Familiar Story Structures Possess an Evolutionary Edge in Memory

Abla Alaoui-Soce[1] and Diana Tamir[1]

[1] Department of Psychology, Princeton University

Author Note

Abla Alaoui-Soce https://orcid.org/0009-0004-5552-4773

Data, analyses, and supplemental materials are openly available on the Open Science Framework (https://osf.io/hu7ke/). We have no conflicts of interest to disclose. This paper has not been peer-reviewed.

Correspondence concerning this article should be addressed to Abla Alaoui-Soce, Peretsman Scully Hall, Department of Psychology, Princeton University, 08544, New Jersey, U.S. Email: ablaa@princeton.edu

Abstract

Human beings demonstrate a universal impulse to share and consume stories. Over generations of transmission, within and across cultures, stories have evolved, with some story structures surviving better than others. Here, we investigate the factors theorized to guide this selection. Across three studies, our project tracks the evolution of stories in real time as participants tell and retell a story for 5 days in a row. Each study varies the structure of the initial story to manipulate its familiarity and structure. We predict that stories with familiar structures will survive better across retellings. Study 1 tracks the evolution of a story with no recognizable structure using novel story structure metrics. Study 2 compares the evolution of a story with structure and without structure. Study 3 compares the evolution of a familiarly structured story with an unfamiliarly structured story. Results show that all stories become more structurally stable across retellings, with stories moving in a consistent direction (i.e., towards a consistent final form). Stories with more familiar structures show greater evolutionary stability: They are modified less than both unstructured (Study 2) and unfamiliar stories (Study 3), and are either consistently more structurally stable across retellings (Study 2) or start out more stable in earlier retellings (Study 3). By the last generation of recall, people show greater agreement in their recall of more familiarly structured stories than unfamiliarly structured stories (Study 3). This suggests that familiar stories with structural integrity have an evolutionary edge over unfamiliar structures, surviving better in people's memories.

Keywords: stories, schemas, memory, natural language processing, serial reproduction

Public Significance Statement. People spend a great deal of time engaging with stories. Over a lifetime of experience, people learn how stories are supposed to go. In these studies, we ask how this influences memory for new stories: We find that, over several days, new stories that follow a familiar structure (like a Cinderella-type tale) both survive better within individuals' memories and are recalled in more similar forms between people.

Word Count: 12,188

Familiar Story Structures Possess an Evolutionary Edge in Memory

Long before Disney took over the fairytale market, stories like Cinderella, Snow White, and Sleeping Beauty had been part of the Western canon and beyond for generations. One of the earliest known variants of Cinderella, “Rhodopis”, for instance, has been traced back to Greece in the first century B.C.E. (De la Rochere, 2016). Researchers in folklore and mythology have amassed catalogs of folktales in different societies across the world (Thompson, 1955; Uther, 2004). Their work reveals shared traits in the stories within and across societal boundaries, as stories are transmitted ‘vertically’ (through the generations) and ‘horizontally’ (across cultures), respectively (Thompson, 1955; Uther, 2004; da Silva & Tehrani, 2016; Bortolini et al., 2017). Incredibly, some stories, like the “Devil and the Smith,” can be traced back as far as the Bronze Age, thousands of years ago (da Silva & Tehrani, 2016). However, not all stories are successfully transmitted across generations or cultures, and not all stories that are transmitted retain their structure. Stories have undergone an evolution, with some disappearing entirely, others changing enough to become almost unrecognizable, and others still surviving in more recognizable forms. What determines how stories will evolve?

Stories do not evolve of their own accord. They evolve through people. To borrow Terry Pratchett’s metaphor, they exist as “parasitic life form[s]” (Pratchett, 1991), occupying the minds of people. This is most evident when we consider folktales: Typically shared in oral form, they exist within and across the minds of people who pass them along with varying levels of accuracy. Even the transmission of written text or visual media depends on people—as generators, consumers and sharers of those stories. Story evolution relies on the minds that carry them—on human experience with stories. The evolution of story structure has taken place over the course of human history. Yet, using novel story structure measures, we can track this evolution in real time, as stories change within the span of days. Why do some stories survive better in our minds than others?

Story Structure in the Mind

The human mind is not a blank slate into which a story can be inscribed. Instead, the human mind builds structures over time that can both facilitate and shape the encoding and recall of new inputs (Mandler & Johnson, 1977; Alba & Haser, 1983). In 1932, Bartlett conducted a series of seminal studies that introduced the theory of schema and the notion of memory as a reconstructive process. Schemas are cognitive structures that organize knowledge and experience of the world around us. Schemas can be specific to a social and cultural context. In Bartlett's (1932) narrative serial reproduction studies, participants were asked to reproduce narratives over and over, either on their own or in a chain (as in a telephone game). The participants in some of these studies, for example, were asked to recall a culturally unfamiliar story, the Native American story 'War of the Ghosts'. In their retellings, participants modified stories to better fit with their existing schemas (Bartlett,1932). When asked to reproduce the story, they added causal connections where they felt they were 'missing' and deleted elements that did not fit into their pre-existing causal schemas. Researchers have since replicated (Bergman & Roediger, 1999) and built upon Bartlett's work, finding that people recall the features and orders of events in a story as more similar to the standard experience—their 'scripts'—than what was actually described (Bower, Black & Turner, 1979). And, people from different cultures remember stories about people performing common activities as more similar to their respective cultural schemas (Harris et al, 1988).

The extent to which a story fits into a person's existing schemas matters in terms of how well people recall it. In particular, a story can fit into people's schemas in two ways: (1)

Structure: Simply put, does the story's sequence of events make sense? A story with structure will fit better into existing schemas than a story without structure. (2) Familiarity: How often has this sequence of events appeared in stories? A familiar story will fit better into existing schemas than an unfamiliar story.

The existence of familiar schemas depends on people's experience with stories. Repeated exposure to certain story structures shapes our ability to retain and recall new stories. That is, people show better comprehension of and retention of familiar stories (Kintsch & Greene, 1978; Kintsch & van Dick, 1975). The stories that have made their way into the cultural canons are more familiar, and thus may further contribute to how people process any new story. Story experience at earlier developmental stages may be crucial (Piaget, 1926). The interaction between people's schemas and stories is understood as a mutually reinforcing feedback loop: Commonly shared story types contribute to the development of internalized story schemas; this leads to new schematic stories being better retained; which, in turn, leads to schematic stories being more commonly shared.

In the current studies, we likewise expect that stories that fit better with people's existing schemas will be better retained. Specifically, we expect that stories modeled off of forms that have evolved to be commonplace would be more structurally stable across retellings in people's minds. Their frequent appearances across different individual stories and mediums would make them highly familiar to people within a cultural context. You would be hard pressed to find someone, for instance, who has not read, heard, or watched a Cinderella-type story. This repeated exposure means that people know how Cinderella-type stories are supposed to go–we have developed a schema for it. Because of this, such stories should be more stable in people's minds. Stories modeled off of unfamiliar forms, on the other hand, may not find resonance in existing schemas. The same could be said of stories with no clear structure, or stories without a structure that makes sense. Instead, people may modify these unstructured or unfamiliar stories to better fit existing schemas. Stories with unstructured or unfamiliar forms, then, should be less structurally stable in memory. To test this possibility, we will observe story evolution at play, tracking how these stories are modified before they become structurally stable.

Current Research

How does a story's structure and familiarity determine its evolution in a person's mind? To capture story evolution within the lab, we employ Bartlett's (1932) narrative reproduction design: In 3 studies, people read a story and then recall it over and over. These studies offer an advance over prior work in that we can track recall using a novel measure that uses natural language processing to automatically derive the changes in story structure across a person's retellings. This structural similarity measure captures how the representation of events changes across generations of recall, abstracting away from the surface level details of the story. Specifically, it calculates changes in (1) which events are included in the story and (2) the ordering of these events. In addition to tracking structural similarity across a person's retellings, we adapt this measure to compare stories across people. That is, we can compare the structure of stories generated by different participants.

With these measures, we can test the following hypotheses as we follow the evolution of stories across 5 days of recall: First, we expect that stories with structure are more evolutionarily fit than stories without structure. Second, we expect that stories with familiar structures are more evolutionarily fit than stories with unfamiliar structures. That is, structured and familiarly structured stories should survive better in people's minds than unstructured or unfamiliarly structured stories. Survival will be measured in two ways: (i) Stability in a person's recall, or how similar a story is across generations of a person's recall, and (ii) Agreement across people's stories, or how similar a story is across people's recall. Familiar stories are expected to be more stable in memory, undergoing little modification as they are told and retold. This would translate to greater story agreement across people's retellings. In contrast, unfamiliar stories or stories without structure should undergo greater modification before reaching stability. These modifications should lead to greater divergence across people's retellings.

Across 3 studies, we modify the familiarity and structure of stories to test how these factors affect the survival, modification, and evolution of the stories over time.

We employ the structural similarity measure to track the evolution of novel stories in three studies: Study 1 first validates the new structural similarity measure by testing how it tracks the evolution of a novel Nonsense story over time. In Study 2, we test the impact of structure on the evolution of a story by comparing the evolution of a novel story that follows the familiar Cinderella structure (a Structured story) with the evolution of a scrambled version of the novel tale (an Unstructured story). In Study 3, we test the impact of familiarity on the evolution of a story by comparing the evolution of that same familiar story with that of an equally structured story, but unfamiliar story. Together, these studies establish why some stories survive better in our minds than others. Structure and familiarity lend stories their evolutionary edge over unstructured or unfamiliarly structured stories.

Study 1: Evolution of a Nonsense Story

Study 1 has two aims. The first is methodological: We validate the new measure of structural similarity as we track how a story evolves across retellings. The second aim is to investigate the evolution of a story with no distinguishable structure. The story used in Study 1 was designed to read like fluent nonsense. It depicts a mishmash of randomly succeeding events, so as to not align with people's established story schemas. Our goal was to study how the nonsense story would change across retellings given that participants should not be able to rely on strong priors (or schemas) to facilitate encoding and recall. We therefore expected to see shifts in the representation of events across retellings that we could track with our measures. We predicted that the story would be unstable at first, changing more across retellings as it is modified to become more stable. In the absence of a clear structure in the initial story to scaffold recall, the story has to be changed in order to better fit pre-existing schemas.

Methods

Participants

Participants (N =199) were recruited using Mechanical Turk (www.cloudresearch.com) to complete a 5 day study. We set a target sample size of 50, after attrition. Of the 199 participants who completed Day 1, 71 participants completed all 5 days of the task. Participants (N = 18) were excluded based on the following a priori exclusion criteria: (i) not meeting task requirements, namely by typing different words and not writing the story at all; (ii) not writing a sufficient amount of text. This was formalized by excluding participants with word counts below 1 standard deviation from the mean (M=126.50, SD=48.53). These exclusions left us with a final usable sample size of 55. In preparation for data analyses, all stories were cleaned in the following ways: (1) Spelling errors and typos were corrected. (2) Meta-commentary was removed. This included phrases like "if I remember correctly", "I recall", "I believe", "I think", as well as longer sentences about the difficulty of the task.

