grandpa, who had been bedridden for almost a year, was at home with my sister
30. I had my share of mistakes and misdeeds
31. He hadn't been able to leave his bed for months without assistance
32. There were only two of us siblings, my sister and me
33. I didn't hear any voice this time
34. But I am asking you with all my heart to heal my grandpa
35. My grandpa passed away a few years later, but his spirit lives on in our hearts
36. I looked around once more, but still found no one
37. He celebrated his ninetieth birthday surrounded by family and friends
38. That day marked the beginning of his recovery
39. We became more grateful for the time we had together
40. I continued with my work, trying to ignore the voice
41. It was just an overwhelming feeling that I needed to go
42. I knelt down in front of the altar
43. I thought about my grandpa and how much he meant to our family
44. Please help him regain his strength and guide us in taking care of him
45. To my surprise, I found my grandpa sitting in the living room, talking to my sister
46. I often thought back to that day in the church
47. I decided to take a break and walk to the church
48. I couldn't be sure, but it definitely changed my perspective on faith and the power of prayer
49. After saying my prayer, I returned to the bookstore

A.4 List of Nouns

The following list of 32 nouns in exactly this order was presented to each participant in the Random List of Nouns experiment.

Example 1

'FAN', 'POLICE', 'SURVEY', 'COFFEE', 'ISLAND', 'CALF', 'ACID', 'SUITCASE', 'TIDE', 'BULLET', 'BUILDING', 'TELEVISION', 'QUEEN', 'BEDROOM', 'LIGHTNING', 'MOUNTAIN', 'LEMON', 'SEAT', 'CITY', 'VICTIM', 'LIST', 'CHAIN', 'STEM', 'ATOM', 'STAKE', 'DEER', 'OPERATOR', 'CROSS', 'FLOOR', 'INFANT', 'INDIAN', 'PASSAGE'

A.5 Story used for LLM reliability assessment

The following narrative was used for comparison of recall performance evaluation by humans versus GPT in fig. 1.

The story was generated with an older version of GPT (`text-davinci-003`) using the following prompt seeded by the boyscouts narrative ([Story 7](#), in paragraph form without the numbering):

Prompt 5 Panic Attack Story Generation

This is an oral narrative about a life-threatening event:
{story}
Create a different, longer and more detailed oral narrative with a similar narration style about another life-threatening event.

yielding the following narrative:

Argument 5 Panic Attack Narrative

Well, I was at work one day and I was doing my usual job - working on the warehouse floor, stocking boxes and organizing merchandise. I had been working for a few hours and it was getting late. I had been feeling a bit strange, a bit lightheaded, but chalked it up to hunger and not having eaten yet that day. But then, as I was walking around the warehouse returning items to their proper places, I started to feel a strange tightness and pressure in my chest. I stopped what I was doing and leaned against a nearby wall for a bit, but the pressure only got stronger. And I could feel my heart pounding in my chest. Next thing I know, I started seeing spots and I could feel sweat dripping down my forehead and my arms. I tried to take some deep breaths, but I just couldn't catch my breath. I was really scared and didn't know what to do. Suddenly I heard a voice coming from across the warehouse. It was the manager, who had seen me and called out to me - it seems someone had noticed me and was checking to see if I was alright. I tried to yell back that something was wrong, but I couldn't find my voice. The manager started walking towards me and I could feel my world going dark. I started to drop to the ground, but luckily the manager got to me in time and caught me just before my legs gave out. He took me to the staff room and sat me down in a chair. He asked me questions about what was wrong and what happened. He got me a glass of cold water and told me to rest for a bit. After a few minutes, the pressure in my chest slowly eased up and I started to feel better. It turned out I had been having a panic attack, but I escaped from the situation with the help of the quick action of the manager. It was never a fun experience, but it really could have been much worse, and it taught me to check in with myself and make sure to properly rest and eat to avoid further episodes.

The story was subsequently segmented using the prompt:

Prompt 6 Segmentation Prompt

This is the original story:
{[Argument 5](#)}
Split the sentences of the original story into the maximum number of small pieces, such that each piece contains minimal information.

The segmentation that comes up is the following:

Completion 5 Panic Attack Segmentation

1. I was at work.
2. I was working on the warehouse floor.
3. I was stocking boxes and organizing merchandise.
4. I had been working for a few hours.
5. It was getting late.
6. I had been feeling a bit strange.
7. I was feeling lightheaded.
8. I chalked it up to hunger and not having eaten yet that day.
9. I was walking around the warehouse returning items to their proper places.
10. I started to feel a strange tightness and pressure in my chest.
11. I stopped what I was doing.
12. I leaned against a nearby wall.
13. The pressure only got stronger.
14. I could feel my heart pounding in my chest.
15. I started seeing spots.
16. I could feel sweat dripping down my forehead and my arms.
17. I tried to take some deep breaths.
18. I couldn't catch my breath.
19. I was really scared.
20. I didn't know what to do.
21. I heard a voice coming from across the warehouse.
22. It was the manager.
23. Someone had noticed me.
24. The manager was checking to see if I was alright.
25. I tried to yell back that something was wrong.
26. I couldn't find my voice.
27. The manager started walking towards me.
28. I could feel my world going dark.
29. I started to drop to the ground.
30. The manager got to me in time.
31. He caught me just before my legs gave out.
32. He took me to the staff room.
33. He sat me down in a chair.
34. He asked me questions about what was wrong and what happened.
35. He got me a glass of cold water.
36. He told me to rest for a bit.
37. The pressure in my chest slowly eased up.
38. I started to feel better.
39. It turned out I had been having a panic attack.
40. I escaped from the situation with the help of the quick action of the manager.
41. It was never a fun experience.
42. It really could have been much worse.
43. It taught me to check in with myself.
44. It taught me to properly rest and eat to avoid further episodes.

Note that the segmentation in this particular case is not verbatim, the clauses have been transformed to first person while retaining most important information. If a verbatim segmentation is desired however, it can be achieved by utilizing the same prompt with `gpt-4` although the number of clauses will be slightly different.

B Additional Results

B.1 Reliability of LLM scoring

Here we report additional details of the reliability for LLM narrative scoring.

In scoring the recalls, the human scorers (who are the authors on this paper) did not develop overly sophisticate scoring protocols. The directions were intuitive, essentially boiling down to “did this clause from the original story appear in the recalled story”. Despite the ambiguity of this procedure, there was very high inter-scorer correlation.

