AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2022, Vol. 151, No. 8, 1833–1842
https://doi.org/10.1037/xge0001171

The Speed of Stories: Semantic Progression and Narrative Success

Henrique Laurino Dos Santos and Jonah Berger
The Wharton School, University of Pennsylvania

Why are some narratives more successful? Although this question has ancient roots, studying it empirically has been challenging. We suggest that semantic progression (i.e., semantic similarity between adjoining portions of a narrative) might shape audience responses but that this role changes over the course of a narrative. Specifically, although slower semantic progression (i.e., greater semantic similarity between adjoining portions) is beneficial at the beginning of narratives, faster semantic progression is beneficial toward the end. To test this possibility, we used natural language processing and machine learning to analyze over 40,000 movie scripts. Consistent with our theorizing, deep-learning-based embeddings find that movies with slower semantic progression early and faster semantic progression later are evaluated more positively. Analysis of over 10,000 TV episodes finds similar results. Overall, these findings shed light on what makes narratives engaging, deepen understanding of what drives cultural success, and underscore the value of emerging computational approaches to understand human behavior.

Keywords: cultural success, natural language processing, narratives, semantic progression

Supplemental materials: https://doi.org/10.1037/xge0001171.supp

Narratives are an integral part of everyday life. Early humans shared epic poems and stories of the hunt, and today we read books, watch movies, and consume content online.

But why are some narratives more successful than others?

Everyone from psychologists and philosophers to linguists and narrative theorists have long been interested in quantifying properties of narratives (e.g., Carroll, 1990; Cutting, 2016; Pennebaker, 2018; Propp, 1968). In *Poetics* (trans. Aristotle & Heath, 1996), Aristotle proposed that stories tend to have common structures, and Freytag (1900) later suggested a dramatic arc or pyramid of rising and falling action. Campbell (1949) theorized that there was an archetypical hero's journey, and Kurt Vonnegut's rejected master's thesis (Vonnegut, 2010) argued that stories could be divided into eight common shapes. Recent work has begun to address this topic empirically, examining variation in sentiment (Reagan et al., 2016) and categories of language (Boyd et al., 2020).

But while research is starting to identify patterns of language across narratives, there has been less attention to whether and how particular narrative features might shape their success. Popular perspectives argue that certain ways of writing can make narratives more successful (e.g., Coyne, 2015; McKee, 1997; Snyder, 2005), for example, but the little empirical work that has been done found no relationship between story structures and popularity (Boyd et al., 2020).

Building on research on semantic similarity, however, we suggest that the pace or speed of a narrative's semantic progression might shape audience response. To test this possibility, we used natural language processing and machine learning to analyze over 40,000 movie scripts. The results shed light on narrative engagement, what drives cultural success (Berger & Heath, 2005; Berger & Milkman 2012; Berger & Packard 2018; Kashima, 2008; Schaller & Crandall, 2004), and the value of computational approaches to understanding human behavior.

The Speed of Semantic Progression

Textbooks, books, and other texts can be described by their cohesion (Graesser et al., 1994, 2011), or how semantically related chunks (e.g., adjoining paragraphs) are to one another (Foltz, 2007). Compared with a paragraph about pine trees, for example, a paragraph about another type of tree should be more semantically similar, or related, than a paragraph about a bank robbery. Along these lines, research finds that adjoining paragraphs of textbooks are more semantically related than paragraphs that are further away and adjoining paragraphs within a chapter are more similar than those on either side of a chapter break (Foltz et al., 1998; see Foltz, 2007 for a review). Further, averaging across all adjoining chunks of a text provides insight into the how easy or difficult it should be for a reader to consume that text. More difficult textbooks, for example, tend to involve larger semantic jumps (i.e., lower similarity or greater distance) between adjoining chunks of text (Foltz et al., 1998).

Whereas semantic relatedness is often applied in the context of education and learning (e.g., how coherence impacts comprehension), we suggest that the same ideas might aid in understanding the evaluation of narratives. Objects that cover a greater distance in

This article was published Online First July 4, 2022.

All the authors contributed equally and are listed in reverse alphabetical order.

Correspondence concerning this article should be addressed to Jonah Berger, The Wharton School, University of Pennsylvania, 700 JMHH 3730 Walnut Street, Philadelphia, PA 19104, United States. Email: jberger@wharton.upenn.edu

the same amount of time can be described as moving faster. Using the notion of semantic relatedness, the same can be said for narratives. Some narratives move more slowly, dwelling on semantically related concepts for longer periods, whereas others move more quickly, jumping between content that is less semantically related.

Consequently, semantic progression can be connected to the notion of pacing, or the speed at which a narrative is told (Hume, 2005; Turco, 1999). Although speed might also conjure up the amount of action, or how quickly things cut between scenes, here we focus specifically on speed at it relates to *semantic* progression, or the speed at which the *content* of the narratives unfolds. The speed of semantic progression can be defined as how quickly the content of discourse (e.g., a narrative) moves between adjoining chunks. By comparing equally long chunks of text across narratives (e.g., 250-word blocks), narratives whose adjoining chunks are more semantically related can be described as moving more slowly, whereas narratives whose adjoining chunks are less semantically related can be described as moving faster.

But which should be liked more: slower semantic progression or faster?

For ideas to make sense, they must be at least somewhat coherent. Slower semantic progression should require less cognitive work to follow (Monahan et al., 2000) and make it easier for people to track what is happening (McNamara et al., 1996). That said, faster progression juxtaposes different concepts which could increase surprise and stimulation and make narratives more engaging (Gergen & Gergen, 1986). So, which is better?

Speed Within a Narrative

We suggest that the answer may depend on the part of the narrative that is being considered. Although one could argue that faster semantic progression itself is either good or bad (i.e., people like faster or slower paced stories overall), we suggest that the speed of semantic progression within a narrative should also impact responses.