All participants were U.S. residents fluent in English. Participants received a total of $3.50 for completing all days of the task (10 cents on Day 1, 20 cents on Day 2, 30 cents on Day 3, 40 cents on Day 4, and an additional $2.50 for completing all 5 days). Participants in this and all subsequent studies provided informed consent in accordance with the Princeton University Institutional Review Board. Participant data (pre-exclusions and cleaning, as well as post-exclusions and cleaning) and analysis code for this and all subsequent studies can be accessed here (https://osf.io/hu7ke/).

Procedure

Participants completed the task over 5 days. The task was set up on Qualtrics (www.qualtrics.com). On the first day, participants listened to a nonsense story lasting 2 minutes and 22 seconds, 378 words long. To ensure that participants had no memory aids, they were asked to not take any notes and to not replay the story. The nonsense story follows Sophia, a psychic, as she goes through a series of nonsensical adventures (See *Supplement*). After confirming that they had listened to the story all the way through, participants were asked to recall the story as best they could. Participants were encouraged to write as much as possible.

On days 2-5 of the study, participants were asked to recall the story they had listened to on the first day. Participants were invited back to complete the task on days 2-5 at the same time each day. They were allowed a maximum of 12 hours to complete the task. This ensured that the time interval between each re-telling could not be less than 12 hours nor exceed 36 hours. By Day 5, successful participants produced 5 "generations" of the story they listened to on the first day, 1 generation for each day.

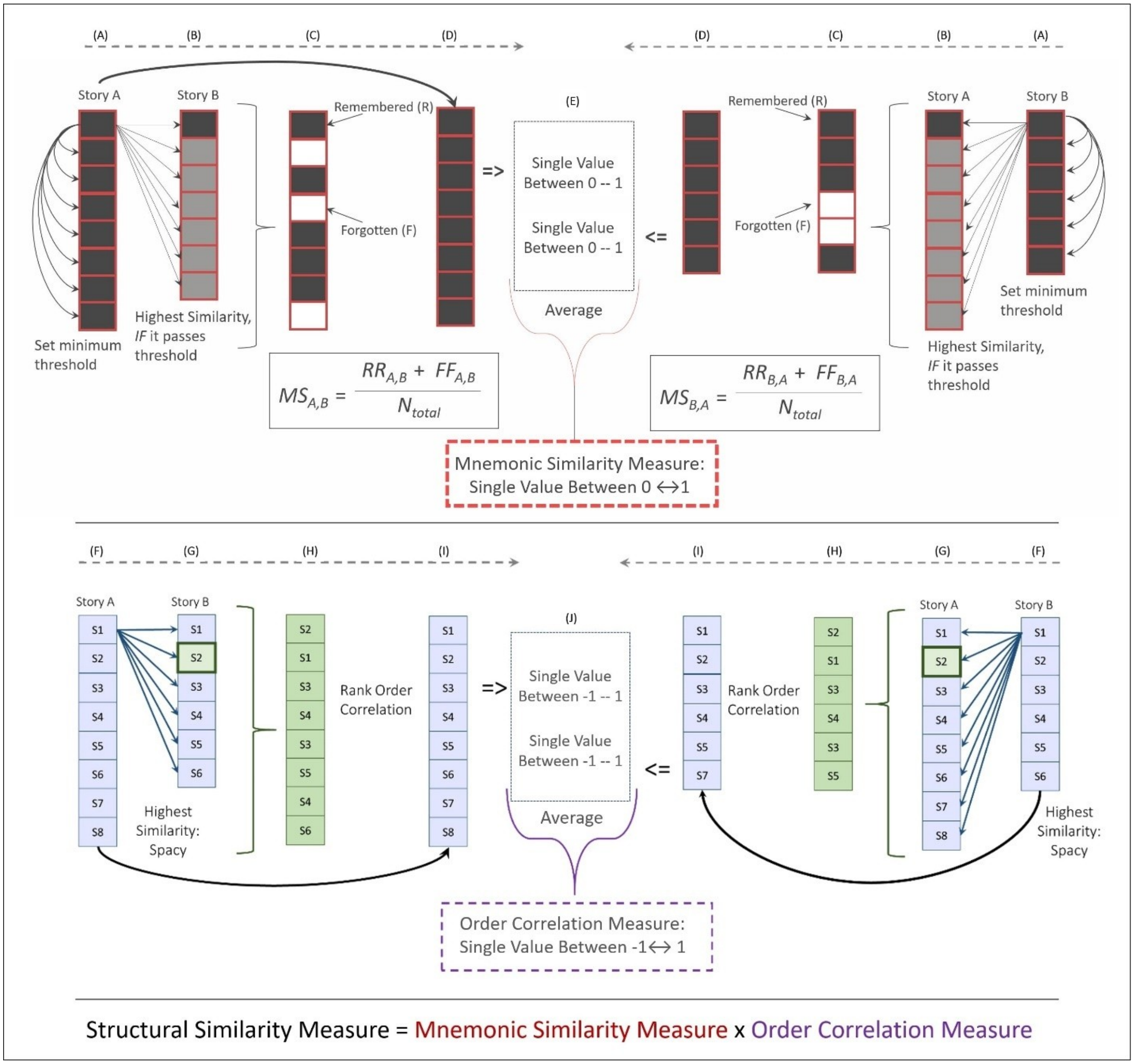
Analyses

Structural Similarity Measure. We developed a structural similarity measure that aims to capture how a story's structure changes between retellings. This measure captures both a story's content and a story's sequence. To capture how event content is changing, we track items remembered and forgotten in common across generations of retelling the story. This is referred to as the ‘mnemonic similarity’ (Coman et al., 2016). To capture how sequence is changing, we track changes to the order in which these events appear, using a novel event order correlation measure. Both the content and sequence aspects of the structural similarity measure use Spacy’s semantic similarity calculator as a basis for deriving similarity (Honnibal & Montani, 2017). Spacy determines similarity by comparing word vectors, multi-dimensional representations of a word’s meaning.

We capture changes in content between two stories (e.g., Story A and Story B) as follows (*Figure 1, A-E)*: First, we break each story into its component sentences, a proxy for events. Then, for each sentence in Story A, we use Spacy’s measure of semantic similarity to compare it to every other sentence in Story A. The highest similarity value will henceforth count as the minimum similarity required a sentence in Story B must exceed in order to qualify as the same event (*Figure 1A*). For instance, if Sentence 1 in Story A is most similar to Sentence 4 in Story, with a Spacy similarity value of 0.70, then 0.70 is the minimum similarity a sentence in Story B must pass to be marked as the same event as Sentence 1. This first step allows us to establish a stand-in for event identity for each sentence. Next, we compare Sentence 1 in Story A to each sentence in Story B. If the sentence in Story B with the highest similarity value passes the minimum similarity threshold set up in the first step, then we count it as the same event (*Figure 1B*). We can therefore mark that event (the one expressed in Sentence 1 of Story A) as *Remembered* in Story B. If not, we mark that event as *Forgotten* in Story B. We then go through the same steps for all the other sentences in Story A, therefore deriving a list of events (as defined in Story A) that are remembered and forgotten in Story B (*Figure 1C*). We then use Coman et al.’s (2016) mnemonic similarity calculation (number of events remembered in common + number of events forgotten in common, divided by total number of events) to derive the mnemonic similarity between Story A and Story B (*Figure 1D*).

Figure 1

Structural Similarity Measure



Note. Figure 1 illustrates the process to calculate the structural similarity measure between two stories, Story A and Story B. The top panel follows the steps to calculate the mnemonic similarity component: The left-hand side follows the steps (A-D) for Story A as the reference story and right-hand side (mirrored) follows the steps (A-D) for Story B as the reference story. The mnemonic similarity value is an average of the values gotten from both (E). The bottom panel follows the steps to calculate the order correlation component: The left-hand side follows the steps (F-I) for Story A as the reference story and right-hand side (mirrored) follows the steps (F-I) for Story B as the reference story. The order correlation value is an average of the values gotten from both (J). The final structural similarity measure value is obtained by multiplying the mnemonic similarity and order correlation measures.

This process is then repeated in full, now starting with Story B as the reference, and comparing each of its component sentences to the Story A sentences. The final mnemonic similarity value is calculated as the average of the two resulting mnemonic similarities (*Figure 1E*). This ensures symmetry in the way similarity is measured: Story A is just as similar to Story B as Story B is to Story A.

We capture changes in sequence–or event order–between two stories *(e.g., Story A and Story B)* as follows (*Figure 1, F-J*): We first divide each generated story into sentences (*Figure 1F*). Then, for each sentence in Story A, we use Spacy's similarity measure to find the sentence in Story B that is most similar (*Figure 1G*). Matches are relative; there is no lower threshold in similarity for assigning a sentence match, so long as the match is higher than all the other sentence similarities. This process is repeated for all the sentences in Story A (*Figure 1H*). Several sentences in Story A may be assigned the same sentence match in Story B. For example, if Story A includes the sentences, "*Sara wants to study physics."* and "*She worked hard on a physics experiment for a science fair.*", then both sentences might be best matched by the following single sentence in Story B: "*Sara dreamed of studying physics, so she worked hard on a project for a science fair*." Indeed, semantic similarity values according to Spacy (on a scale of 0 to 1) are high for both sentence comparisons–measured at 0.87 and 0.94, respectively. Once all sentences in Story A have a match, we calculate the Spearman's rank order correlation between sentence order in Story A and the order of its matches in Story B (*Figure 1I*). As for the mnemonic similarity value, this process is then repeated with Story B as the reference. The final order correlation value is calculated as the average of the two resulting correlation coefficients. (*Figure 1J*).

We then combine the order correlation measure with the mnemonic similarity measure (Mnemonic Similarity x Order Correlation), which together form the Structural Similarity Measure. This measure of structure similarity is used for the comparisons across generations, the comparisons of each generation to the last generation, and the comparisons of each generation to the initial story. This same analysis technique was used for all studies reported here.

Validating the Structural Similarity Measure. Before we could make use of the structural similarity measure, we first validated it as both useful and informative by testing the extent to which it agreed with human judgements of similarity.