Table S3: A comparison between gpt-4, mean human ($\bar{h}$), and individual human scorers (h_1, h_2, h_3). Each entry in the table gives the correlation coefficient (r-value) between the corresponding row and column label.

r	$\bar{h}$	h_1	h_2	h_3
gpt-4-0613	0.94	0.92	0.90	0.90
$\bar{h}$		0.95	0.98	0.97
h_1			0.89	0.86
h_2				0.96

B.1.1 Drift in Performance over time

The models offered through the OpenAI API are known to undergo regular updates. We found that within the span of two months, the correlation coefficient with mean human performance rose by 0.05. We compare gpt-4 called via OpenAI’s API on 05/23/23, the same model gpt-4 called less than two months later on 07/03/23, with the deprecated model gpt-4-0613 (called after 06/13/23). Fortunately, the frozen model gpt-4-0613, which we use throughout the paper, is essentially as good as the most recent version.

Table S4: A comparison on gpt-4 scoring compiled at different times of the year. Evidently, the model improves over time in approximating the mean human scoring, with the most recent model achieving a correlation closest to one. The entries of the table show correlation coefficients (r-values) between the corresponding row and column variables.

r	gpt-4 (05/23/23)	gpt-4-0613	gpt-4 (07/03/23)
$\bar{h}$	0.89	0.937	0.944
gpt-4 (05/23/23)		0.94	0.92
gpt-4-0613			0.99

B.1.2 Comparison between different LLMs

We compared different LLMs available through the OpenAI API. In order of increasing number of learned parameters (model size), these were GPT-3 (text-davinci-003), GPT-3.5 (or ChatGPT, gpt-3.5-turbo) and GPT-4 (gpt-4-0613). The same scoring Prompt 3 was used for all of these models. The results are visualized in Figure S1. We find that GPT-4 is qualitatively better than the smaller models (compare correlation between human and LLM scoring in Figure S1B). Furthermore, increasing size does not seem to be sufficient to increase performance, as illustrated by the severe drop in performance by GPT-3.5. As seen in Figure S1(A,B), GPT-3.5 appears to score more generously than the other models, resulting in a systematic upward bias of nearly all recall probabilities. It is possible that this reflects the sensitivity of these models to prompts, and that with appropriate tuning of the scoring prompt, the performance of GPT-3.5 could be improved.

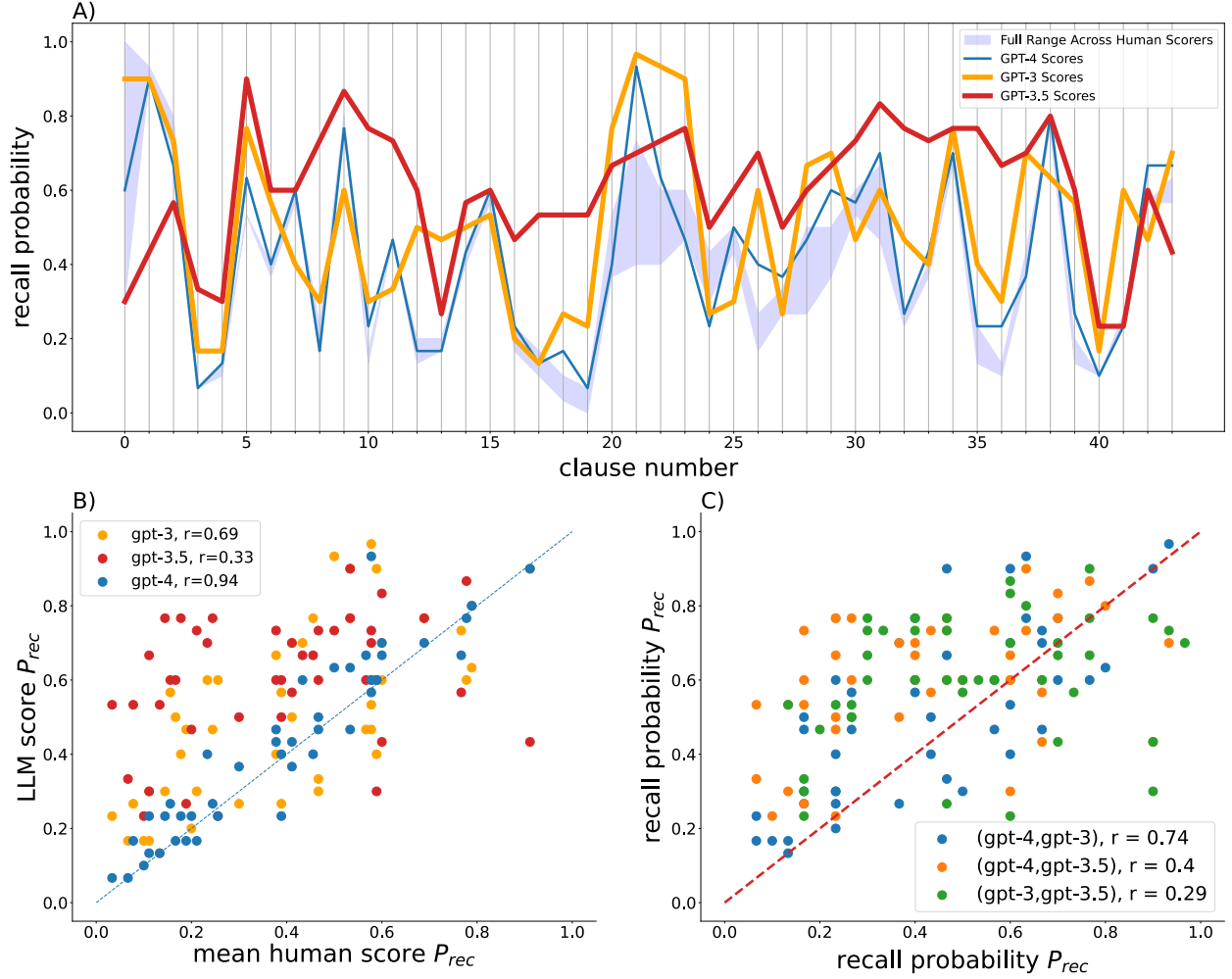


Figure S1: *Reliability of different LLMs for scoring narratives* We observed that performance on human recall scoring was not monotonic in the size of the LLM. (A) shows P_{rec} for each clause as computed by LLMs of three different sizes, in order of increasing size: GPT-3 (OpenAI’s API model `text-davinci-003`), GPT-3.5 (or ChatGPT, model `gpt-3.5-turbo`), and GPT-4 (`gpt-4-0613`). GPT-3.5 scores appear systematically biased to be higher than the mean human scores, whereas GPT-3 roughly follows the human trend. (B) shows correlations between these models and mean human performance. (C) shows inter-LLM correlations.