At the beginning of a book, movie, or any other narrative, the canvas is blank. The audience doesn't know anything about the characters or context. Consequently, the beginning must set the stage (Cutting, 2016). It must outline these details, and do so in a way that builds a base, or jumping off point, for the rest of the narrative (MacEwan, 1900). Indeed, descriptions of people, places, and things peak at the beginning of stories (McClure & Enderle, 2018), as do prepositions and articles (Boyd et al., 2020), which helps the audience understand what is going on (Morrow, 1990). A story might start by talking about "the house at the end of the road," for example, but once the reader becomes familiar with the context, simply refer to the house as *it*.

Consequently, we suggest that early on, slower semantic progression (i.e., greater similarity between adjoining narrative chunks) might boost success. Like teammates in a relay race, if the second runner starts slowly enough, the first runner can still catch them and hand off the baton. But if that second runner starts at full speed, the first person will never be able to catch up. The same might be true in narratives. If they more too fast, too early on, the audience might get lost.

A variety of literatures support this prediction. Children's stories, jokes, and even music often begin by repeating a similar concept (Loewenstein & Heath, 2009; Rozin et al., 2006). In the Three Little Pigs, for example, one pig builds a straw house, the wolf blows it down, and something very similar happens to a second pig (i.e., he builds a house of sticks, and the wolf blows it down). In jokes, a certain thing happens when a priest walks into a bar, and a similar thing then happens to a nun. This semantic similarity grounds the audience and helps build expectations. Similarly, developmental psychology research suggests that when trying to learn something, children often prefer seeing the same stimulus again and again so they can deepen their understanding. Exposure to the same, or similar content, is easier to process and requires less cognitive effort (Monahan et al., 2000; Zajonc, 2001), which might be particularly important at the beginning of a narrative when an audience knows little about the world being created and the characters in it.

Although slower might boost success early on, this might shift as narratives advance. Once the audience has met the characters and understands the context, the plot must progress (MacEwan, 1900). Relationships must develop, things must happen, and challenges must be overcome. Indeed, measures of plot progression (e.g., pronouns and auxiliary verbs) start small but pick up later in narratives (Boyd et al., 2020).

Consequently, we suggest that toward the end of the narrative, faster pacing (i.e., lower semantic similarity between adjoining chunks) might be beneficial. Consistent with this notion, whereas jokes and folk tales start with repetition to form expectations, these expectations are then broken by the final contrasting event (e.g., but when the wolf tied to blow down the house of bricks . . .), facilitating surprise and engagement (Loewenstein et al., 2011; Loewenstein & Heath, 2009; Rozin et al., 2006).[1] Treading the same ground again and again can get boring and once people understand something, they often want to move on to something else (Flavell et al., 2001). Similarly, work on comprehension and learning finds that textual coherence helps novices, but lower coherence might be better for more knowledgeable readers because they already know enough to fill in the blanks (McNamara et al., 1996).

The Current Work

In sum, we suggest that speed of semantic progression might have different effects on narrative evaluations depending on the point in the narrative. Similar things should be easier to process and require less cognitive effort, and such ease of processing can increase liking and evaluation (Alter & Oppenheimer, 2009). Consequently, early on, when characters, setting and everything else are still novel, narratives should be evaluated more positively when the speed of semantic progression is slower (i.e., greater semantic similarity between adjoining chunks). Slower speeds should help the audience understand what is going on, which should increase engagement and evaluations.

Toward the end of the narrative, however, the opposite should occur. Once the audience is already familiar with the characters and setting, they have the cognitive structures and schemas in place that should make it easier to incorporate novel information. More novel things should be more stimulating (Flavell et al., 2001) and thus faster

[1] Note that though our work is certainly related to the structures described by Rozin et al. (2006) and Loewenstein and Heath (2009), there are some important differences. Although they focus just on short content (i.e., jokes and ads), we consider content that is much longer and where similarity is more complex. Although they dichotomize similarity (i.e., either things are similar or dissimilar), we allow similarity to be continuous and develop a method for empirically quantifying similarity.

semantic progression (i.e., lower semantic similarity between adjoining chunks) might be evaluated more positively.

Said another way, there is a tension between novelty and familiarity. Being exposed to the same thing repeatedly gets boring, and novelty provides stimulation which can increase evaluations. At the same time though, if something is too novel, it might be difficult to understand. Consequently, whereas a blend of novelty and familiarity is often beneficial (Berger et al., 2012), which aspect is valued more should depend on position in the narrative, At the beginning of a narrative everything is new, so there is less need to provide additional stimulation, and greater semantic similarity should facilitate understanding. At the end of a narrative, however, the characters and setting are already laid out, and so greater semantic leaps might provide beneficial stimulation.

To test whether the speed of semantic progression has different effects at different parts of a narrative, we use Natural Language Processing. We embed the dialogue of more than 40,000 movies into vectors that characterize semantic content. Then, we analyze the relationship between adjoining chunks of narrative and how it relates to cultural success. In addition to providing multiple controls and robustness checks, to test the generalizability of the effects, we also examine whether they hold among 10,000 TV episodes.

Note that we focus on semantic progression rather than specific details of the content. Two narratives might have completely different characters, setting, and plot, for example, but their semantic progression might be quite similar (i.e., slow pacing initially and faster at the end). We test whether such speed of semantic progression might impact audience response.

Empirical Analysis of Over 40,000 Movies

Data

To collect data on the semantic content of narratives, we used the English OpenSubtitles2018 corpus of OPUS (https://opus.nlpl.eu/OpenSubtitles-v2018.php; Tiedemann & Lison, 2016). It contains the dialogue, or words actors say throughout a movie's script. The text files were cleaned to remove irrelevant information (e.g., metadata on what software was used to encode subtitles or opening credits).

To capture audience responses, we collected movie ratings from IMDb.com. Each movie is reviewed by multiple people on a scale ranging from 1 to 10, and IMDb reports the mean score. We focused on popular opinion, rather than critics ratings, because we are interested in general response. The semantic progression of the movie should influence audiences' evaluations of it, and online ratings provide a measure of such evaluations

To control for other movie features that might impact success, we collected metadata like runtime, production year, and movie genre (e.g., action or comedy; see Table S1 in the online supplemental material for prevalence of different genres). To focus on movies, analysis focused on content that was at least 30 min long and contained at least 2,500 words. This resulted in a dataset of 42,472 unique movies.