We used the stories generated by participants in Study 1 to validate the structural similarity measure. We asked a new set of MTurk participants (*N*=212) to each rate the similarity of 5 pairs of stories on a scale of (1) Extremely Different to (7) Extremely Similar). To ensure that the validation captured the full range of similarity that would be measured in each study, we selected the pairs of stories in the following way: For each participant, one pair included the initial nonsense story, one pair was of stories from two different participants at the same generation of recall, one pair was of stories from the same participant from different generations of recall, and two pairs were of stories from two different participants at two different generations of recall. Each generation of recall would also appear at least once for each participant. The same pairs of stories appeared at least 3 times across all participants to ensure multiple ratings on the same pair. In total, 153 participants rated all the stories they were given, and we were able to collect at least 3 similarity ratings on 94 different pairs of stories. These repeated similarity ratings were averaged to derive a single similarity rating measure per pair of stories. We then tested these averaged human similarity ratings against: (1) The mnemonic similarity measure on its own, ($r(92)=0.57$, $p<0.001$), (2) the order correlation measure on its own, ($r(92)=0.53$, $p<0.001$), and (3) the structural similarity measure which combines both, ($r(92)=0.67$, $p<0.001$). Not only do we find significant correlations for all three, we confirm that the combined structural similarity measure does indeed align more closely with human ratings of similarity, ($r(92)=0.67$, $p<0.001$) than both the mnemonic similarity measure ($z=2.11$, $p<0.001$) and rank order correlation measure ($z=3.51$, $p<0.001$) (Hittner, May & Silver, 2003; Diedenhofen & Musch, 2015).

In addition to testing the structural similarity measure against human ratings of similarity, we wanted to make sure that the measure is not simply picking up on differences in length between stories. Or at least, not more so than would a human rater. To check for this, we tested the structural similarity measure against differences in word count and differences in sentence count. We also compared these correlations against human ratings' correlation with differences in word and sentence count. Controlling (separately) for the effects of word count and sentence count differences, we find similar correlations between our structural similarity measure and human ratings of similarity, ($r(92)=0.67$, $p<0.001$) and ($r(92)=0.67$, $p<0.001$), respectively (Kim, 2015). In addition, we find no significant correlations between the structural similarity measure and either the word count ($r(92)= -0.08$, $p=0.46$ (ns) or sentence count difference ($r(92)= -0.11$, $p=0.28$ (ns)). Nor do we find any significant correlations between human ratings and either the word count ($r(92)= -0.08$, $p=0.42$ (ns)) or sentence count difference ($r(92)= -0.05$, $p=0.63$(ns)). In other words, story similarity, as calculated by the structural similarity measure and as judged by human raters, is not determined by differences in story length.

Measuring the Evolution of Structural Similarity. Having validated the structural similarity measure, we could now employ it in different ways to capture story evolution. To track how the structure of the story evolved as it is recalled, we designed the analyses to detect three components: Stabilization, Consistency and Modification.

Stabilization is defined as an increase in similarity over time, and is measured as change in pairwise similarity from Generation 1 through 5. For example, we would expect Generation 2 and Generation 3 to be more similar to one another than Generation 1 and Generation 2. We expect that recalls stabilize over time as each generation moves towards a consistent form.

Consistency is defined as change in a unique direction, and is measured as change in similarity to the Generation 5, from Days 1-4. That is, we use the final generation, the story produced on the last day of the task (Day 5) as the standard against which the earlier stories are compared. We expect to see consistent stabilization, such that generations become more similar to the generation produced on the final day.

Modification is defined as the degree of change undergone by the initial story given to participants. Thus, we track the similarity of the initial story to each generation (Days 1 to 5). We expect that stories with structures that are more culturally familiar will better fit into people's schemas and therefore undergo less modification across generations.

Results

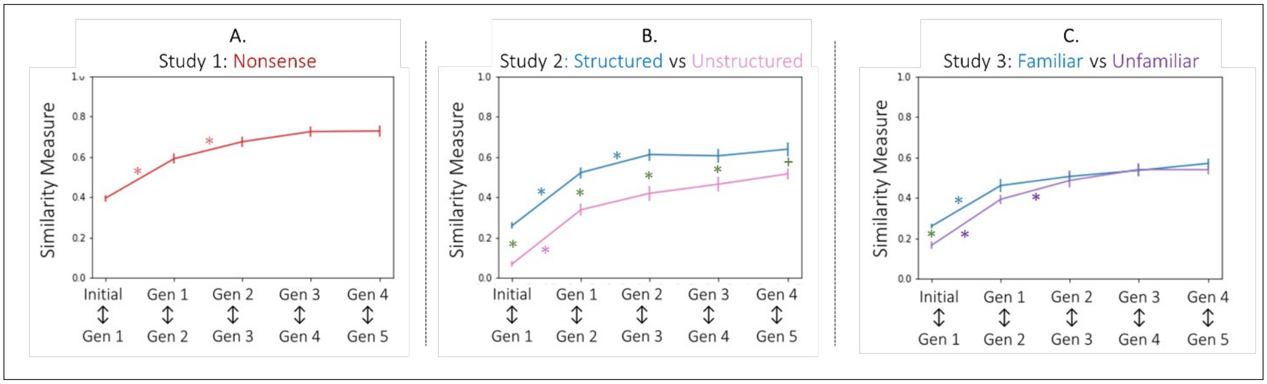
This study tests how story structure anchors recall. Study 1 asks how a story evolves in the *absence* of structure, such that people would have no obvious pre-existing schema to guide its encoding and recall. We do this by looking at how a nonsense story changes across retellings, with large changes reflecting instability, and small changes reflecting greater stability. We expected that, while the story may start more unstable, changing more across earlier generations, it would become more stable in subsequent generations. In addition, we expected to find a uniform trend such that retellings would become more similar in a consistent way, tracked by similarity to the last generation produced on Day 5.

Stabilization: Similarity Across Generations

We first measured stability, or how similar each generation is to each subsequent generation of recall (*Figure 2A*). We ask whether the story becomes more stable over retellings. We find a main effect of generation on stability, ($F(3.26,176.04)=62.681$, $p<0.001$, $\eta^2_G=0.324$), such that stories in earlier generations increased in structural similarity before settling into high similarity values. Similarity between the Initial story and Generation 1 ($M=0.39$, $SD=0.12$) is significantly lower than the similarities between all subsequent generations [pairwise: Gen1 and Gen2, ($M=0.59$, $SD=0.19$, $t(54)= -7.87$, $p<0.001$, $d = -1.05$, 95% CI [-0.27, -0.12]; Gen2 and Gen3, $M=0.67$, $SD=0.18$, $t(54)= -11.53$, $p<0.001$, $d = -1.53$, 95% CI [-0.35,-0.21]; Gen3 and Gen4, $M=0.72$, $SD=0.19$, $t(54)= -14.77$, p<0.001, $d = -1.96$, 95% CI [-0.39,-0.26]; Gen4 and Gen5, $M=0.73$, $SD=0.20$, $t(54)= -16.28$, $p<0.001$, $d = -2.16$ 95% CI [-0.39,-0.27]] (Singmann et al., 2015). (All pairwise t-tests, here and moving forward, are Bonferroni corrected for multiple comparisons. All Cohen's d effect sizes are hedges corrected.) We next see that similarity between Generation 1 and Generation 2 is lower than the similarity between all subsequent generations [pairwise: Gen2 and Gen3, $t(54)= -3.82$, $p=0.003$, $d = -0.51$, 95% CI [-0.15, -0.02];; Gen3 and Gen4, $t(54)= -4.83$, p<0.001, $d = -0.64$, 95% CI [-0.22, -0.05]; Gen4 and Gen5, $t(54)= -4.69$, $p<0.001$, $d = -0.62$, 95% CI [-0.22, -0.05];]. However, as early as Generation 2, the story settles into high similarity across generations, with similarity between Generation 2 and Generation 3 not differing from similarity between any subsequent generation. At that point, structural similarity levels remain around 0.7, suggesting that the stories are highly stable across retellings. This could reflect that only minor changes are happening across generations, as recollection is not perfect across days.

Figure 2

Stabilization: Similarity Across Generations



Note. Results for Stability, or the Similarity Across Generations, Studies 1-3. As stories are told across generations, they become more and more stable. In *B-C*, the green asterisks mark the significant differences between the two Story conditions in each study. The green crosses mark marginal differences. The color matched asterisks (e.g., red asterisks for Study 1) represent significant differences between the structural similarity comparisons for adjacent generations. Error bars represent standard error.

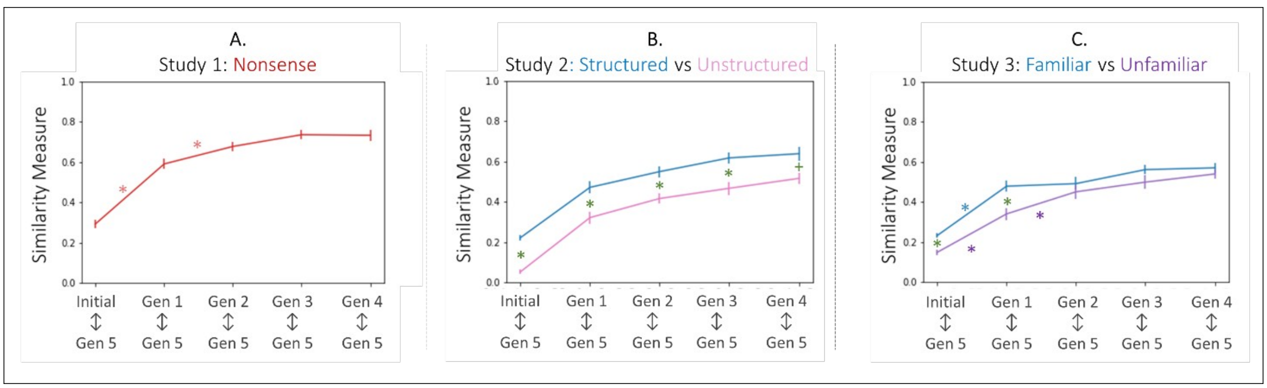
Consistency: Similarity to Last Generation

Here, we ask whether the story is changing in a consistent direction by tracking similarity to the last generation of recall on day 5 (*Figure 3A*). The stories' similarity to the last generation increases across earlier retellings before settling into higher similarity to the last generation (main effect of generation: $F(3.51,189.55)=114.028$, $p<0.001$, $\eta^2_G=0.474$). The structural similarity between the Initial story and Generation 5 (M=0.29, SD=0.14) is significantly lower than all the similarities between each subsequent generation and Generation 5 [pairwise: Gen1 and Gen5, M=0.59, SD=0.19, $t(54)$= -11.51, p<0.001, d = -1.53, 95% CI [-0.37, -0.22]; Gen2 and Gen5, M=0.68, SD=0.17, $t(54)$= -16.21, p<0.001, d = -2.16, 95% CI [-0.45, -0.31]; Gen3 and Gen5, M=0.73, SD=0.17, $t(54)$= -17.14, p<0.001, d = -2.28, 95% CI [-0.52, -0.37]; Gen4 and

Gen5, *M*=0.73, *SD*=0.20, $t(54) = -17.23$, $p < 0.001$, $d = -2.29$, 95% CI [-0.52, -0.36]]. Mirroring the results of similarity across generations, the same holds for the similarity between Generation 1 and Generation 5 versus similarity between each subsequent generation and Generation 5 [pairwise: Gen2 and Gen5: $t(54) = -4.56$, $p < 0.001$, $d = -0.61$, 95% CI [-0.14, -0.03]; Gen3 and Gen5: $t(54) = -5.55$, p<0.001, $d = -0.74$, 95% CI [-0.22, -0.07]; Gen4 and Gen5: $t(54) = -5.48$, $p < 0.001$, $d = -0.73$, 95% CI [-0.22, -0.07]]. By Generation 2, however, we no longer see differences in similarity to the last generation. This suggests that by Generation 2, the story has already settled into a relatively stable structural form, consistent with the final retelling on Day 5.