B.2 Compressed Recall in Longer Stories

Each recall can be split up into a number of clauses. We denote the mean number of clauses used in recalls by C . Figure S2A compares C to the mean number of recalled clauses, defined previously and denoted by R . If every clause used by the participant were recalling a single clause from the original story, then the points would lie on a straight line, corresponding to $C = R$. For shorter stories, we find in fact that $C > R$, and people appear to use more clauses than necessary. However, for the longest story ($L = 130$), there is a marked compression, in which participants on average use 75% as many clauses as they are scored to have recalled. It is possible that this reflects a tendency of GPT-4 to score generously when it encounters summaries or compressed descriptions in the recall. Figure S2B shows how both R and C scale with the estimate for the mean number of encoded clauses M . The linear fit to R is presented for comparison (also shown in Figure 2), to illustrate the deviation of C from linear scaling at longer stories.

We used **Prompt 7** with GPT-4 at $T = 0$ to segment recalls. An example of this applied to **Argument 4** is given in **Completion 6**.

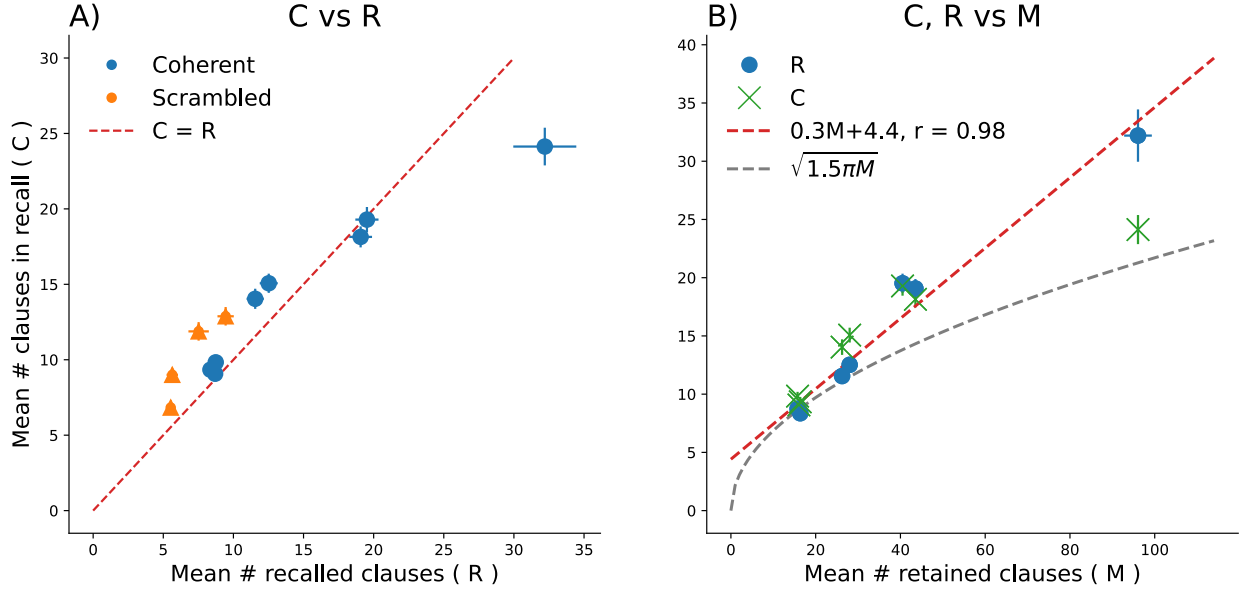


Figure S2: *Compression of Recalls*. Here we analyze how the length of recalls relates to story length and memory. A) The vertical axis shows the mean number of clauses C written for recall of each story, averaged over the participant population. C for each recall was determined using the segmentation prompt given in this section. The horizontal axis gives R , the mean number of clauses from the narrative that were scored as being recalled. B) Plotting R (blue dots) and C (green crosses) vs M , as well as the linear fit (red dashed) to (R, M) . For the longest narrative, C deviates significantly from the linear fit to R , suggesting that a more conservative scoring of recalls would yield a sublinear scaling with M .

Prompt 7 Segmentation Prompt

Provide a word-for-word segmentation of the following narrative into linguistic clauses, numbered in order of appearance in the narrative:
 {narrative}

Completion 6 Prompt 7 using Argument 4

1. A boy was with the scoutmaster
2. and he was going up and down
3. and on the third time of going up the pier
4. he had cramps
5. and shouted 'help'
6. but no-one believed him,
7. even the scoutmaster looked on
8. by until this guy came out of nowhere
9. and grabbed the boy
10. and saved him.

B.3 Lack of Output Interference in Recognition Experiments

Similar to (Criss et al., 2011), we measure the strength of output interference in recognition experiments by tracking the d' measure as a function trial number. The discrimination measure used here is conventionally called simply d' and is given by the difference in z-score between the hit probability P_h (i.e. true positive rate) and the false alarm probability P_f (i.e. false positive rate). The z-transform is defined

$$z(p) = \sqrt{2} \operatorname{erf}^{-1}(2p - 1), \quad (3)$$

and the discrimination measure is

$$d' = z(P_h) - z(P_f). \quad (4)$$

We plot d' over the course of the recognition trials in Figure S3 separately for coherent stories and scrambled stories. Output interference is characterized by a d' which decreases with trial number, indicating a diminishing ability to discriminate new vs old items.