Computing Speed of Semantic Progression

First, we split each movie into chunks, where each chunk (except the last one) has the same number of words. There is no obvious right answer for how long chunks should be, so we relied on prior work as well as the nature of the data. Most movies range from between 3,000 and 20,000 words, and given the need to break them up into at least five parts (Freytag, 1900), and have at least three chunks per part (to compute average semantic progression), we used 250-word blocks for our main analysis. Results are the same, however, for larger and smaller blocks (i.e., 150 and 350 words; see the Robustness and Alternative Explanations section).

Second, we determine the semantic similarity of adjoining chunks. Early work in this area used latent semantic analysis (see Foltz, 2007 for a review), but recent advances in computer science provide even more advanced approaches. Unsupervised embedders take words, sentences, or even whole documents, and represent them as vectors in a high-dimensional space. Word embeddings, for example, are based on Firth's (1957, p. 11) suggestion that "you shall know a word by the company it keeps." If two words are often surrounded by similar words, they probably have very similar meanings. Embedding algorithms "organize" words in a multidimensional space, where each word receives a numerical vector, and words with similar meanings or uses are closer together (Bhatia, 2017). Extensions of these algorithms allow larger chunks of text (i.e., paragraphs or whole documents) to be transformed into numerical vectors with the same interpretation.

Most sentence or paragraph-level embedders use a common approach to word-level embeddings but differ in how they combine them to get a vector for the whole chunk. Given its performance in human scoring benchmarks, we use Google's Universal Sentence Encoder (Cer et al., 2018), but the results are robust to other approaches (i.e., doc2vec; see the Robustness and Alternative Explanations section). The Universal Sentence Encoder offers a few different deep-learning models pretrained on a variety of large corpi (e.g., English-language Wikipedia and online reviews); and for its simplicity, speed, and robust results, we chose Google's pretrained Deep Averaging Network (DAN; Iyyer et al., 2015). DAN first embeds individual words in a text (as word2vec), then takes their weighted average for classification and loss function computation (see the online supplemental material for more detail).

Following prior work (e.g., Foltz et al., 1998) we measure the semantic similarity of adjoining blocks of text using cosine similarity (also see Bhatia, 2017; DeFranza et al., 2020). This provides marginal advantages over euclidean distance with added resilience against a few degenerate cases sometimes seen in text data (though both approaches lead to similar results). It is defined as follows:

$$cos\theta = \frac{\vec{a} \cdot \vec{b}}{||\vec{a}||\,||\vec{b}||}.$$

This yields similarity values between −1 and 1, where two identical vectors would have similarity 1, and larger cosine similarity values indicate greater semantic similarity (as they do in Latent semantic analysis).

Prior work (Bhatia, 2017; Garg et al., 2018; Kozlowski et al., 2019), as well as a series of validation tests (see below) demonstrate that this measure captures semantic similarity as well as human perceptions of similarity. Within the same movie, for example, adjoining blocks are, on average, scored as more semantically similar than nearby blocks, which are scored as more similar than blocks that are further away, which are scored as more similar than blocks from a completely different movie.

Third, by averaging across pairs of adjoining chunks, we calculate the average speed of semantic progression for different parts of each movie. A great deal of prior work suggests that movies follow a five-act structure (Freytag, 1900), so the main analysis break movies up into five parts, but results are robust to larger or smaller numbers as well (see the Robustness and Alternative Explanations section). For each part (i.e., one fifth of the movie), we calculate the average similarity between adjacent chunks (i.e., Chunk 1 and Chunk 2, Chunk 2 and Chunk 3, and so on) within that part. Taking the opposite of this captures speed of semantic progression. If two chunks are more semantically dissimilar (i.e., lower cosine similarity between them), it means that greater semantic distance was covered in the same amount of time, and thus indicates a faster speed of semantic progression.

Finally, linear regression predicted average movie rating based on the average speed of semantic progression of each part. See Table S2 in the online supplemental material for a correlation matrix of the variables.

Results

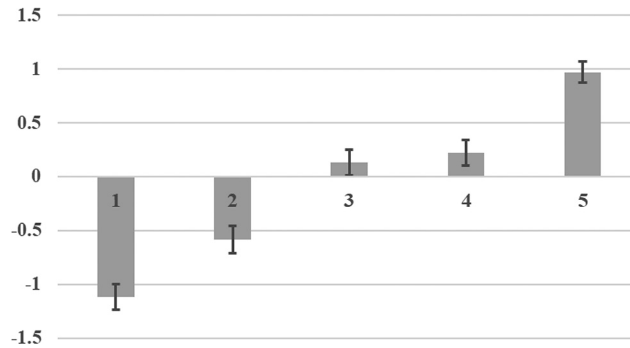
Results indicate that while movies with faster semantic progression early on (i.e., in the first part) are evaluated less positively, movies with faster semantic progression at the end (i.e., last part) are evaluated *more* positively (see Table 1, Model 1). Speed of semantic progression initially has a negative relationship with evaluations (i.e., greater speed or more distance between adjoining chunks is detrimental), but this relationship increases through the rest of the movie, eventually becoming significantly positive by the end (i.e., greater speed or more distance between adjoining chunks is linked to higher ratings; see Figure 1).

Table 1
Speed of Semantic Progression and Audience Response

Effect	Part	(1) Base	(2) Plus Controls
Speed of semantic progression	Part 1	−1.113***	−0.230*
		(0.12)	(0.09)
	Part 2	−0.583***	−0.116
		(0.13)	(0.10)
	Part 3	0.133	0.054
		(0.12)	(0.10)
	Part 4	0.225	−0.065
		(0.12)	(0.09)
	Part 5	0.972***	0.309***
		(0.10)	(0.08)
Controls			
Year			−0.017***
			(0.00)
Runtime			0.007***
			(0.00)
Genre		No	Yes
Budget			−0.024***
			(0.00)
Rating			0.207***
			(0.00)
R^2		0.004	0.402
Bayesian information criterion		129,400	113,900
Observations		42,472	42,472

Note. Values represent the relationship between each feature and movie ratings.
* $p < .05$. *** $p < .001$.