Figure 3

Consistency: Similarity to Last Generation



Note. Results for Consistency, or the Similarity to the Last Generation, Studies 1-3. As stories are told across generations, they become more similar to the last generation. In *B-C*, the green asterisks mark the significant differences between the two Story conditions in each study. The green crosses mark marginal differences. The color matched asterisks (e.g., red asterisks for Study 1) represent significant differences between the structural similarity comparisons for adjacent generations. Error bars represent standard error.

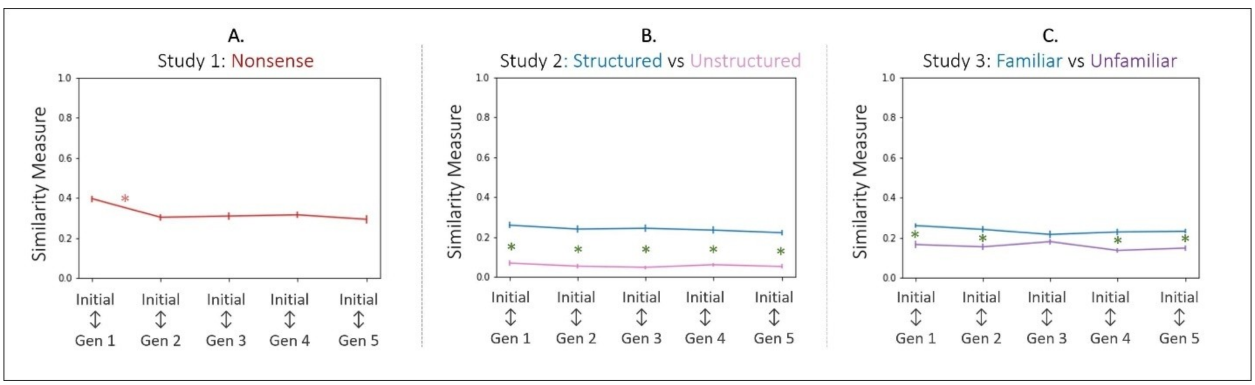
Modification: Similarity to Initial Story

Lastly, we looked at similarity to the initial story to track the degree to which participants are modifying the story they are given (*Figure 4A*). There is a main effect of generation

($F(3.28,177.16)=13.762$, $p<0.001$, $\eta^2_G=0.074$), driven by the first generation. The structural similarity between the Initial story and Generation 1 ($M=0.39$, $SD=0.12$) is significantly higher than all the similarities between the Initial story and each subsequent generation [pairwise: Initial and Gen2, $M=0.30$, $SD=0.12$, $t(54)=6.74$, $p<0.001$, $d = 0.90$, 95% CI [0.05, 0.13]; Initial and Gen3, $M=0.31$, $SD=0.15$, $t(54)=4.61$, $p<0.001$, $d = 0.61$, 95% CI [0.03, 0.14]; Initial and Gen4, $M=0.31$, $t(54)=5.36$, $SD=0.13$, p<0.001, $d = 0.71$, 95% CI [0.04, 0.12]; Initial and Gen5, $M=0.29$, $SD=0.14$, $t(54)=5.23$, $p<0.001$, $d = 0.69$, 95% CI [0.05, 0.16]].The stories' similarity to the initial story does not change across later retellings. This could suggest that by the second generation of recall, the major structural modifications have been made to the story.

Figure 4

Modification: Similarity to Initial Story



Note. Results for Modification, or the Similarity to the Initial Story, Studies 1-3. In *B-C*, the green asterisks mark the significant differences between the two Story conditions in each study. The color matched asterisks (e.g., red asterisks for Study 1) represent significant differences between the structural similarity comparisons for adjacent generations. Error bars represent standard error.

Discussion

Here we tracked the evolution of a nonsense story and found that the nonsense story reaches high levels of stability as early as a person's second retelling. This suggests that participants are efficiently modifying the initial nonsense story early on to transform their stories into more stably recollected forms. Indeed, most of the modifications to the initial story occurred during the first or second generation of recall. This does not mean that all subsequent generations are identical. Rather, this suggests that the event content and event sequence of the story remain highly stable from that point forward.

In this first study, we saw that we can use our measure to track the changes in stories across generations. We first validated the measure against human judgements of similarity, finding promising alignment between the two, and confirmed that the measure is not swayed by potentially trivial textual changes (*e.g.*, length differences). Then, by testing this measure on the progression of nonsense stories, we have established a baseline for expected trends in the evolution of stories across generations. That is, this progression illustrates how we would expect a groundless story, schematically speaking, to change and stabilize across retellings. Armed with this measure, then, we can go on to explore how different factors guide story evolution.

Study 2: Evolution of a Structured vs Unstructured Story

In Study 2, we test the influence of structure on story recall: We ask whether structure gives stories an evolutionary edge, such that they survive better in memory. To do this, we manipulate structure by comparing the evolution of a structured story against that of an unstructured story. The structured story is modeled off of the Cinderella-type tale, a tale for which we expect participants to have strong structural priors. The unstructured story is derived by scrambling the structured story, in an attempt to capture absence of structure (as in Study 1).

Methods

Participants

Participants ($N = 188$) were recruited using Prolific (www.prolific.co; Palan & Schitter, 2018). As in Study 1, we set a target sample size of 50, after attrition. Of the participants who completed Day 1, 64 participants completed all 5 days of the task. Participants not lost to attrition were excluded for not meeting task requirements: For example, writing less than the required count, copy and pasting words over and over, or not writing the story at all, typing in different words. This left us with a final sample size of 56. Participants were randomly assigned to one of two conditions, the Structured condition ($N = 31$) and Unstructured condition ($N = 25$). The different condition counts are due to different attrition and exclusion rates.

All participants were U.S. residents fluent in English. Participants received a total of $7.60 for completing all days of the task ($2.20 on Day 1, 1.10 on Days 2 to 5 each, plus an additional $1.00 for completing all 5 days).

Procedure

Participants completed the task over 5 days. As in Study 1, the task was set up on Qualtrics. On the first day, participants read and recalled one of two stories: 1) A Structured story and 2) An Unstructured version of the same story. The Structured story followed Sara, a high schooler who works towards a science fair that will allow for entrance into a university (See *Supplement*). The story was designed to match a well-established story schema for the target demographic of the task. Specifically, this story was written by experimenters so as to match the main event structure of the Cinderella story, a folktale considered part of Western canon. The Unstructured story contained the same sentences as the canonical story, but these sentences were randomly shuffled, so as to break the continuity of the story (See *Supplement*). Each story was 758 words long. After reading the story, participants were asked to confirm having read the story completely. Next, participants recalled the story they had just read. Participants were encouraged to write as much as possible. They needed to write at least 350 words to submit their response on each day of the task.

As in Study 1, on days 2-5 of the study, participants were invited back, at the same time each day, to re-recall the story they read on the first day. They had 12 hours to participate. Thus, by the end of the study, participants who stayed throughout produced 5 generations of the story they read on the first day, with Generation 1 on Day 1, and Generation 5 on Day 5.

Analyses

Story Dispersion. In Study 2, in addition to tracking changes within participants' retellings, we also compare stories across participants. We measure the similarity of all the participants' stories at each generation of recall, and measure their closeness based on our structural similarity measure. We expect that stories with more recognizable structures are recalled more similarly across participants than stories with unrecognizable structures, with these differences becoming clear by the last generation of recall.

To measure dispersion, we needed to map our stories into a shared 2-dimensional space. To do this, we calculated the structural similarity between all the stories generated across Study 2 and the upcoming Study 3 at each generation (1-5), including stories excluded for length, as well as the initial stories given to participants. We also included stories generated in another study (not reported here) in our comparisons. These stories were retellings across participants (*i.e.*, in a chain) based on the same initial stories as those used in Study 3. (These stories can be found on https://osf.io/hu7ke/.) The more stories we could compare, the more accurately we could map them in relation to each other. These comparisons created a similarity matrix of all the stories, which we converted into a distance matrix. We then applied Multidimensional Scaling to this distance matrix to map all the stories in a 2-d space. Within this space, we can measure dispersion–how spread out the stories are–using Standard Distance Deviation: $SDD = sqrt(sum((x - mean[x])^2 + (y - mean[y])^2) / N(stories))$, where x is the x-dimension of each story in 2-d space and y is the y-dimension of each story in 2-d story space. Here we report the primary dispersion metric of interest, at the last generation of recall, comparing dispersion for the Structured vs. Unstructured story. We used permutation tests to determine whether the difference was significant. The SDD for the remaining generations, and comparisons across conditions are reported in the Supplemental Materials.

Results

This study tests how structure (structured vs unstructured) shapes story evolution. For each of our three metrics of evolution – Stability, Consistency and Modification – we first follow the evolution of each story type, the Structured Story and the Unstructured Story, and then compare the two conditions to one another.

Stability: Similarity Across Generations

We first tested story stability, measured as structural similarity across subsequent generations of recall (*Figure 2B*). We ask first whether the stories in each condition (Structured and Unstructured) become more stable across retellings [main effect of generation: $F(3.47,187.22)= 119.22$, $p<0.001$, $\eta^2_G=0.476$]. Evaluating stability in the Structured condition, we find, as expected, an increase in stability over subsequent recalls (main effect of generation: $F(3.01, 90.37)= 66.62$, $p<0.001$, $\eta^2_G =0.431$). Specifically, the similarity between the Initial story and first generation (M=0.26, SD=0.09) is significantly lower than similarity between all subsequent generations [pairwise: Gen1 and Gen2, (M=0.52, SD=0.15), $t(30)= -13.09$, $p<0.001$, d = -2.29, 95% CI [-0.33, -0.20]; Gen2 and Gen3, (M=0.61, SD=0.17), $t(30)= -12.73$, $p<0.001$, d = -2.22, 95% CI [-0.44, -0.26]; Gen3 and Gen4, (M=0.61, SD=0.19), $t(30)= -10.71$, $p<0.001$, d = -1.88, 95% CI [-0.44, -0.25]; Gen4 and Gen5, (M=0.64, SD=0.19), $t(30)= -12.10$, $p<0.001$, d = -2.12, 95% CI [-0.46, -0.29]]. In addition, the similarity between Generations 1 and 2 is significantly lower than the similarity between Generation 2 and 3 ($t(30)= -4.19$, $p=0.002$, $d = -0.73$, 95% CI [-0.16, -0.02]) and the similarity between Generation 4 and 5 ($t(30)= -4.07$, $p=0.003$, $d = -0.71$, 95% CI [-0.20, -0.03]). There was no significant difference between the similarity of Generations 2 and 3, and any of the subsequent generations. Thus, story structure rapidly settles by the second retelling of the story, with a high level of similarity across all subsequent generations.