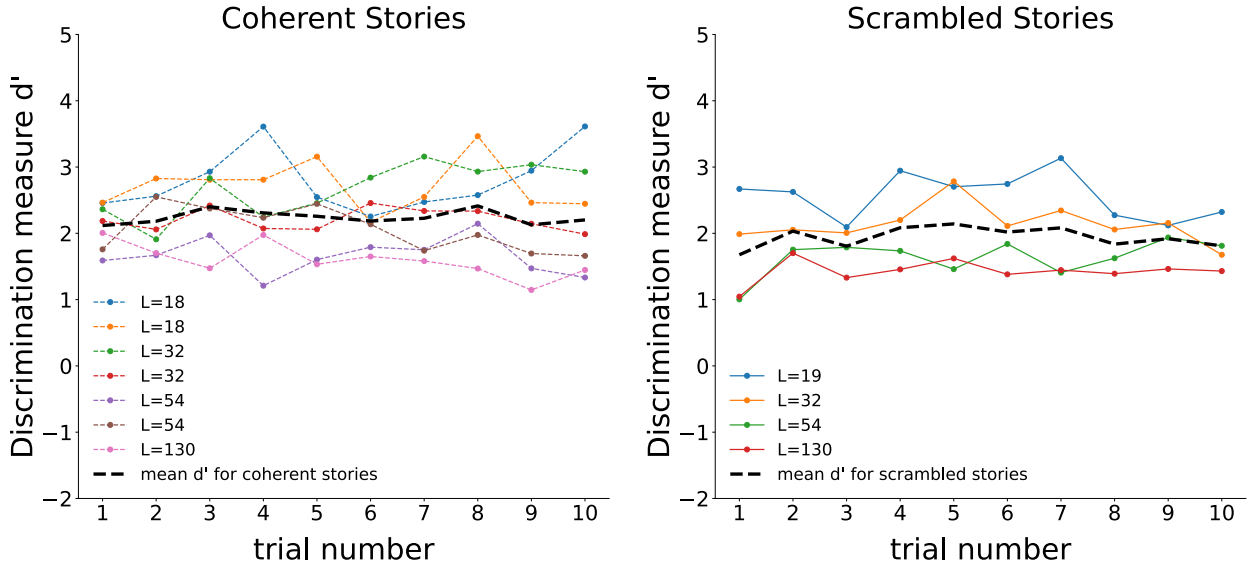


Figure S3: Discrimination measure shows good memory for true clauses which does not change significantly as the recognition trial progresses.

B.4 Recall Probability Curves and Distributions

Here we plot recall probability per clause as a function of the serial position of presentation. We refer to these below as recall probability curves. We compare recall probability curves for narratives to data obtained for random word

lists. Figure S4a shows the recall probability curve for coherent (blue), scrambled (red) stories, alongside the recall probability curve for free recall of a random word list (black dashed). Importantly, for the scrambled recall, the serial position corresponds to the position of *presentation*. Two main points can be gleaned from these figures. First, whereas for random lists there is a pronounced primacy effect, we see no such structure in both coherent and scrambled stories. In fact, highly recalled clauses appear at all positions in the story. Another observation is that the distribution of P_{rec} for clauses in a narrative covers a much broader range than P_{rec} for words in a list. This can be seen by looking at the cumulative distribution function of $F(p) = P(P_{rec} > p)$, which is the probability to find $P_{rec} > p$. As shown in Figure S5A, $F(p)$ drops to zero around $p \approx 0.6$ for random lists, while showing a more gradual decay for coherent narratives. In contrast, Figure S5B shows that recall is significantly impaired for scrambled stories, with $F(p)$ showing a more precipitous fall to zero similar to random word lists.

However, the recall probability curves for scrambled stories reveals a more subtle effect. While the overall distribution of recall probabilities may be similar to (if not worse than) random lists, there are still spikes in P_{rec} throughout the bulk of the narrative, suggesting that these texts are not processed as if they were a random list of clauses. In Figure S6, we replot the recall of only the scrambled stories from Figure S4 according to the serial position of the clause *in the coherent story*. Below this, we show the correlation between recall probability of a clause appearing in the coherent story versus the scrambled version. While low, the correlations are positive and statistically significant, suggesting that participants are able to identify memorable clauses even in the scrambled scenario, and use this in attempting to construct a coherent recall. This active process of selection might explain why average recall of clauses for scrambled narratives appears *worse* than recall of words in random lists (i.e. why the scrambled narrative data (orange triangles) lie below the square-root scaling (dashed gray) in Figure 2).

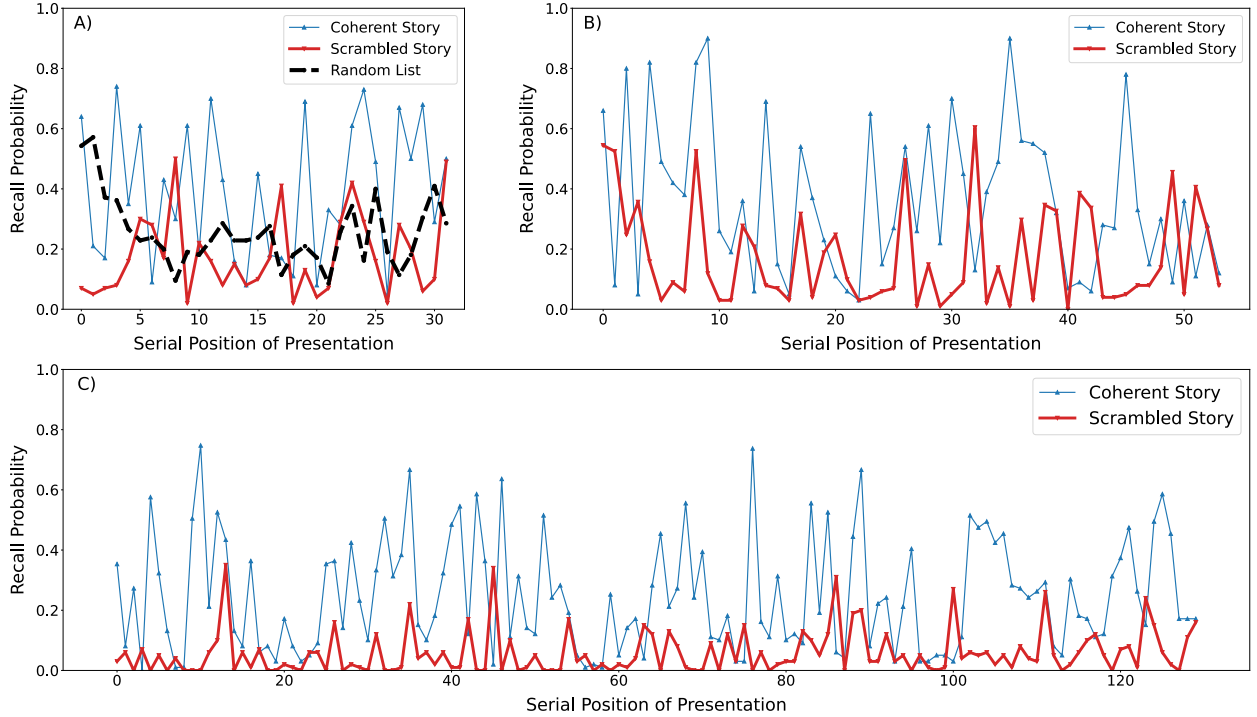


Figure S4: *Recall probability per item for narratives and random lists of words.* (A) shows the recall probability per clause as a function of the serial position of presentation for a coherent narrative (blue) and the scrambled version of this same narrative (red). As a comparison, we plot the recall probability per *word* as a function of serial position of presentation (black dashed). While the recall probability for the list shows a marked primacy effect, with a general suppression of recall in the bulk of the list, the recall of coherent narratives shows many spikes in P_{rec} throughout the narrative. And while the scrambled stories have overall much lower recall probabilities, they still show large peaks within the bulk of the presentation. (B) and (C) compare coherent and scrambled narratives of increasing length.

