



Genre-typical narrative arcs in films are less appealing to lay audiences and professional film critics

Taleen Nalabandian¹ · Molly E. Ireland¹

Published online: 30 November 2018
© Psychonomic Society, Inc. 2018

Abstract

People tend to like stimuli—ranging from human faces to text—that are prototypical, and thus easily processed. However, recent research has suggested that less typical stimuli may be preferred in creative contexts, such as fine art or music lyrics. In an archival sample of movie scripts, we tested whether genre-typicality predicted film ratings as a function of rater role (novice audience member or expert film critic). Genre-typicality was operationalized as the profile correlations between linguistic arcs (across five segments, or acts) for each script and within-genre averages. We predicted (1) that critics would prefer more disfluent (genre-atypical) films and general audiences would prefer fluent (genre-typical) films, and (2) that these differences would be most pronounced for genres expected to be more entertaining (e.g., action/adventure) than challenging (e.g., tragedy). Partly consistent with our hypotheses, the results showed that critics gave higher ratings to action/adventure films with less typical positive emotion arcs. However, regardless of audience-member or professional-critic status, higher ratings were attributed to films that were more genre-atypical (or disfluent), in terms of analytic thinking, narrative action, and emotional tone, across all genres except family/kids films. Such findings support the growing literature on the appeal of disfluency in the arts and have relevance for researchers in psychology and computer science who are interested in computational linguistic approaches to attitudes, film, and literature.

Keywords Prototypicality · Genre · Film · Language · Computerized text analysis

The narratives of novels, films, and plays both reflect and influence human behavior. Authors directly and indirectly theorize on psychological processes in their work, describing how and why people think, feel, and behave the way they do (Cutting, 2016)—leading some to conclude that authors are often more experienced observers of human behavior than psychologists (Rosenberg, 2013). Fictional narratives not only lend insight into human behavior; in some cases, fiction influences behavior as well. Examples include violent media eliciting aggression (Bushman & Anderson, 2001), fan fiction as a gateway to sexual development (Mixer, 2018), and popular television’s influence on accent change (e.g., the London-based series

EastEnders has been implicated in rising Cockney accents among young Glaswegians; Stuart-Smith, Pryce, Timmins, & Gunter, 2013). More broadly, reading fiction theoretically involves simulating social worlds, which in turn can improve emotion regulation and social skills (Mar & Oatley, 2008). In part because of the impacts that narrative fiction has on individuals and cultures, it is important to examine the process by which people evaluate films. Understanding the narrative preferences of audiences has relevance for more basic research on human cognition, affect, and behavior, as well. Specifically, studying how language patterns in films relate to critics’ and audiences’ preferences adds to the growing literature on how mathematical properties of attitude objects (e.g., the averageness or typicality of human faces and song lyrics) relate to attitudes (Berger & Packard, 2018; Trujillo, Jankowitsch, & Langlois, 2014). In the present study, we attempt to uncover which types of narrative arcs in films individuals tend to enjoy more, employing novel linguistic measures of genre-typicality adapted from earlier computational linguistic research on narratives in novels and short stories (Blackburn, 2015; Malin et al., 2014).

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13428-018-1168-7>) contains supplementary material, which is available to authorized users.

✉ Taleen Nalabandian
taleen.nalabandian@ttu.edu

¹ Texas Tech University, Lubbock, TX, USA

Narrative Arc Theory

Narrative arcs have been studied for millennia. Aristotle (c. 335 BCE/1961) proposed that fictional plots should have three acts—a beginning, middle, and end—that are causally connected and self-contained, imitating how dramatic events unfold in real life. Centuries later, Aristotle's ideal narrative was refined and expanded by Freytag (1894) into a five-arc narrative: exposition, rising action, climax, falling action, and resolution or denouement. On the basis of Freytag's (1894) five-arc narrative, Blackburn (2015) and Malin et al. (2014) developed Narrative Arc Theory, which describes the typical linguistic trajectories found in fictional novels and short stories using the Linguistic Inquiry and Word Count (LIWC; Pennebaker, Booth, Boyd, & Francis, 2015) software, a widely used dictionary-based text analysis program. After dividing each text into five equal segments by word count, LIWC calculates the percentage of words in each segment that fall into various linguistic (e.g., articles, personal pronouns), cognitive (e.g., insight, causation), and affective (e.g., positive emotion, anger) categories.