Next, we tested for evidence of stabilization within the Unstructured story, ($F(4, 96)= 53.54$, $p<0.001$, $\eta^2_G=0.534$). For the Unstructured story, similarity between the Initial story and first generation ($M=0.07$, $SD=0.08$) is significantly lower than similarity between all subsequent generations [pairwise: Gen1 and Gen2, ($M=0.34$, $SD=0.16$), $t(24)= -9.94$, $p<0.001$, $d = -1.93$, 95% CI [-0.34, -0.20]; Gen2 and Gen3, ($M=0.42$, $SD=0.18$), $t(24)= -9.37$, $p<0.001$, $d = -1.81$, 95% CI [-0.45, -0.25]; Gen3 and Gen4, ($M=0.46$, $SD=0.17$), $t(24)= -11.62$, $p<0.001$, $d = -2.25$, 95% CI [-0.50, -0.30]; Gen4 and Gen5, ($M=0.52$, $SD=0.15$), $t(24)= -14.67$, $p<0.001$, $d = -2.83$, 95% CI [-0.54, -0.35]]. In addition, the similarity between Generation 1 and Generation 2 is lower than the similarity between Generation 3 and 4 ($t(24)= -3.39$, $p=0.02$, $d = -0.66$, 95% CI [-0.22, -0.03]) and the similarity between Generation 4 and 5 ($t(24)= -5.26$, $p<0.001$, $d = -1.01$, 95% CI [-0.27, -0.08]). After Generation 2, similarity levels between subsequent generations stay pretty consistent throughout, with no significant or trending differences between the structural similarity of later generations. Similar to the Structured Story, the structural similarity of the Unstructured Story settles quickly within the first two-retellings.

We then compared the story evolution of the Structured story to that of the Unstructured story. We expected to find higher rates of stabilization for the Structured story. Indeed, having the story follow a clear structure allowed participants to achieve a higher and more consistent level of stabilization. Across all generations, we see greater similarity for the Structured story than the Unstructured story [main effect of story: $F(1,54) = 26.056$, $p<0.001$, $\eta^2_G = 0.221$]. (Note: For the comparison between Generation 4 and Generation 5, similarity is marginally higher for the Structured story than the Unstructured story, ($F(1,54)=6.875$, $p=0.055$, $\eta^2_G = 0.113$, 95% CI [0.03,0.22]. See *Supplement* for differences across story conditions for each generation comparison.)

Consistency: Similarity to Last Generation

In addition to tracking changes in stability, we measured whether the changes are in a consistent direction (*Figure 3B*). To do this, we track the structural similarity of each generation to the final generation of recall on Day 5. If structural similarity increases this suggests that there is consistency in the stories' evolution [main effect of generation: $F(3,162.14)= 166.36$, $p<0.001$, $\eta^2_G = 0.545$]. As expected, Structured stories not only stabilize quickly, they also consistently evolve in a unique direction, toward their final form. This can be seen by looking at structural similarity between each Generation and the Generation produced on the last day (Generation 5), where consistency increases with retellings ($F(2.72,81.56)= 91.087$, $p<0.001$, $\eta^2_G = 0.493$): The similarity between the Initial Story and Generation 5 (M=0.22, SD=0.09) is lower than all subsequent generations' similarity to Generation 5 [pairwise: Gen1 and Gen5, M=0.47, SD=0.17, $t(30) = -9.96$, $p<0.001$, $d = -1.74$, 95% CI [-0.33, -0.17]; Gen2 and Gen5, M=0.55, SD=0.15, $t(30) = -14.44$, $p<0.001$, $d = -2.53$, 95% CI [-0.39, -0.26]; Gen3 and Gen5, M=0.62, SD=0.16, $t(30) = -16.97$, $p<0.001$, $d = -2.97$, 95% CI [-0.48, -0.32]; Gen4 and Gen5, M=0.64, SD=0.19, $t(30) = -13.56$, $p<0.001$, $d = -2.37$, 95% CI [-0.51, -0.33]]. The similarity between Generation 1 and 5 is lower than all subsequent generations' similarity [pairwise: Gen2 and Gen5, $t(30) = -4.25$, p=0.002, $d = -0.74$, 95% CI [-0.14, -0.02]; Gen3 and Gen5, $t(30) = -5.52$,

$p<0.001$, $d = -0.97$, 95% CI [-0.22, -0.07]; Gen4 and Gen5, $t(30) = -4.92$, $p<0.001$, $d = -0.86$, 95% CI [-0.26, -0.07]] . The similarity between Generation 2 and 5 is lower than all subsequent generations' similarity [pairwise: Gen3 and Gen5, $t(30) = -3.63$, $p=0.01$, $d = -0.64$, 95% CI [-0.13, -0.003]; Gen4 and Gen5, $t(30) = -3.43$, $p=0.018$, $d = -0.60$, 95% CI [-0.16, -0.02]]. The similarity between Generation 3 and 5 versus between Generation 4 and 5 are not significantly different. This suggests that by Generation 3 (Day 3 of the task), participants have settled on a relatively stable structure of the story in memory.

The Unstructured stories likewise consistently evolve in a unique direction, toward their final form ($F(3.01,72.36)= 75.874$, $p<0.001$), $\eta^2_G = 0.617$. As with the Structured stories, the similarity between the Initial story and Generation 5 ($M=0.05$, $SD=0.06$) is lower than the similarity between all subsequent generations and Generation 5 [pairwise: Gen1 and Gen5, $M=0.32$, $SD=0.15$, $t(24)= -8.07$, $p<0.001$, $d = -1.56$, 95% CI [-0.36, -0.18]; Gen2 and Gen5, $M=0.42$, $SD=0.12$, $t(24)= -14.59$, $p<0.001$, $d = -2.83$, 95% CI [-0.44, -0.29]; Gen3 and Gen5, $M=0.47$, $SD=0.16$, $t(24)= -11.56$, p<0.001, $d = -2.24$, 95% CI [-0.50, -0.32]; Gen4 and Gen5, $M=0.52$, $SD=0.15$, $t(24)= -14.00$, $p<0.001$, $d = -2.71$, 95% CI [-0.56, -0.36]]. The same is true for the similarity between Generation 1 and Generation 5 compared to subsequent generations [Gen2 and Gen5, $t(24)= -3.89$, $p=0.007$, $d = -0.75$, 95% CI [-0.16, -0.03]; Gen3 and Gen5, $t(24)= -4.95$, $p<0.001$, $d = -0.96$, 95% CI [-0.23, -0.06]; Gen4 and Gen5, $t(24)= -5.45$, $p<0.001$, $d = -1.06$, 95% CI [-0.30, -0.09]]. Lastly, the similarity between Generations 2 and 5 is lower than the similarity between Generations 4 and 5 ($t(24)=-3.85$, $p=0.008$, $d =-0.75$, 95% CI [-0.18, -0.02]).

Importantly, for both the Structured and Unstructured story, we never find that the similarity to the last generation decreases. This suggests that the story is moving consistently towards a 'final form', rather than changing in inconsistent ways.

Finally, we compared the evolution of the Structured story to that of the Unstructured story. Across all generations, we find greater similarity to the last generation for the Structured story than the Unstructured story [main effect of story: $F(1,54) = 22.48$, $p<0.001$, $\eta^2_G = 0.203$], suggesting that the Structured story remains consistently closer to its final form through all retellings. (Note: For the comparison between Generation 4 and Generation 5, similarity is marginally higher for the Structured story than the Unstructured story, ($F(1,54)=6.875$, $p=0.055$, $\eta^2_G = 0.113$, 95% CI [0.03,0.22]. See *Supplement*.)

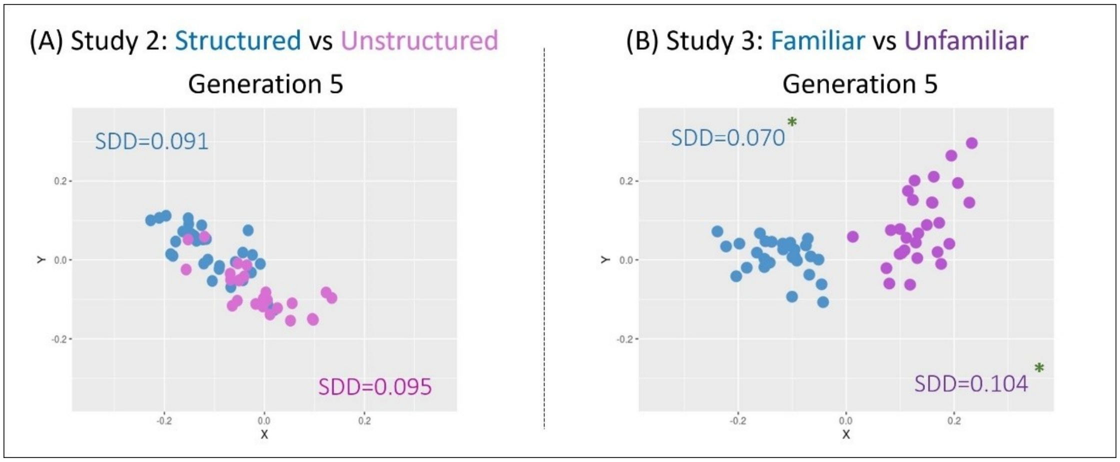
Modification: Similarity to the Initial Story

Lastly, we tracked how the initial story is modified across retellings. To do this, we compare the structural similarity of each generation to the initial story (*Figure 4B*). For both the Structured and Unstructured story, we do not find a main effect of generation, suggesting that there are no differences in the similarity to the initial story across generations. That is, independently for each story (Structured or Unstructured), the initial story is equally preserved across generations.

We then compare across story conditions to test whether structure influences the degree to which the initial story is modified in memory. As expected, we see greater similarity to the initial story for the Structured story than the Unstructured story (main effect of story: $F(1,54)=104.35$, $p<0.001$, $\eta^2_G = 0.554$; pairwise: Initial and Gen1, $F(1,54)=67.38$, $p<0.001$, $\eta^2_G = 0.555$, 95% CI [0.14, 0.24]; Initial and Gen2, $F(1,54)=64.67$, $p<0.001$, $\eta^2_G = 0.545$, 95% CI [0.14, 0.23]; Initial and Gen3, $F(1,54)=68.51$, $p<0.001p<0.001$, $\eta^2_G = 0.559$, 95% CI [0.15, 0.24]; Initial and Gen4, $F(1,54)=63.68$, $p<0.001$ $\eta^2_G = 0.541$, 95% CI [0.13, 0.22]; Initial and Gen5, $F(1,54)=72.77$, $p<0.001$, $\eta^2_G = 0.574$, 95% CI [0.13, 0.21]). That is, the initial Structured story is better preserved in memory across generations than the initial Unstructured story. (See *Supplement* for differences across story conditions for each generation comparison.)