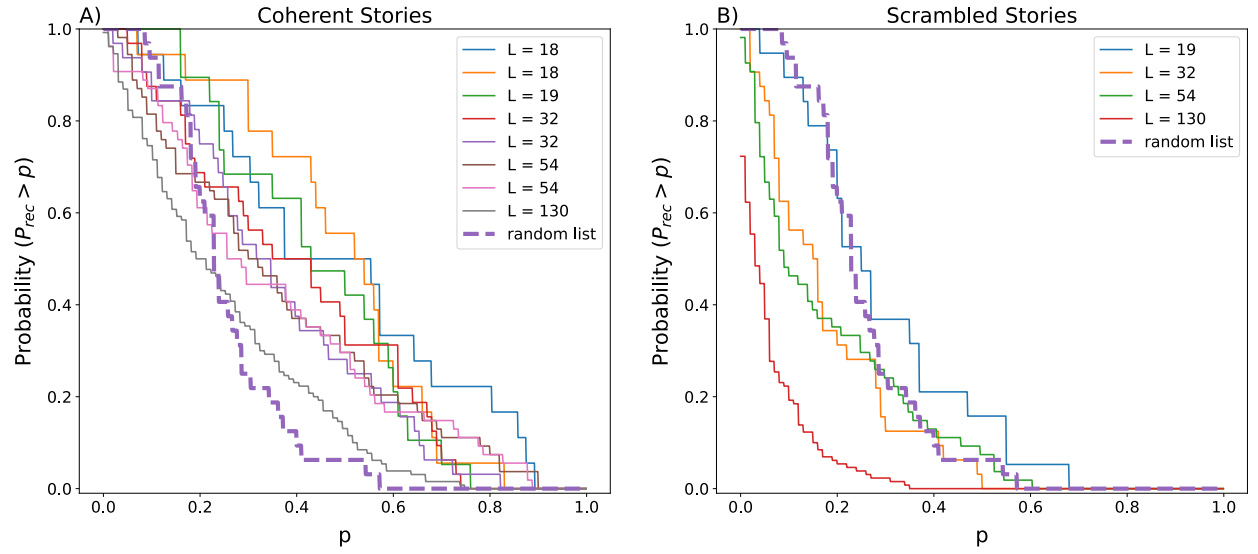


Figure S5: *Cumulative distribution of recall probabilities* (A) As compared to random lists (purple dashed), coherent narratives have a slower decline in the cumulative distribution function, indicating a larger number of clauses with a higher recall probability. (B) shows this behavior is reversed for scrambled stories, where the recall probability of clauses appears to uniformly drop, in some cases well below the corresponding distribution for recall of a random list.

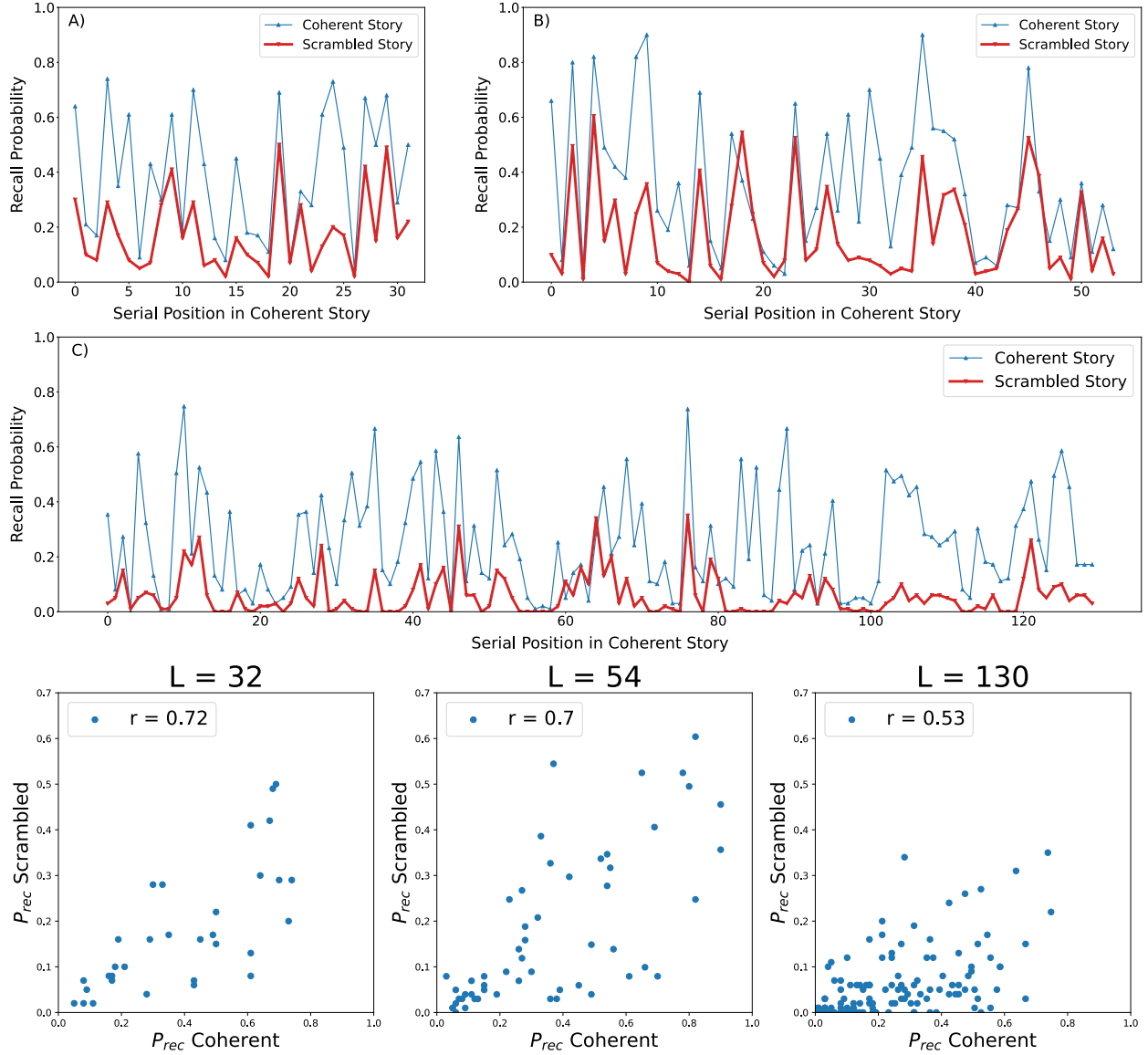


Figure S6: Descrambling of narratives in recall. (A-C) show recall probability as a function of the serial position in the coherent story. Note that for the scrambled P_{rec} (red curves), this is *not* the presentation order, but rather the original order of the clauses. Plotting the curves in this way, we observe the spikes in recall probability appear to coincide between coherent and scrambled recall. Below, we show correlations between P_{rec} for a given clause in the coherent and scrambled presentation.