Figure 1
Audience Response Based on Speed of Semantic Progression Across Movie Parts



Note. Results based on raw coefficients and standard errors from Model 1.

While one could wonder whether these effects are somehow driven by the modeling approach used, estimates for each movie part (i.e., one fifth of the narrative) are independent, so no functional form is being forced on the results. Instead, the fact that estimates are negative initially, but reduce and become positive at the end is consistent with the notion that slower pacing boosts success at the beginning of movies, whereas faster pacing helps toward the end.

Controls

These initial results are consistent with our theorizing, but one could wonder whether they are driven by other factors. Consequently, we include various control variables (i.e., year of release, runtime, genre, budget, and number of ratings) to test alternative explanations and robustness.

First, one might wonder whether release year is somehow driving the results. Maybe more recent movies are both more highly rated, for example, and use certain plot structures. Although it is unclear how such an explanation would explain variation in semantic progression within movies, to control for this possibility, we control for *Year of Release*. Using continuous form or dummy variables yields similar results, so for simplicity the main results include a continuous version.

Second, one could wonder whether longer movies have faster or slower semantic progression toward the end and might also receive differential ratings. Consequently, we control for *runtime* in minutes.

Third, one could wonder whether certain genres of movies might tend to receive higher ratings and have different patterns of semantic progression. Consequently, we use dummy variables to control for IMDb's *genre* tags (e.g., action, comedy, or horror). Some movies were tagged with multiple genres, and thus count in multiple groups.

Fourth, one could wonder whether blockbuster type movies tend to use certain plot structures and receive higher or lower ratings.

We control for this possibility in the following two ways: by *budget* and by *number of ratings* for each movie, controlling for the log of each because it gives a more normal distribution of values.

Even controlling for all these factors, however, the relationships between speed of semantic progression and success persist (see Table 1, Model 2). Movies with faster semantic progression early on are evaluated less positively, but movies with faster semantic progression toward the end are evaluated *more* positively

Robustness and Alternative Explanations

We also test robustness in a number of other ways, including (1) the blocking approach, (2) number of parts, (3) embedding approach, (4) modeling approach, (5) blocks per part, (6) mean similarity across parts, (7) removing stop words, (8) comparing only movies of similar lengths, and (9) production side factors. Across all these different specifications, the results still persist.

First, one could wonder whether the results are somehow driven by the blocking approach used. The main model used 250-word blocks, but to test this possibility, we also examine 150- and 350-word blocks. Results remain the same (see Table 2). Movies with faster semantic progression early on are evaluated less positively, but movies with faster semantic progression toward the end are evaluated *more* positively.

Second, one could wonder whether the results are somehow driven by the number of parts or acts used. Following prior work (Freytag, 1900), the main model used a five-act structure, but some work has suggested seven-act structures and, though less common, even four- or six-act structures. To test the robustness of the effect, we also test a four-, six-, and seven-part structure (i.e., analyzing the speed of semantic progression within one quarter of each movie, in sixths or sevenths). Results remain the same (see Table S3 in the online supplemental material).

Third, one could wonder whether the results are somehow driven by the embedding approach used. Our main analyses relied on USE-DAN because it well captures semantic similarity while being simpler, faster, and more robust than alternative methods, but to test robustness, we use doc2vec (Le & Mikolov, 2014). We used a pretrained doc2vec model (Lau & Baldwin, 2016) to calculate the position of each movie chunk and reran our main analyses. Results remain the same (see Table 2).

Fourth, one could wonder whether the results are somehow driven by treating each part (i.e., one fifth of the movie) as independent. Maybe audiences value high similarity at the beginning of a narrative *or* low similarity at the end, for example, but do not care whether both are present. There might also be other interactions that are stronger than the effects of individual acts or parts. To test this possibility, we cluster movies based on their shapes and examine whether movies with certain shapes are liked more (see the online supplemental material). Results remain the same.

Along those lines, one could wonder whether it would make more sense to look at change in narrative speed over the entire movie. To examine this alternate measure, we take the slope of the line between act number and average speed in the act for each movie. A positive slope indicates that the speed of semantic progression is generally increasing across the movie. Consistent with the notion that people like slower speed early and faster speed late, results indicate that movies are liked more when the speed of semantic progression increases more over the course of the movie (coefficient = 1.15, SE = .29, $p < .001$).

Fifth, one could wonder whether the method of dividing up the text could be driving the results. The results are robust to different blocking sizes, but the number of blocks or chunk per part can also vary. Although a 10,000-word movie would be broken up into five parts (i.e., acts) of 2,000 words, each of which composed of eight 250-word chunks, other lengths are slightly more uneven. A 11,000-word movie, for example, has forty-four 250-word chunks, which means that whereas the first four parts would each have nine chunks, the final part would have only eight. To test whether slightly differing numbers of chunks for later parts or acts in some cases could somehow drive the results, we removed the last chunk of parts with higher counts (so all parts have an equal number) and reran the model. Results remain the same (see Table 2).