Figure 5

Story Dispersion



Note. This figure illustrates the Story Dispersion results for Study 2 (A) and Study 3 (B). Every story generated by participants is mapped onto a 2-d space, and categorized by color. (A): In Study 2, the stories generated in the Structured condition are in blue (*N*=31). The stories generated in the Unstructured condition are in pink (*N*=25). (B): In Study 3, the stories generated by participants in the Familiar condition are in blue (*N*=27). The stories generated in the Unfamiliar condition are in purple (*N*=28). SDD above refers to the Standard Distance Deviation of each story grouping, color matched to its condition. Asterisks indicate that the condition SDDs are significantly different from one another.

Story Dispersion

So far, all the measures reported track changes within individual participants' memory. These measures capture the first aspect of story survival. The second aspect of story survival has to do with story retellings across people. A story with an evolutionary edge should have greater agreement in recall across participants, such that the stories remain similarly structured across people. To test this, we measure story dispersion by the last generation of recall (*Figure 5A*), and ask whether the Structured story is more tightly dispersed than the Unstructured story. By generation 5, we do not find differences in dispersion for the Structured (*SDD*=0.091) versus the

Unstructured (*SDD*=0.095), *p*=ns. Against expectations, this suggests that participant recall is just as similar across those who recalled the Structured story as those who recalled the Unstructured story. In addition, as shown in *Figure 5A*, there is clear overlap between the distribution of the Structured story points (in blue) and Unstructured story points (in pink), suggesting that participants across both conditions converged upon a similar story structure.

Discussion

Study 2 tested the impact of story structure on story evolution. The results from Study 2 suggest that the Structured story represents a more stable version of a story than the Unstructured story. Similarity across generations, similarity to the last generation, and similarity to the initial story were higher for the structured story. This provides evidence that stories with clear structures allow for greater stability in memory. The Structured story becomes stable very quickly, settling into a form that is more similar to its initial story. That is, the initial story is subject to less modification. The Unstructured story is less stable, and is subject to greater modification—as represented by the lower structural similarity values.

We did not find differences in the dispersion of final stories across the two conditions. That is, cross-participant recall is just as similar–or as dispersed–for the Structured story as for the Unstructured story. This suggests that the participants converge on a final story to a similar extent for both the Structured and Unstructured story. Further, there is clear overlap in the mapping of both stories by generation 5, suggesting that participants converged on a similar final story for both conditions. Given that the Unstructured story is a scrambled version of the Structured story, we speculate that participants are essentially unscrambling the former into the latter. Indeed, participants in the Unstructured condition of the study often made the story coherent by unscrambling the story, even as early as Generation 1. In other words, people may have unscrambled the Unstructured story in similar ways because of the widespread familiarity with the Cinderella-form, showing even further sensitivity to its intended structure. We directly address the effect of structural familiarity in Study 3 by comparing the evolution of this same Structured story, modeled off of Cinderella, and a structured but Unfamiliar story.

Stories are typically considered to be *coherent* sequences of events. In the absence of a clear structure, an unstructured story could arguably be discounted as a story. However, storytelling as it occurs in the world varies in structure from the nonsense of a toddlers' ramblings to the highly codified structure of, for instance, serialized detective dramas. These results from Study 2 shed light on the effects of structure, both in the case of a clear structure and of a structure broken apart by scrambling.

Study 3: Evolution of a Familiar vs Unfamiliar Story

Study 3 tests how having a culturally *familiar* structure shapes the evolution of stories across retellings. Specifically, we test how having a familiar structure enables participants to recall a story with more stability over time, and, how having a familiar structure leads to more similar recall across participants. To do so, we compare the evolution of two stories that have comparably coherent structure. One story follows the familiar structure of the Cinderella tale (the same story used in Study 2), whereas the other follows an unfamiliar structure, modified to be different from the Cinderella-type sequence of events. If structural familiarity, above and beyond coherence, contributes to differences in (1) the way stories evolve and (2) the agreement in recall between participants, this would suggest that existing story schemas facilitate recollection.

Methods

Participants

As with Study 2, participants (*N* = 128) were recruited using Prolific, with a target sample size of 50. Of the participants who completed Day 1, 55 usable participants completed all 5 days of the task. Participants were excluded using the same exclusionary criteria as in previous studies. Participants were randomly assigned to one of two conditions, the Familiar condition (*N* = 27) and Unfamiliar condition (*N* = 28).

All participants were U.S. residents fluent in English. Participants received a total of $7.60 for completing all days of the task, following the same payment scheme as in Study 2.

Procedure

As in Study 1 and 2, participants completed the task over 5 days. The procedure matched that of Study 2, with one major change in manipulation. Participants either read the Familiar Story (Word Count=758), the same story modeled off of the Cinderella-type story as used in Study 2, or an unfamiliar story (Word Count=756). The Familiar story here is the same as the Structured story in Study 2. The Unfamiliar story was written to include many events similar to those in the Familiar story, and to match as closely as possible for length, reading ease (Flesch Reading Ease for Familiar Story=72.297, for Unfamiliar Story=72.246) and cohesion (Deep cohesion percentile for Familiar Story=29.81, for Unfamiliar Story=26.76) according to Coh-Metrix measures (Graesser et al, 2004). It was made unfamiliar by changing Sara's role from that of science fair participant to that of a science fair organizer. This story thus no longer easily mapped onto any existing popular story type.

As in Study 1 and 2, by the end of Study 3, participants who stayed throughout produced 5 generations of the story they read on the first day (Generations 1 to 5).

Analysis

Study 3 applied the same analysis techniques used in Study 2 to track the structural changes of the stories across generations of retellings and story agreement across participants (story dispersion).

Results

Study 3 asks how structural familiarity shapes the evolution of stories: We first track the evolution of each story type (Familiar and Unfamiliar) and compare these evolutions to one another. We then look at story dispersion, which captures agreement across participants.

Stability: Similarity Across Generations

We first measured stability, asking whether structural similarity increases across retellings (main effect of generation: $F(4,212)=119.26$, $p<0.001$, $\eta^2_G = 0.452$), (*Figure 2C*). In the Familiar condition, we replicate the pattern identified in Study 2 ($F(4,104)=48.040$, $p<0.001$, $\eta^2_G = 0.406$). As expected, the similarity between the Initial Story and Generation 1 ($M=0.26$, $SD=0.05$) is lower than the similarity between all subsequent generations [pairwise: Gen1 and Gen2, $M=0.46$, $SD=0.16$, $t(26)= -6.84$, $p<0.001$, $d = -1.28$, 95% CI [-0.28, -0.12]; Gen2 and Gen3, $M=0.51$, $SD=0.14$, $t(26)= -9.18$, $p<0.001$, $d = -1.72$, 95% CI [-0.34, -0.15]; Gen3 and Gen4, $M=0.54$, $SD=0.16$, $t(26)= -9.84$, $p<0.001$, $d = -1.84$, 95% CI [-0.36, -0.19]; Gen4 and Gen5, $M=0.57$, $SD=0.13$, $t(26)= -12.74$, $p<0.001$, $d = -2.38$, 95% CI [-0.38, -0.24]]. The similarity between Generations 1 and 2 is lower than the similarity between Generations 3 and 4 ($t(26)= -3.14$, $p=0.042$, $d = -0.59$, 95% CI [-0.15, -0.001]) and between Generations 4 and 5 ($t(26)= -4.97$, $p<0.001$, $d = -0.93$, 95% CI [-0.18, -0.04]). Lastly, the similarity between Generations 2 and 3 is marginally lower than the similarity between Generations 4 and 5 ($t(26)= -2.95$, $p=0.066$, $d = -0.55$, 95% CI [-0.13, 0.001]). There is no difference between the similarity of Generations 3 and 4 and of Generations 4 and 5. Generally speaking, similarity across generations increases across retellings, indicating that the Familiar story is evolving to become more stable in memory.

We find similar increases in stability across generations for the Unfamiliar story (F(2.93,79.09)=72.837, p<0.001, η^2_G= 0.496). The similarity between the Initial Story and Generation 1 (M=0.17, SD=0.09) is lower than the similarity between all subsequent generations [pairwise: Gen1 and Gen2, M=0.39, SD=0.13, t(27)= -9.02, p<0.001, d = -1.66, 95% CI [-0.31, -0.15]; Gen2 and Gen3, M=0.49, SD=0.10, t(27)= -9.37, p<0.001, d = -1.72, 95% CI [-0.41, -0.23] ; Gen3 and Gen4, M=0.54, SD=0.16, t(27)= -12.23, p<0.001, d = -2.25, 95% CI [-0.46, -0.29]; Gen4 and Gen5, M=0.54, SD=0.13, t(27)= -15.64, p<0.001, d = -2.87, 95% CI [-0.44, -0.30]]. The same is true for the similarity between Generation 1 and 2, as compared to all subsequent generations [pairwise: Gen2 and Gen3, t(27)= -3.18, p=0.037, d = -0.58, 95% CI [-0.17, -0.02] ; Gen3 and Gen4, t(27)= -5.67, p<0.001, d = -1.04, 95% CI [-0.22, -0.07]; Gen4 and Gen5, t(27)= -6.11, p<0.001, d = -1.12, 95% CI [-0.21, -0.08]]. For the following generations (Gen 2 with Gen3 onwards), there are no significant differences between similarity across generations. This pattern of evolution suggests that the Unfamiliar story is also evolving to become more stable in memory.

Next, we compared the evolution of the Familiar and Unfamiliar story: As expected, we find a difference in stability in the first generation of retellings. Similarity across generations is higher for the Familiar story for the comparison between the Initial story and Generation 1 (F(1,53)=20.820, p<0001, η^2_G= 0.282). (*Figure 2C*). For subsequent generations, there are no differences between the two story lineages. In other words, the Unfamiliar story starts out less stable in participants' memory. However, as early as Generation 2, the Unfamiliar story closes the gap, matching the Familiar Story in terms of stability across generations. Indeed, there is a marginal interaction effect between story and generation (*F(4,212)*=2.37, *p*=0.054, η^2_G= 0.016). (See *Supplement*.)