B.5 Semantic Similarity and Recall

In the main text, we compared a measure of semantic similarity for each clause with its population averaged recall probability. The main text showed this for only a few narratives. Here, we display the data comparing this measure with recall probability for all narratives, as well as some summary figures. We also compare across a few different embedding models, some of which are publicly available.

The text-to-vector embedding models take a block of text as input (no longer than 4096 tokens in length), and produce a real vector output, which we refer to as the semantic vector embedding of the text. The OpenAI API provides access to three text embedding models, each of which produces normalized vector embedding of dimension 1536 (`text-embedding-3-small`, `text-embedding-ada-002`) and 3072 (`text-embedding-3-large`). We compare to an open-access embedding model from mixedbread.ai (`mxbai-embed-large-v1`) which produces unnormalized vector embeddings of dimension 1024.

The cosine similarity score we compute takes the vector embedding of a single clause $\mathbf{E}(c_i)$, and computes the cosine similarity with the vector embedding of the entire narrative $\mathbf{E}(C)$, where $C = c_1 c_2 \dots c_L$. The cosine similarity score is

$$S_i = \frac{\mathbf{E}(c_i) \cdot \mathbf{E}(C)}{\|\mathbf{E}(c_i)\| \|\mathbf{E}(C)\|}, \quad (5)$$

where the numerator shows the dot-product of the two vectors, and $\|\mathbf{x}\| = \sqrt{\sum_i x_i^2}$ denotes the norm of the vector $\mathbf{x}$.

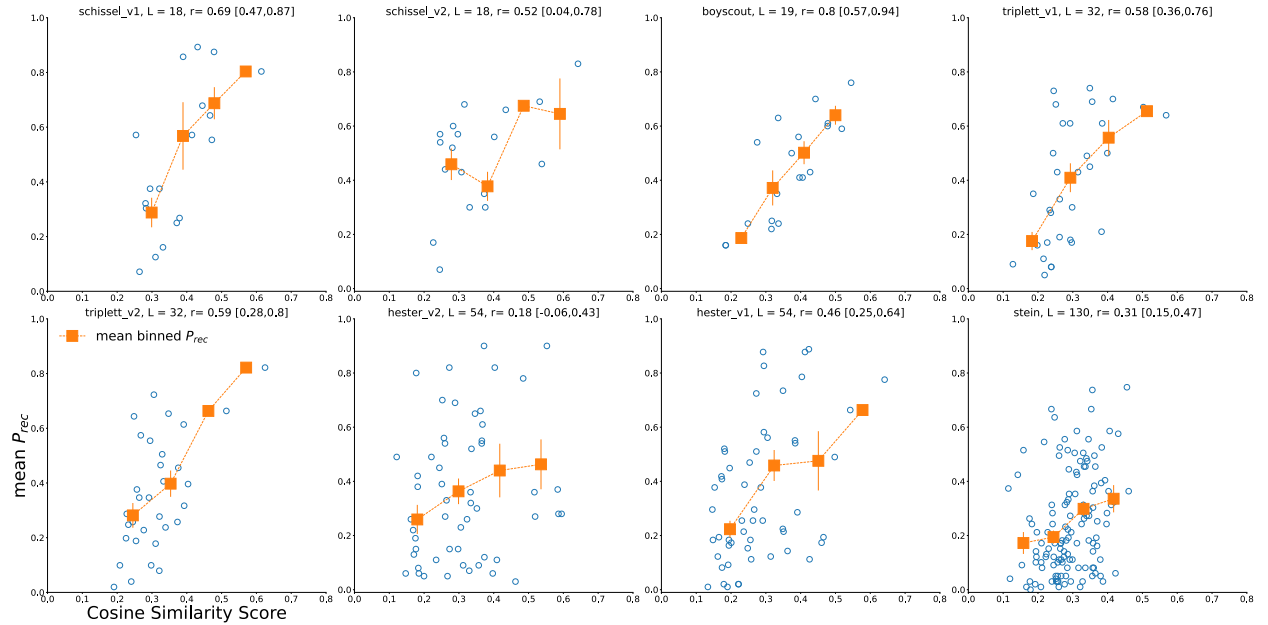


Figure S7: *Recall probability vs cosine similarity scores for all narratives* In each plot, we plot all L points ($P_{rec}(c_i), S_i$) (blue circles). To compute the orange square points, we bin the full range of similarity scores into four equal intervals, and compute the average P_{rec} in each interval. The horizontal coordinate for these points is taken to be the midpoint of the bin interval. The plots are labeled by the corresponding story name, along with the length L , and the correlation coefficient computed between P_{rec} and the similarity score. The 95% confidence interval for the correlation coefficient is indicated in brackets $[r_{low}, r_{high}]$, and is computed with bootstrap using 1000 samples.

Note that in Figure S7, each narrative has a different range of similarity scores. This likely has to do with the overall length of the narrative, though evidently length alone does not fully account for the variability. Furthermore, each narrative has a different range of recall probabilities. In order to compare across narratives, we compute z-scores of P_{rec} and S for a given narrative, and combine these z-scores across narratives in Figure S8. There is an overall strong correlation which survives when averaging across narratives. Using z-scores also allows a meaningful comparison between different models, and we show two OpenAI models, along with an open access model from mixedbread-AI. A comparison of raw similarity scores is shown in Figure S8D, where a model’s particular bias is evident. For instance, `text-embedding-ada-002` tends to produce similarity scores always above 0.7, indicating a strong clustering of

vector embeddings inside a cone. The Mixedbread-AI embeddings produce a much larger range of embedding directions, but also show a weaker correlation with OpenAI’s ‘3-small’ embeddings.

Finally, in Figure S9, we compute the correlation coefficient between recall probability and similarity score for each narrative individually (as in Figure 5), and compare across different LLM embedding models. All models show the same trend of decreasing correlation with increasing narrative length, discussed in the main text. Overall, the OpenAI models achieve higher correlation with lower p-value.