Sixth, one could wonder whether the mean level of semantic progression across parts could be driving the results. Maybe audiences always like the same speed of semantic progression throughout a

Table 2
Robustness Tests

Effect	Part	Blocking Approach 150	Blocking Approach 350	Word2Vec Embedding	Removing Excess Chunks
Speed of semantic progression	Part 1	−0.205* (0.09)	−0.159* (0.09)	−0.706*** (0.10)	−0.294* (0.09)
	Part 2	−0.207 (0.11)	−0.150 (0.09)	−0.281* (0.11)	−0.176 (0.09)
	Part 3	0.070 (0.11)	−0.106 (0.09)	0.032 (0.11)	0.017 (0.09)
	Part 4	−0.062 (0.11)	−0.069 (0.09)	0.002 (0.11)	0.019 (0.09)
	Part 5	0.481*** (0.10)	0.262*** (0.07)	0.324*** (0.09)	0.303*** (0.08)
Controls					
Year		−0.017*** (0.00)	−0.017*** (0.00)	−0.017*** (0.00)	−0.017*** (0.00)
Runtime		0.007*** (0.00)	0.007*** (0.00)	0.007*** (0.00)	0.007*** (0.00)
Genre		Yes	Yes	Yes	Yes
Budget		−0.024*** (0.00)	−0.024*** (0.00)	−0.024*** (0.00)	−0.024*** (0.00)
Rating		0.207*** (0.00)	0.207*** (0.00)	0.209*** (0.00)	0.203*** (0.00)
R^2		0.402	0.402	0.403	0.402

Note. Values represent the relationship between each feature and movie ratings.
* $p < .05$. *** $p < .001$.

narrative, but actual speed of semantic progression used tends to be higher at the beginning and lower at the end, leading to the observed relationships (i.e., lower than average is preferred at the beginning and higher than average is preferred at the end). But this is not the case. There is no difference between the mean speed of semantic progression at the beginning and end, $t(43{,}078) = -.93$, $p = .82$. Further the mean level of speed of semantic progression is similar (i.e., within one standard deviation) across the movie. Consequently, variation in actual speed of semantic progression across the movie has trouble explaining the results.

Seventh, one could wonder whether the results could be driven by different types of language emerging across a story. Narratives tend to move from low to high rates of pronouns (Boyd et al., 2020), for example, and one could wonder whether the resulting changes in weighting of background function words is driving the effect. To test this possibility, we rerun the main analysis, but remove stop words (e.g., *I*, *me*, *an*, and *the*) from the text. Although the size of the exact coefficients changes slightly, the results remain the same. Movies are liked more when they move slowly at the beginning and faster at the end. This casts doubt on the possibility that language shifts are driving the effect, but future work might examine the consequences of shifts in language across narratives more generally.

Eighth, one could wonder whether the results were somehow driven by comparing movies of different lengths. This is not the case. Even looking at movies of similar lengths (i.e., between 25 and 35 chunks, which is around the mode of the distribution) results remain the same.

Ninth, one could wonder whether production side factors could somehow be driving the effect. We already controlled for production side measures like budget, but maybe certain screenwriters tend to make "better" movies, for example, and tend to follow certain pacing structures and that is driving things.

But this does not seem to be the case. We identified famous screenwriters that have written several films (i.e., George Lucas, Quentin Tarantino, and Ingmar Bergman), found all the movies in the dataset that they had written, and tested whether movies by the same screenwriter tended to have a consistent pattern of semantic progression. They did not. Some of George Lucas' movies, for example, increased in semantic progression across the course of the narrative, whereas others decreased. Even Quentin Tarantino, a writer known for his distinctive style, has considerable variety in his semantic progression (see Figure S1 in the online supplemental material).

This variation also casts doubt on the possibility that these findings are already known in the industry. If faster pacing being helpful late but detrimental early was already known, one would expect famous screenwriters to use this approach. But they do not. Further, one could argue that even if screenwriters do not know it, studios do, and so the invest more money in films that follow that structure. But as shown in Figure S1 in the online supplemental material, not all famous movies follow this structure, casting doubt on that possibility as well.

Validation

Although the results are consistent with our theorizing, one could wonder whether the measure of semantic progression is truly capturing the similarity between content. A great deal of prior work documents that embeddings distance captures semantic similarity (e.g., Bhatia, 2017; Garg et al., 2018; Kozlowski et al., 2019), but to underscore this point, we provide a few additional validation tests in our data (see the online supplemental material for additional comparison with the Semantic Textual Similarity benchmark).

Semantic Similarity Between Different Chunks

If our measure is truly picking up semantic similarity, as we suggest, then following Foltz et al. (1998), adjoining chunks of a movie should be scored as more semantically similar than chunks that are further away. Similarly, chunks of one movie should be scored as more semantically similar to chunks of that same movie than chunks of a completely different movie.

To test these possibilities, we draw 9,000 target chunks randomly (without replacement) across all movies. For each target chunk, we randomly select four comparison chunks: an adjacent chunk, a chunk five positions away from the initial chunk, a chunk five to 10 positions away, and a chunk from another movie. Then, we compute the distance between the target chunk and each of these four types of chunks and average them for each type across all 9,000 target chunks.

Results support the notion that our measure captures semantic similarity. As predicted, adjacent chunks from the same movie are more semantically similar, on average, than chunks from the same movie within five positions, which are more semantically similar than chunks selected anywhere from the same movie, which are more semantically similar than chunks from different movies ($M_{\text{Adjacent}} = .52 < M_{\text{5Away}} = .47 < M_{\text{5-10Away}} = .44 < M_{\text{Diff Movie}} = .35$; $ts > 22$, $ps < .0001$).

Relation to Human Perceptions

To provide further evidence that this measure captures semantic similarity, we also tested whether it is related to perceived similarity. We randomly picked 50 movies, randomly picked a target chunk from each, and then picked two comparison chunks. To do so, we computed the distance between the target chunk and all other chunks in the same movie, and randomly picked a chunk from the bottom quartile of similarity and a chunk from the top. Then, for each target chunk, three hypothesis-blind research assistants coded which comparison chunk was more similar. Majority rule determined which chunk humans judged as closer.

Results indicate that our measure of semantic similarity was reasonably related with human perceptions of similarity. Human judgment agreed with word embeddings 74% of the time. As a baseline, the joint probability of agreement between judges (i.e., the probability that evaluations from two judges on the same triplet is the same) was 72%. This indicates that our semantic similarity measure is reasonably consistent with human similarity perception and similarly reliable.