Consistency: Similarity to Last Generation

Next, we look at whether the stories are changing in a consistent direction by tracking changes in similarity to Generation 5 (*Figure 3C*). For the Familiar story, the retellings consistently evolve toward their final form (*F*(4,104)=68.30, *p*<0.001, η^2_G= 0.476): The similarity between the Initial Story and Generation 5 (*M*=0.23, *SD*=0.06) is, once again, lower than all subsequent generations' similarity to Generation 5 [pairwise: Gen1 and Gen5, *M*=0.48, *SD*=0.15, *t*(26)= -9.53, *p*<0.001, *d* = -1.78, 95% CI [-0.33, -0.17]; Gen2 and Gen5, *M*=0.49, *SD*=0.17, *t*(26)= -8.84, *p*<0.001, *d* = -1.65, 95% CI [-0.35, -0.17]; Gen3 and Gen5, *M*=0.56, *SD*=0.12, *t*(26)= -16.99, *p*<0.001, *d* = -3.17, 95% CI [-0.41, -0.25]; Gen4 and Gen5, *M*=0.57, *SD*=0.13, *t*(26)= -15.14, *p*<0.001, *d* = -2.83, 95% CI [-0.40, -0.27]]. In addition, the similarity between Generation 1 and Generation 5 is lower than the similarity between Generations 3 and 5 (*t*(26)= -3.79, *p*=0.009, *d* = -0.71, 95% CI [-0.15, -0.01]) and the similarity between Generations 4 and 5 (*t*(26)= -3.70, *p*=0.001, *d* = -0.69, 95% CI [-0.16, -0.03]). Lastly, the similarity between Generation 2 and 5 is lower than the similarity between Generations 4 and 5 (*t*(26)= -3.17, *p*=0.038, *d* = -0.59, 95% CI [-0.15, -0.01]). By Generation 3, however, the stories settle into a high similarity to Generation 5 with no significant differences as compared to later generations' similarity to Generation 5.

For the Unfamiliar story, we find similar trends (*F*(3.05,82.43)=82.43, *p*<0.001, η^2_G= 0.468): The similarity between the Initial Story and Generation 5 (*M*=0.15, *SD*=0.06) is lower than all subsequent generations' similarity to Generation 5 [pairwise: Gen1 and Gen5, *M*=0.34, *SD*=0.16, *t*(27)= -6.65, *p*<0.001, *d* = -1.22, 95% CI [-0.27, -0.11]; Gen2 and Gen5, *M*=0.45, *SD*=0.19, *t*(27)= -9.12, *p*<0.001, *d* = -1.68, 95% CI [-0.39, -0.21]; Gen3 and Gen5, *M*=0.50, *SD*=0.19, *t*(27)= -10.74, *p*<0.001, *d* = -1.97, 95% CI [-0.43, -0.27]; Gen4 and Gen5, *M*=0.54, *SD*=0.13, *t*(27)= -17.21, *p*<0.001, *d* = -3.16, 95% CI [-0.46, -0.33]]. The same holds for the similarity between Generation 1 and Generation 5, as compared to subsequent generations [pairwise: Gen2 and Gen5, *t*(27)= -4.39, *p*=0.002, *d* = -0.81, 95% CI [-0.18, -0.04]; Gen3 and Gen5, *t*(27)= -6.63, *p*<0.001, *d* = -1.22, 95% CI [-0.23, -0.09]; Gen4 and Gen5, *t*(27)= -9.87, *p*<0.001, *d* = -1.81, 95% CI [-0.26, -0.13]]. Lastly, the similarity between Generation 2 and Generation 5 is lower than the similarity between Generations 4 and 5 (*t*(27)= -4.11, *p*=0.003, *d* = -0.75, 95% CI [-0.16, -0.02]). By Generation 3, once again, there is no increase in similarity to the final form: That is, similarity between Gen 3 and Gen 5 is no different than the similarity between Gen4 and Gen5.

For both the Familiar and Unfamiliar story, similarity to the final form never decreases across retellings. This suggests (as in Study 1 and Study 2) that, as the stories are told and retold, they change in a consistent direction (main effect of generation: *F*(3.37,178.68)=139.624, *p*<0.001, η^2_G= 0.465).

Next, we compared the evolutions of the Familiar and Unfamiliar story. We find a main effect of story, driven by the earlier generations (*F*(1,53)=5.12, *p*=0.028, η^2_G= 0.061). That is, the differences between stories only hold for earlier generations: Similarity to last generation is higher for the Familiar versus Unfamiliar story for similarity between Initial and Generation 5 (*F*(1,53)=25.16, *p*<0.001, η^2_G= 0.322, 95% CI [0.05, 0.12]) and between Generation 1 and Generation 5 (*F*(1,53)=10.97, *p*=0.001, η^2_G= 0.171, 95% CI [0.05, 0.22]; *Figure 3C*). For subsequent generations' similarity to Generation 5, there are no differences between story conditions. Once again, the Unfamiliar story closes the gap with the Familiar story. Once again, this is highlighted by an interaction effect between story and generation (*F(3.37,178.68*)=3.17, *p*=0.027, η^2_G= 0.018). (See *Supplement*.)

Modification: Similarity to Initial Story

Last amongst the within participant measures, we look at how the initial story is modified across retellings (*Figure 4C*). As in Study 2, we do not find a main effect of comparison between generations, suggesting that there are no differences in the similarity to the initial story across generations.

However, we do, once again, find a main effect of story (*F*(1,53)=23.83, *p*<0.001, η^2_G= 0.212), with the Familiar story (*M*=0.23, *SD*=0.02) being more similar to the initial story than is the Unfamiliar story (*M*=0.16, *SD*=0.02). This effect is found in all generations, except for Generation 3 [Initial and Gen1: Familiar (*M*=0.26, *SD*=0.06) vs Unfamiliar (*M*=0.17, *SD*=0.09), *F*(1,53)=20.82, *p*<0.001, η^2_G= 0.282, 95% CI [0.05, 0.14]; Initial and Gen2: Familiar (*M*=0.24, *SD*=0.09) vs Unfamiliar (*M*=0.15, *SD*=0.08), *F*(1,53)=15.42, *p*<0.001, η^2_G= 0.225, 95% CI [0.04, 0.13]; Initial and Gen4: Familiar (*M*=0.23, *SD*=0.08) vs Unfamiliar (*M*=0.14, *SD*=0.07), *F*(1,53)=19.32, *p*<0.001, η^2_G= 0.267, 95% CI [0.05, 0.13]; Initial and Gen5: Familiar (*M*=0.23, *SD*=0.06) vs Unfamiliar (*M*=0.15, *SD*=0.06), *F*(1,53)=25.16, *p*<0.001), η^2_G= 0.322, 95% CI [0.05, 0.12]]. These differences suggest that, as the story is told and retold, the Unfamiliar story is subject to greater modification than the Familiar story. (See *Supplement*.)

Story Dispersion

Finally, we compared how stories are told across participants using the story dispersion measure. We looked at the dispersion of stories between participants in the last generation of recall. We find a significant difference in dispersion, with the Familiar story having lower dispersion (*SDD*=0.070) than the Unfamiliar Story (*SDD*=0.104), *p*=0.046 by Generation 5

(*Figure 5B*). These differences start to appear by Generation 4 (See *Supplement*). This suggests that recall is more similar across participants for the Familiar story than the Unfamiliar story. In other words, the Unfamiliar story recall diverges more across participants.

Unlike in Study 2, we find no overlap in the distribution of the Familiar story points (in blue) and Unfamiliar story points (in purple). The two groups of stories occupy distinct parts of the 2-d story space we generated (*Figure 5B*). This suggests that, as the stories are told and retold, they do, indeed, remain two distinct story lineages.

Discussion

In Study 3, we asked how stories with varying degrees of structural familiarity fare within an individual's memory. We found that structural familiarity allows for greater initial structural stability. Structurally familiar stories start out more stable in participants' memory. This suggests that a familiarly structured story, as one modeled after the Cinderella-type tale, latches better onto memory. The familiarly structured story does not require as much modification in order to become stable in participants' memory. Unfamiliar stories, on the other hand, are modified more in individuals' memory, as demonstrated by lower levels of stability in earlier retellings. However, across retellings, the unfamiliar stories become more stable, suggesting that these modifications lead the stories toward more stable forms in people's memory. Indeed, participants' memory processes seem to rapidly and efficiently modify the unfamiliarly structured story into a more stable form: As early as Generation 2, there is no longer any difference between the stability of the Familiar and Unfamiliar story. While this is not addressed in this series of studies, future research could investigate the kinds of modifications that allow for greater stability. What changes occurred in participants' memory that allowed the unfamiliar story to become more stable across retellings?

Participants differed in how they modified the Familiar and Unfamiliar stories. There was a greater divergence in recall for the unfamiliarly structured stories than for more familiarly structured stories. By the last generation of recall, the Unfamiliar story was more dispersed in story space than the Familiar story. The familiarly structured story, in contrast, was recalled with greater agreement across participants, pointing at an evolutionary edge: Not only is it subject to less modification, it also recalled more faithfully across people. A Cinderella tale remains a Cinderella tale no matter who recounts it.

General Discussion

Stories have developed, over history, into a consistent set of shapes that define a culture's modern narrative repertoire. How do stories evolve and take shape over retellings? Across 3 studies, we address this question by varying the initial structure (structured or unstructured) and familiarity (familiar or unfamiliar) of stories. We used cultural commonness as a proxy for familiarity, selecting the Cinderella structure as an extremely common and, therefore, highly familiar story structure. As expected, we find that structured and familiar stories are more stable and subject to less modification in people's memory. Unstructured and unfamiliar stories, on the other hand, underwent greater modification in people's memory before becoming comparably stable.

Our results also show that people converge in how they recall familiarly structured stories. By the fifth generation of recall, the stories generated by participants are more structurally similar to each other when the initial story was familiarly structured than when it was unfamiliarly structured. This result builds upon prior findings to show that, not only is recall for familiarly structured stories more stable *within* people (Bartlett 1932, Harris et al, 1988), it is also more consistent *across* people. Structured and familiar stories fit better into *individuals'* schemas, *and* these individual schemas are shared across people. People's recollections for familiarly structured stories converge, whereas the greater modifications across retellings cause people's recollections for unfamiliarly structured stories to diverge. The studies reported here cannot provide insight into exactly what these modifications are. However, previous research (Bartlett, 1932; Kintsch & Greene, 1978) suggests that unfamiliar stories are modified to better fit people's schemas. To the extent that recall accuracy reflects schematic fit, then the current results likewise suggest that stories are modified over retellings to better fit into people's schemas. These schemas, however, may not be as consistent across people as those that align with familiarly structured stories: Study 3 finds that these modifications lead to greater differences in recall for an unfamiliarly structured story, suggesting that, without the shared Cinderella structure, recall may be guided by different schemas across people.

Familiar story structures may serve as a kind of shared cultural schema, such that people within a cultural context share similar story structures. Differences in individuals' experiences with stories will certainly lead to differences in their story schemas, but there is overlap–a shared story canon. Why are *these* the story structures that entered the canon? In other words, why are these the stories that became familiar? Perhaps it is simply that these are the stories that have been told over and over again, passed down through the generations. Alternatively, it is possible that there are characteristics of certain story structures that make them better stories, above and beyond mere familiarity.