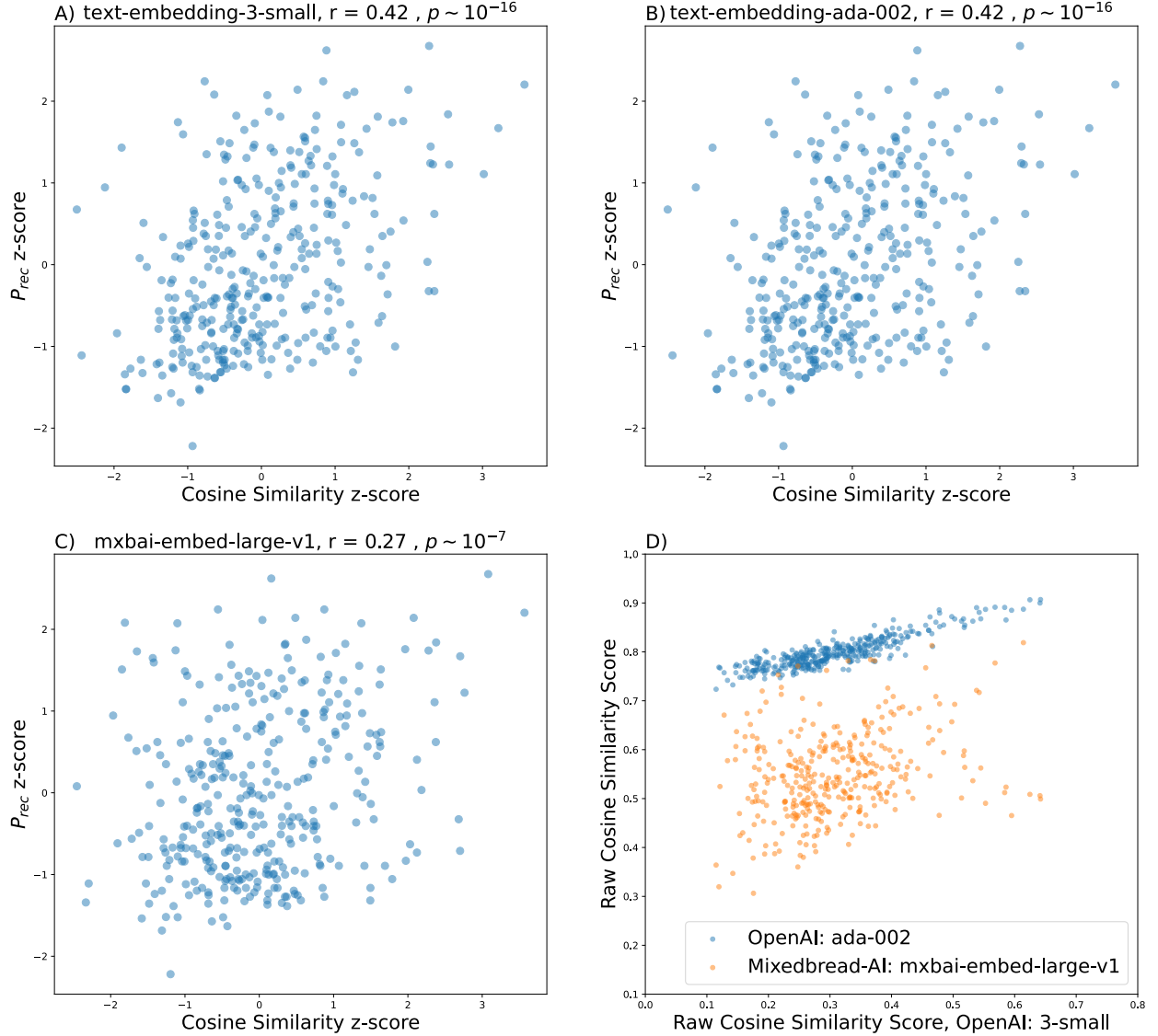


Figure S8: *Recall probability vs Similarity scores for all narratives* Here we compare different embedding models by calculating z-scores for recall and similarity values and pooling all of the data. We see very strong correlation between the OpenAI models, and a weaker but still good correlation with the open-sourced model. Not shown here is the correlation between `text-embedding-3-large` and P_{rec} , which is $r = 0.44$ at $p \sim 10^{-18}$, only marginally better than the smaller and more efficient models. p-value is computed using two-sided Wald test.

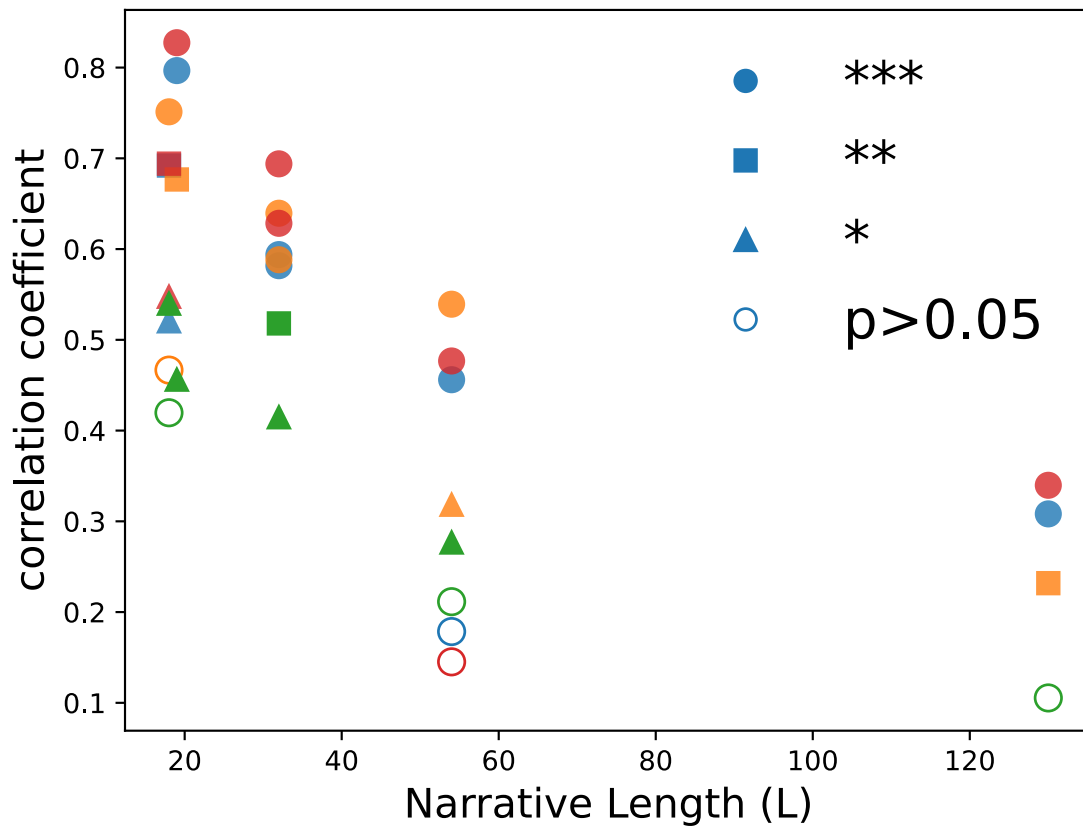


Figure S9: *Embedding model comparison* Here we show the correlation coefficient and statistical significance for each model on each narrative separately. The colors code for the embedding model: blue (OpenAI text-embedding-3-small), orange (OpenAI text-embedding-ada-002) and green (mixedbread.ai mx-bai-embed-large-v1). p-value is computed using two-sided Wald test.