Ancillary Test: Empirical Analysis of Over 10,000 TV Episodes

Results of the main study are consistent with our theorizing, but one could wonder whether the findings are replicable. Even though we included multiple controls, and robustness checks, maybe there is something unique to movies that is driving the effect. Consequently, to further test robustness, and generalizability, we conduct an ancillary test in another domain. Specifically, we collect over ten thousand episodes from over 300 TV shows. We test whether

TV shows are liked more when the speed of semantic progression is slower at the beginning but faster at the end.

A major global media company provided a dataset of closed captions for English language TV shows. This included everything from multiseason thrillers to episodic children's cartoons. To capture audience response, as in the main study, we used IMDb.com to collect ratings for each episode of each TV show. To control for other features which might impact success, we collected metadata at both the show level, such as the channel it aired on and genre (e.g., *comedy* and *crime*), and episode level, such as the season of the episode and episode number within the season. Episode-level information allows us to control for the possibility that later episodes of a show are liked more (or less). This resulted in a dataset of 10,578 unique episodes from 290 TV shows.

As noted in the main study, computing average semantic progression requires at least three chunks per part of the narrative. TV shows are much shorter than movies (M = 4,321 words vs. M = 8,827), so breaking episodes into five parts and using 250-word chunks of would mean that many episodes would have fewer than three chunks a part. We address this in three ways. First, the main analysis uses 150-word chunks and three parts. This ensures that most narratives have a similar number of chunks per part as movies. Second, to ensure that the results aren't somehow driven by the chunk size, we also examine 250-word chunks. Third, to ensure that the results aren't somehow driven by the number of parts, we also examine what happens when more parts are used.

Across all three versions, results are similar to those found for movies (see Table 3). While episodes with faster semantic progression early on are evaluated *less* positively, episodes with faster semantic progression at the end are evaluated *more* positively. In the main analysis, for example, speed of semantic progression initially has a negative relationship with evaluations, but this relationship increases through the rest of the show, eventually becoming significantly positive by the end (Model 1). Results are the same using larger chunks (Model 2) or dividing shows into more parts (Model 3).

Analysis of over 10,000 TV episodes underscore the findings of the main study. Although TV episodes whose early parts are more semantically similar were evaluated more positively, shows whose later parts are more semantically similar were evaluated less positively. Finding the same results across a different domain speaks to the generalizability of the effect.

Discussion

Academics and practitioners alike have long been interested in what makes narratives engaging. From Aristotle's early ideas about the trajectories of tragedies and comedies, to more recent theories of ways to make screenplays, scripts and stories more popular (e.g., McKee, 1997), different perspectives have made different propositions. But while these ideas have captured the imaginations of everyone from literary theorists to so called script-doctors, actual empirical tests have been few and far between. Consequently, less is known about what makes some narratives more successful.

Machine Learning based Natural Language Processing of over 40,000 movies sheds light on this question. Whereas movies whose early parts are more semantically similar are evaluated more positively, movies whose later parts are more semantically similar are evaluated *less* positively. Ancillary analysis of over 10,000 TV episodes finds similar results. Taken together, the findings suggest that whereas slower semantic progression might be beneficial at the beginning of narratives, faster semantic progression is beneficial toward the end.

Although we focused on movies and tv show due to data availability, similar approaches could be applied to any type of narrative (e.g., plays or books). One important moderator, however, might be the way narratives are consumed. Whereas movies and plays tend to be consumed in one sitting, for example, books tend to be read over multiple occasions. Consequently, the impact of pacing might vary. In books, pacing *within* chapters might be important, as starting slowly might helping readers remember what happened in the last chapter. Research could also examine how the speed of semantic progression within a narrative shapes comprehension and memory.

Future work might also more deeply examine the psychological mechanisms behind these effects. As noted, comprehension, processing ease, and stimulation might all play a role. At the beginning of a narrative, much is already quite novel, and so greater semantic similarity (i.e., slower semantic progression) might help the audience understand what is going on and make the content easier to process, bother of which should increase evaluations. At the same time, at the end of the narrative, the characters and settings are no longer new, so providing greater surprise and stimulation through faster semantic progression might boost evaluations.

Work might also examine how speed of semantic progression shapes other relevant outcomes, like comprehension and memory. Does faster speed of semantic progression, for example, decrease understanding and memory for what occurred? This might be particularly relevant for educational texts.

One could also examine whether the effects extend to other types of discourse. If these effects are specific to narratives, things like product manuals or how-to guides should show different patterns. On the other hand, if slower pacing early, and faster pacing later, make any type of content easier to follow, such semantic

Table 3
Speed of Semantic Progression in TV Episodes

Effect	Part	Main model	Larger chunks	More parts
Speed of semantic progression	1	−0.698*** (0.21)	−0.852*** (0.19)	−0.402* (0.20)
	2	0.237 (0.21)	0.305† (0.18)	0.045 (0.20)
	3	0.542** (0.20)	0.263† (0.15)	−0.200 (0.20)
	4			0.655*** (0.17)
R^2		0.204	0.208	0.204
Observations		10,578	10,578	10,578

Note. Values represent the relationship between each feature and TV episode ratings.
†$p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

progression might be beneficial in a broader set of content. Even "non-narratives" like product manuals often follow a building structure where later points might build on earlier ones, and thus semantic progression might play a role.

Work might also examine other narrative features. People often talk about narratives as covering a lot of ground, for example, or going in circles. Research could try to quantify these aspects, looking at ground covered (i.e., the space covered by all the chunks of the narrative) or the degree to which it repeatedly returns to common themes and ideas (Toubia et al., 2021). Two movies might both be fast paced, for example, but one might go back and forth between the same ideas, whereas the other moves on to new topics.

Similar ideas could also be applied to personal narratives. People often use narratives to explain and understand their own lives (McAdams & McLean, 2013). Just as creative people have more distance (i.e., less semantic relatedness) between their thoughts (Gray et al., 2019), faster semantic progression in personal narratives might provide insight into the writer's personality, or even how the act of writing might impact well-being (Pennebaker, 2018).