We propose that stories' success may be related to the role they play in society. Stories are a source of social information. Thus, stories with higher informativeness may fare better, especially when subject to transmission across different minds. According to the concept of the prospective brain, human memory maintains what is useful to predict and navigate future situations (Schacter, Addis & Buckner, 2007). The story structures that are retained in both the single mind and across minds thus should be those that are useful in navigating the expected structure of our world. We suggest that a useful story structure is one that teaches people something about human behavior and the social world: If there is nothing to learn from the story, it may not be seen as useful. Some unexpected elements may be crucial to a successful story. On the other hand, if the story depicting a sequence of behaviors that are too unexpected, it may not be useful either. This proposed curvilinear effect mimics findings by Norenzayan et al. (2006) that suggest that stories with minimally counterintuitive content are more culturally successful. Grimm folktales with fewer counterintuitive elements (*i.e.*, that defy assumptions about objects' behaviors) tend to be more popular than folktales with only intuitive or mostly counterintuitive elements. Norenzayan et al. (2006) argue that this is because minimally counterintuitive content is more memorable–and therefore more likely to be passed on and spread–then entirely intuitive or counterintuitive content. Our proposal also fits with work showing that some coherence gaps help memory and learning for texts, as compared to no gaps or too many gaps (Kintsch, 1994). Story structures that fit this happy medium may have been more memorable, enjoyable and more likely to be passed down. Thus, as an amendment to Terry Pratchett's description of stories as "parasitic life form[s]", we propose that stories are "symbiotic life forms": We carry them across time and space because they are useful to us.

Stories as Symbiotes

Stories serve as a rich source of information about human behavior and our social world. Previous research has suggested that stories provide valuable social simulations (Mar & Oatley, 2008), enabling people to improve their theory of mind and knowledge about social situations (Zunshine, 2006). Fiction reading has been associated with stronger theory of mind ability

(Tamir et al., 2016). In children between four and six, those who have more experience with juvenile fiction do better on theory of mind tasks (Mar et al., 2010). In adults, fiction reading has been causally linked to an increase in theory of mind ability (Kidd & Castano, 2013; Doddell-Feder & Tamir, 2018) and empathy (Bal & Veltkamp, 2013). Perhaps stories providing better sources of social simulation (i.e., the more socially informative stories) would fare better under evolutionary pressures. And indeed, you would be hard pressed to find stories that do not follow the actions and emotions of people or anthropomorphized entities (Zunshine, 2006).

The research reported here does not yet allow us to distinguish the potential influence of social informativeness on story evolution from that of familiarity. We find that the culturally familiar Cinderella-structure is better preserved as it is told and retold, suggesting that it is schematically fit. But we do not yet know why: It is possible that the Cinderella-structure is schematically fit because it is highly familiar. Or, it is also possible that the Cinderella-structure is both schematically fit and culturally widespread because it is a reliable source of social learning. Given the likely co-occurrence of familiarity, structure, and social informativeness in modern, common stories, it is difficult to disambiguate between these factors. Stories could have become widespread, and more highly structured, at least in part, precisely because they depicted socially informative sequences of events.

This story evolution design can allow us to address this question in future research, however. Indeed, there is a lot that can be done in this story space. Future research can disambiguate the impact of structural familiarity from that of internal story characteristics by exploring story forms from different cultural contexts. Are there indeed certain structural or social informativeness characteristics that make even an unfamiliar story more stable in a relatively naïve mind? This could be accomplished by tracking how stories that have "evolutionarily fit" structures in one culture fare in cultures unfamiliar with the story schema. If the success of its structure in one context points to a general advantage, this story should survive well cross-culturally. If the familiarity of the structure matters, however, then this story should not survive well out of its home culture.

Modeling the Structure of Stories

Another avenue of research would be to consider different ways of modeling story structure. The analyses developed in these four studies were designed to track structural similarity across generations, and were intended to handle a constrained space of text. Given a same initial story, there are only so many ways that the story can be modified. However, it gives us little insight into what is actually being modified across generations and what is being maintained. Nor does it tell us what patterns may be found in modification and maintenance across many different story forms—in part because we only use a select few here. How are we to model story structure?

Researchers have identified story regularities within and across cultures to create catalogs and phylogenies of folktales. Such catalogs can be massive, with some indexes categorizing up to 2500 distinct plot types (Uther, 2004). Is this the only way we can make sense of story structure? Or, can we model story structure so as to capture more *fundamental* regularities? In his rejected Anthropology Masters' thesis, Kurt Vonnegut, Jr. proposed his theory about the shape of stories (https://www.youtube.com/watch?v=oP3c1h8v2ZQ). He argued the following: Take any story plot, feed it into a computer, and see it expressed as one of eight simple shapes. He drew graphical representations of story progressions, plotting the main character's fortune—on an axis of ill fortune to great fortune—from story beginning to story end.

Vonnegut's ambition has been shared by many scholars across diverse fields, from literature to anthropology to psychology. Across fields, the core of the ambition remains the same: Stories follow certain common shapes. As complex as they may seem, their underlying structure walks along familiar, well-trodden ground. Scholars have attempted to model story structure in a variety of ways. Some have emulated Vonnegut and attempted to reduce stories into shapes built according to a single dimension. Often, these attempts will plot a story's emotion or valence (e.g., good vs. bad) over time. A recent attempt at cataloging story shapes, for example, scoured fiction from the Project Guttenberg collection to reveal six distinct, basic emotional arcs that represent the basic building blocks of story (Reagan et al, 2016).

Valence, however, may not capture the whole story. Additional dimensions are likely critical to capturing how people represent stories. Future work should project story texts into a multi-dimensional space informed by reliable mental state and action dimensions (Tamir & Thornton, 2018; Thornton & Tamir, 2019). If researchers can map stories as trajectories traversing this space, and, if similarity values based off of these maps align with human judgements of similarity, we may be able to build a 'story space' into which any story may be projected onto and reduced. Furthermore, this story space would be one that is defined by psychologically relevant dimensions. This may provide us further insight into the social and psychological information carried within those stories. Rather than being defined by the one-dimensional good fortune-ill fortune axis, the stories would be mapped onto a higher-dimensional space that may, arguably, better capture how the events therein are actually represented in people's minds. This could, in turn, allow us to make sense of the patterns in story structures and how these might correspond to differing levels of social informativeness.

Conclusion

In the three studies reported here, we employ novel natural language processing analyses to illustrate how familiarly structured stories possess an evolutionary edge over unstructured, unfamiliar stories. All stories used here are new to participants, and yet we find that those with familiar structures not only survive better within individuals' memory, they survive more consistently across people. In future work, we will ask why these are the stories that survive. By moving forward with cross-cultural research and continuing to develop new story analysis techniques, we can ask: What about a story structure makes it a good, strong story? One that will be remembered and passed down through the generations. In our own way, we would be able to answer Vonnegut's hopes: We could feed any story into a computer to find its shape. And in doing so, using psychologically relevant dimensions, we could learn how these stories feed into and inform people's understanding of their social worlds.

References

Alaoui-Soce, A. (2023, May 28). Familiar Story Structures Possess an Evolutionary Edge in Memory. Retrieved from osf.io/hu7ke

Alba, J. W., & Hasher, L. (1983). Is memory schematic?. Psychological Bulletin, 93(2), 203.

Bal, P. M., & Veltkamp, M. (2013). How does fiction reading influence empathy? An experimental investigation on the role of emotional transportation. *PloS one*, *8*(1), e55341.

Bartlett, F.C. (1932).*Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Bergman, E. T., & Roediger, H. L. (1999). Can Bartlett's repeated reproduction experiments be replicated? *Memory & cognition*, *27*(6), 937-947.

Bortolini, E., Pagani, L., Crema, E. R., Sarno, S., Barbieri, C., Boattini, A., ... & Tehrani, J. J. (2017). Inferring patterns of folktale diffusion using genomic data. Proceedings of the National Academy of Sciences, 114(34), 9140-9145.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive psychology*, *11*(2), 177-220.

Coman, A., Momennejad, I., Drach, R. D., & Geana, A. (2016). Mnemonic convergence in social networks: The emergent properties of cognition at a collective level. Proceedings of the National Academy of Sciences, 113(29), 8171-8176.

Da Silva, S. G., & Tehrani, J. J. (2016). Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales. *Royal Society open science*, *3*(1), 150645.

De la Rochère, M. H. D. (2016). *Cinderella across cultures: New directions and interdisciplinary perspectives*. Wayne State University Press.

Diedenhofen, B., & Musch, J. (2015). cocor: A comprehensive solution for the statistical comparison of correlations. PloS one, 10(4), e0121945.

Dodell-Feder, D., & Tamir, D. I. (2018). Fiction reading has a small positive impact on social cognition: A meta-analysis. *Journal of Experimental Psychology: General*, *147*(11), 1713.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers*, *36*(2), 193-202.

Harris, R. J., Lee, D. J., Hensley, D. L., & Schoen, L. M. (1988). The effect of cultural script knowledge on memory for stories over time. *Discourse Processes*, *11*(4), 413-431.

Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. The Journal of general psychology, 130(2), 149-168.

Honnibal, M., & Montani, I. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

Kidd, D. C., & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*, *342*(6156), 377-380.

Kim, S. (2015). ppcor: an R package for a fast calculation to semi-partial correlation coefficients. Communications for statistical applications and methods, 22(6), 665.

Kintsch, W. (1994). Text comprehension, memory, and learning. American psychologist, 49(4), 294.

Kintsch, W., & Greene, E. (1978). The role of culture-specific schemata in the comprehension and recall of stories. *Discourse processes*, *1*(1), 1-13.

Kintsch, W., & Van Dijk, T. A. (1975). Comment on se rappelle et on résume des histoires. *Langages*, (40), 98-116.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive psychology*, *9*(1), 111-151.

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on psychological science*, *3*(3), 173-192.

Mar, R. A., Tackett, J. L., & Moore, C. (2010). Exposure to media and theory-of-mind development in preschoolers. *Cognitive Development*, *25*(1), 69-78.

Norenzayan, A., Atran, S., Faulkner, J., & Schaller, M. (2006). Memory and mystery: The cultural selection of minimally counterintuitive narratives. *Cognitive science*, 30(3), 531-553.

Piaget J (1926) Language and thought of the child. New York: Harcourt Brace

Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22-27.

Pratchett, T. (1991). Witches Abroad. *London: Corgi*.

Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M., & Dodds, P. S. (2016). The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, *5*(1), 31.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature reviews neuroscience*, *8*(9), 657-661.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2015). Package ‘afex’.

Tamir, D. I., Bricker, A. B., Dodell-Feder, D., & Mitchell, J. P. (2016). Reading fiction and reading minds: the role of simulation in the default network. *Social cognitive and affective neuroscience*, *11*(2), 215-224.

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in* *cognitive sciences*, *22*(3), 201-212.

Thompson, S. (1955). Motif-Index of Folk-Literature, Volume 4: A Classification of Narrative Elements in Folk Tales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jest-Books, and Local Legends (Vol. 4). Indiana University Press.

Thornton, M. A., & Tamir, D. (2019). Six dimensions describe action understanding: the ACT-FASTaxonomy.

Uther, H. J. (2004). The types of International Folktales-A classification and bibliography.

Zunshine, L. (2006). Why we read fiction. *Skeptical Inquirer*, *30*(6), 29.