B.6 Clauses with High Recall Probability

We noted that clauses recalled with higher probability tend to encapsulate a summarized version of the presented narrative. While not all nuances are retained, they collectively outline the principal storyline. Presented below are examples from stories of varying lengths. It's important to note that we present the recalled clauses exactly as they appeared in the original story and maintain the same sequence.

Example Story 11 Highly recalled clauses for narrative of length 18 clauses

Original Story

1. My best friend pushed me into the pool.
2. This happened during my cousin's wedding reception,
3. where everyone was dressed to impress.
4. And the reason it happened,
5. she spotted a bee hovering near my face,
6. - this was at a fancy hotel garden -
7. and she tried to save me from being stung.
8. And my aunt had just handed me a glass of champagne,
9. and I warned her to be careful.
10. 'Course friends, y'know, they don't always think things through.
11. So that's when she gave me a little shove,
12. and I tumbled into the water.
13. When I resurfaced, gasping for air,
14. she just started laughing,
15. and she apologized profusely.
16. And... my beautiful dress was ruined,
17. and naturally, the first thing to do was to get out and dry off,
18. and my cousin just says, "Just about a few inches more," she says, "and you'd have landed on the cake."

Highly recalled clauses

- | | |
|-------------------|---|
| $P_{rec} = 0.857$ | 1. My best friend pushed me into the pool. |
| $P_{rec} = 0.893$ | 2. This happened during my cousin's wedding reception, |
| $P_{rec} = 0.893$ | 5. she spotted a bee hovering near my face, |
| $P_{rec} = 0.661$ | 16. And... my beautiful dress was ruined, |
| $P_{rec} = 0.857$ | 18. and my cousin just says, "Just about a few inches more," she says, "and you'd have landed on the cake." |

Example Story 12 Highly recalled clauses for narrative of length 54 clauses

Original Story

1. Let me share a story from when I was twenty-two years old, just fresh out of college
2. I was staying in a small town about five miles from here
3. I had just started my first job at a local bookstore
4. It was a quiet afternoon, and I was alone in the store
5. My grandpa, who had been bedridden for almost a year, was at home with my sister
6. She had to care for him daily, as he couldn't move around on his own
7. There were only two of us siblings, my sister and me

8. My parents had gone out of town for the weekend
9. As I was shelving books, I suddenly heard a voice inside my head
10. It said, "If you go to the old church down the street and light a candle for your grandpa, he will get better
11. I looked around the bookstore, but no one was there
12. I continued with my work, trying to ignore the voice
13. It spoke to me again, repeating the same message
14. I looked around once more, but still found no one
15. I decided to take a break and walk to the church
16. I was only a short distance away, maybe a quarter-mile
17. As I approached the church, I hesitated for a moment
18. I was not a particularly religious person
19. I remembered my grandma always saying that prayers from a sincere heart could bring miracles
20. On the other hand, she also said that prayers without faith were useless
21. I admitted to myself that I was far from being a saint
22. I had my share of mistakes and misdeeds
23. I thought about my grandpa and how much he meant to our family
24. Finally, I went inside the church and found a candle
25. I knelt down in front of the altar
26. I said, "God, I don't know how to pray properly
27. But I am asking you with all my heart to heal my grandpa
28. Please help him regain his strength and guide us in taking care of him
29. After saying my prayer, I returned to the bookstore
30. I resumed my work, shelving books
31. Suddenly, I felt a strong urge to go home
32. It was as if something was pulling me towards the house
33. I didn't hear any voice this time
34. It was just an overwhelming feeling that I needed to go
35. I locked up the store and rushed home
36. To my surprise, I found my grandpa sitting in the living room, talking to my sister
37. He hadn't been able to leave his bed for months without assistance
38. My sister told me that he had suddenly felt a surge of energy and was able to stand up and walk
39. That day marked the beginning of his recovery
40. He gradually regained his strength and independence
41. It was a turning point in our lives
42. We became more grateful for the time we had together
43. I often thought back to that day in the church
44. I wondered if my prayer had anything to do with my grandpa's recovery
45. I couldn't be sure, but it definitely changed my perspective on faith and the power of prayer
46. My grandpa lived for many more years after that incident
47. He celebrated his ninetieth birthday surrounded by family and friends
48. We cherished every moment we had with him
49. That experience happened almost twenty years ago

- 50. It remains a vivid memory in my mind
- 51. My grandpa passed away a few years later, but his spirit lives on in our hearts
- 52. Sometimes, when I'm feeling lost or uncertain, I think back to that day
- 53. I remember the power of faith and the miracles it can bring
- 54. And I am reminded that we are never truly alone, even when it seems like we are

Highly recalled clauses

- | | |
|-------------------|---|
| $P_{rec} = 0.710$ | 1. Let me share a story from when I was twenty-two years old, just fresh out of college. |
| $P_{rec} = 0.830$ | 3. I had just started my first job at a local bookstore. |
| $P_{rec} = 0.850$ | 5. My grandpa, who had been bedridden for almost a year, was at home with my sister. |
| $P_{rec} = 0.860$ | 9. As I was shelving books, I suddenly heard a voice inside my head. |
| $P_{rec} = 0.920$ | 10. It said, "If you go to the old church down the street and light a candle for your grandpa, he will get better." |
| $P_{rec} = 0.720$ | 15. I decided to take a break and walk to the church. |
| $P_{rec} = 0.700$ | 31. Suddenly, I felt a strong urge to go home. |
| $P_{rec} = 0.900$ | 36. To my surprise, I found my grandpa sitting in the living room, talking to my sister. |
| $P_{rec} = 0.780$ | 46. My grandpa lived for many more years after that incident. |