Importantly, we do not mean to suggest that the script captures everything in a movie. Audio-visual elements such as music and cinematography certainly play an important role. But capturing these aspects, and linking them to audience reactions, is far from trivial. We utilized the largest available movies script database, but there is much less moment-to-moment data available on cinematography or audio-visual aspects. Further, even if one could eventually construct such a dataset, quantification might be challenging. Less is known about how to measure a movie's visual evolution of a movie (cf. Cutting, 2016), or how to quantify its soundtrack. Hopefully future work can examine these questions in greater detail.

These other aspects also help put the size of the observed effect in context. Unlike laboratory experiments, where everything else is carefully controlled, we focused on real-world content evaluations in the noisy field. Thousands of things likely impact such evaluations (e.g., cinematography, music, the actors, acting quality, and the director), and given these other contributing factors, it would be quite surprising if speed of semantic progression alone explained a huge portion of the variance. After all, a movie could have no name actors, a terrible plot, bad acting and directing, and still move slowly at the beginning and fast at the end. That said, the fact that an effect of semantic progression emerges even with everything involved speaks to its persistence. Further, the fact that TV episodes show similar results speaks to the effect's generalizability.

We also do not mean to suggest that semantic progression is the only element that contributes to narrative success. The experience of being transported by a story, use of rhetorical devices, and invoking emotion should all also play a role (e.g., Berger et al., 2021; Green & Brock, 2000). In addition to these, and potentially other aspects, however, we suggest that the overall narrative structure, and speed of semantic progression is important to consider.

More generally, these results speak to the role of psychological processes in cultural success. Just as sociocultural background shapes psychological processes (Markus & Kitayama, 1991), the reverse is also true; When shared across individuals, psychological processes can act as a selection mechanism, shaping whether cultural items succeed or fail (Berger & Heath, 2005; Berger & Milkman 2012; Berger & Packard 2018; Kashima, 2008; Schaller & Crandall, 2004). Just as minimally counterintuitive narratives are more memorable and popular (Norenzayan et al., 2006), in this case, the speed of semantic progression within a narrative might shape its success.

Finally, these findings highlight the value of Natural Language Processing and emerging computational approaches to study human behavior (see Berger & Packard, 2022 for a recent review). Language in an integral part of everyday life, and a wealth of related data from social media posts and online reviews to song lyrics, and movie scripts is now available. By quantifying features of textual data, natural language processing can unlock a range of interesting questions.

Context

We have recently been using natural language processing to extract behavioral insight from textual data. Some of this work has focused on cultural success. Why are some songs, books, movies, and even academic papers more successful than others? When shared across individuals, psychological processes can act as a selection mechanism, shaping whether cultural items succeed or fail. Consequently, textual features which shape things like memory, evaluation, or social transmission can impact success. Atypical songs, for example, are more popular (Berger & Packard 2018), as are those that use second person pronouns (because they remind people of close others, Packard & Berger 2020). Here we studied how the speed of semantic progression might shape the evaluation of movies (and tv episodes). The speed of semantic progression should impact how easy information is to process, as well as things like surprise, which together in turn should impact evaluations. These findings shed light on why narratives succeed and fail and how natural language processing can be used to study culture.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Aristotle, & Heath, M. (1996). *Poetics*. Penguin Books.

Berger, J. A., & Heath, C. (2005). Idea habitats: How the prevalence of environmental cues influences the success of ideas. *Cognitive Science*, *29*(2), 195–221. https://doi.org/10.1207/s15516709cog0000_10

Berger, J., Eric, B., Alex, B., & Yao, Z. (2012). From Karen to Katie: Using baby names to study cultural evolution. *Psychological science*, *23*(10), 1067–1073.

Berger, J., Kim, Y., & Meyer, R. (2021). What makes content engaging? How emotional dynamics shape success. *The Journal of Consumer Research*, *48*(2), 235–250. https://doi.org/10.1093/jcr/ucab010

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, *49*(2), 192–205. https://doi.org/10.1509/jmr.10.0353

Berger, J., & Packard, G. (2018). Are atypical things more popular? *Psychological Science*, *29*(7), 1178–1184. https://doi.org/10.1177/0956797618759465

Berger, J., & Packard, G. (2022). Using natural language processing to understand people and culture. *American Psychologist*. Advance online publication. https://doi.org/10.1037/amp0000882

Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, *124*(1), 1–20. https://doi.org/10.1037/rev0000047

Boyd, R. L., Blackburn, K. G., & Pennebaker, J. W. (2020). The narrative arc: Revealing core narrative structures through text analysis. *Science Advances*, *6*(32), Article eaba2196. https://doi.org/10.1126/sciadv.aba2196

Campbell, J. (1949). *The hero with a thousand faces*. Pantheon Books.

Carroll, N. (1990). *The philosophy of horror*. Routledge.

Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). Universal sentence encoder for english.

Proceedings of the 2018 conference on empirical methods in natural language processing: System demonstrations (pp. 169–174). Association for Computational Linguistics.

Coyne, C. (2015). Harriet Wolf's seventh book of wonders. *Library Journal*, *140*(15), 65.

Cutting, J. E. (2016). Narrative theory and the dynamics of popular movies. *Psychonomic Bulletin & Review*, *23*(6), 1713–1743. https://doi.org/10.3758/s13423-016-1051-4

DeFranza, D., Mishra, H., & Mishra, A. (2020). How language shapes prejudice against women: An examination across 45 world languages. *Journal of Personality and Social Psychology*, *119*(1), 7–22. https://doi.org/10.1037/pspa0000188

Firth, J. R. (Ed.) (1957). A synopsis of linguistic theory, 1930–1955. *Studies in linguistic analysis* (pp. 1–31). Blackwell.

Flavell, J. H., Miller, P. H., & Miller, S. A. (2001). *Cognitive development*. Prentice Hall.

Foltz, P. W. (2007). Discourse coherence and LSA. In T. K. Landauer, D. S. McNamara, S. Dennis, W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 167–184. Lawrence Erlbaum Associates Publishers.

Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes*, *25*(2–3), 285–307. https://doi.org/10.1080/01638539809545029

Freytag, G. (1900). *Freytag's technique of the drama: An exposition of dramatic composition and art by Dr. Gustav Freytag* (E. J. MacEwan, Trans.). Scott Foresman and Company.

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Gergen, K. J., & Gergen, M. M. (1986). *Narrative form and the construction of psychological science*. Praeger Publishers/Greenwood Publishing Group.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-metrix: Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, *40*(5), 223–234. https://doi.org/10.3102/0013189X11413260

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371–395. https://doi.org/10.1037/0033-295X.101.3.371

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, *79*(5), 701–721. https://doi.org/10.1037/0022-3514.79.5.701

Gray, K., Anderson, S., Chen, E. E., Kelly, J. M., Christian, M. S., Patrick, J., Huang, L., Kenett, Y. N., & Lewis, K. (2019). "Forward Flow": A new measure to quantify free thought and predict creativity. *The American Psychologist*, *74*(5), 539–554. https://doi.org/10.1037/amp0000391

Hume, K. (2005). Narrative speed in contemporary fiction. *Narrative*, *13*(2), 105–124. https://doi.org/10.1353/nar.2005.0010

Iyyer, M., Manjunatha, V., Boyd-Graber, J., & Daumé, H., III. (2015). Deep unordered composition rivals syntactic methods for text classification. In C. Zong & M. Strube (Eds.), *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (Volume 1: Long Papers)* (pp. 1681–1691). Association for Computational Linguistics. https://aclanthology.org/P15-1162

Kashima, Y. (2008). A social psychology of cultural dynamics: Examining how cultures are formed, maintained, and transformed. *Social and Personality Psychology Compass*, *2*(1), 107–120. https://doi.org/10.1111/j.1751-9004.2007.00063.x

Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, *84*(5), 905–949. https://doi.org/10.1177/0003122419877135

Lau, J., & Baldwin, T. (2016). An empirical evaluation of doc2vec with practical insights into document embedding generation. *Proceedings of the 1st Workshop on Representation Learning for NLP*, (pp. 78–86).

Le, Q. V., & Mikolov, T. (2014). Distributed representations of sentences and documents. *Proceedings of the 31st International Conference on Machine Learning*, *32*(2), 1188–1196. https://proceedings.mlr.press/v32/le14.html

Loewenstein, J., & Heath, C. (2009). The repetition-break plot structure: A ognitive influence on selection in the marketplace of ideas. *Cognitive Science*, *33*(1), 1–19. https://doi.org/10.1111/j.1551-6709.2008.01001.x

Loewenstein, J., Raghunathan, R., & Heath, C. (2011). The repetition-break plot structure makes effective television advertisements. *Journal of Marketing*, *75*(5), 105–119. https://doi.org/10.1509/jmkg.75.5.105

MacEwan, G. (1900). *Freytag's technique of the drama*. Scott Foresman.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

McAdams, D. P., & McLean, K. C. (2013). Narrative identity. *Current Directions in Psychological Science: A Journal of the American Psychological Society*, *22*(3), 233–238. https://doi.org/10.1177/0963721413475622

McKee, R. (1997). *Story: Substance, structure, style, and the principles of screenwriting*. Harper Collins.

McClure, D., & Enderle, S. (2018). Distributions of function words across narrative time in 50, 000 Novels. In *Digital Humanities* (pp. 242–245).

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43. https://doi.org/10.1207/s1532690xci1401_1

Monahan, J. L., Murphy, S. T., & Zajonc, R. B. (2000). Subliminal mere exposure: Specific, general, and diffuse effects. *Psychological Science*, *11*(6), 462–466. https://doi.org/10.1111/1467-9280.00289

Morrow, D. G. (1990). Spatial models, prepositions, and verb-aspect markers. *Discourse Processes*, *13*(4), 441. https://doi.org/10.1080/01638539009544769

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. https://doi.org/10.1207/s15516709cog0000_68

Packard, G., & Berger, J. (2020). Thinking of you: How second-person pronouns shape cultural success. *Psychological Science*, *31*(4), 397–407. https://doi.org/10.1177/0956797620902380

Pennebaker, J. W. (2018). Expressive writing in psychological science. *Perspectives on Psychological Science*, *13*(2), 226–229. https://doi.org/10.1177/1745691617707315

Propp, V. (1968). *Morphology of the folktale*. University of Texas Press. https://doi.org/10.7560/783911

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), 1–12. https://doi.org/10.1140/epjds/s13688-016-0093-1

Rozin, P., Rozin, A., Appel, B., & Wachtel, C. (2006). Documenting and explaining the common AAB pattern in music and humor: Establishing and breaking expectations. *Emotion*, *6*(3), 349–355. https://doi.org/10.1037/1528-3542.6.3.349

Schaller, M., & Crandall, C. S. (2004). *The psychological foundations of culture*. Erlbaum.

Snyder, B. (2005). *Save the Cat: The last book on screenwriting you'll ever need*. Michael Wiese Productions.

Tiedemann, J., & Lison, P. (2016). OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the 10th international conference on language resources and evaluation* (pp. 923–929). European Language Resources Association (ELRA).

Toubia, O., Berger, J., & Eliashberg, J. (2021). How quantifying the shape of stories predicts their success. *Proceedings of the National Academy of Sciences*, *118*(26), e2011695118. https://doi.org/10.1073/pnas.2011695118

Turco, L. (1999). *The book of literary terms: The genres of fiction, drama, nonfiction, literary criticism, and scholarship,* University Press of New England.

Vonnegut, K. (2010). Shapes of stories. https://www.youtube.com/watch?v=oP3c1h8v2ZQ

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science: Journal of the American Psychological Society*, *10*(6), 224–228. https://doi.org/10.1111/1467-8721.00154

Received January 8, 2021
Revision received October 20, 2021
Accepted November 17, 2021 ■