Narrative Arc Theory has explored how categorization words, narrative action words, cognitive-processing words, and negative and positive emotion words follow prototypical trajectories or arcs throughout fictional narratives (Blackburn, 2015; Malin et al., 2014; see Table 1). Categorization words—made up of articles and prepositions—convey analytic thinking (Pennebaker, Chung, Frazee, Lavergne, & Beaver, 2014) and are highest in the expositional first act of a narrative (Blackburn, 2015; Malin et al., 2014). Constructing the characters and laying out the setting of a narrative requires naming objects or entities (entailing higher rates of articles) and describing how they relate to each other (entailing higher rates of prepositions). Categorization words decrease as the plot advances, because the setting and characters are already known. Following the exposition is the rising action, wherein narrative action words (personal and impersonal pronouns, adverbs, auxiliary verbs, conjunctions, and negations) increase in order to portray known characters (more often referred to by pronouns such as *he* and *she* than by name) carrying out actions to advance the plot (Blackburn, 2015; Malin et al., 2014). The rising action leads to the climax, reflected in a peak in cognitive-processing words (including insight, causation, discrepancy, tentativeness, certainty, and differentiation language categories), which help depict characters' thoughts and motivations. Finally, as the falling action leads to the resolution, rates of both negative and positive emotion words increase, reflecting attempts to find meaning in the narrative as a whole (Blackburn, 2015; Malin et al., 2014).

An analysis of the same script sample used in the present study recently extended narrative arc theory to films, showing that narrative arcs in films are similar to those found in novels and short stories, with a few key differences (Nalbandian,

Table 1 Examples of LIWC 2015 language categories for statistical analysis

Language Categories	Examples
Categorization	
Articles	<i>a, an, the</i>
Prepositions	<i>in, about, upon</i>
Narrative Action	
Impersonal	<i>another, someone, whoever</i>
Personal	<i>he, she, we, you</i>
Adverbs	<i>already, while, briefly</i>
Auxiliary verbs	<i>must, have, become</i>
Conjunctions	<i>also, because, nevertheless</i>
Negations	<i>can't, no, never</i>
Cognitive Processes	
Insight	<i>accept, perspective, realize</i>
Causation	<i>basis, infer, conclude</i>
Discrepancy	<i>abnormal, impossible, yearn</i>
Tentativeness	<i>undecided, unsure, wonder</i>
Certainty	<i>absolute, obvious, confident</i>
Differentiation	<i>alternative, distinguish, exception</i>
Emotion	
Positive	<i>joy, hugs, romantic</i>
Negative	<i>dismay, unkind, lame</i>

Categories and examples are from the most recent version of LIWC (Pennebaker et al., 2015), and composite narrative arc dimensions are from Blackburn (2015)

Iserman, & Ireland, 2018). That work established common narrative arcs across dramatic film scripts but did not find that critics' or audiences' attitudes toward films were predicted by the narrative arcs for categorization words, narrative action, cognitive processing, or emotional tone. These null results may have been due to the fact that the analyses did not include genre—a major narrative element—or linguistic measures of typicality, but rather simply regressed audience and critic ratings of films on various language categories as a function of script segment or act. Thus, in the present research we further explored whether and under what conditions films' popularity relates to genre-typicality in the language categories that define narrative arcs. More specifically, we examined genre-typical language by measuring the frequency of narrative arc theory (Blackburn, 2015; Malin et al., 2014) language categories that are typical of each film's designated genre throughout the five acts of the film.

Examining film (preferences) through genre

Genre is a cognitive schema that offers individuals information regarding categories of movies, novels, and music (Cutting, 2016; Shevy, 2008). Movie audiences likely use

genre in much the same way that people use any other categories in everyday life. Individuals naturally tend to categorize stimuli in order to simplify their environments, which facilitates the development of schemas or mental representations of the world (MacWhinney, 2015). Genres are essentially basic-level (i.e., salient and commonly used) categories that people use to simplify their movie-watching experience—including when deciding which films to watch, forming expectancies, and ultimately judging a film's quality. As the long history of genre theory attests, taxonomizing fiction is far from straightforward, and many questions remain about the nature of genre, including whether genres are natural kinds (having objective reality independent of human perception) or socially constructed, and thus tied to particular cultures and times (Chandler, 1997; Frow, 2014).

At minimum, researchers interested in the intersection of art and behavioral science tend to agree that a folk psychology of genre guides consumer behavior. That is, people have relatively reliable intuitions about what various fictional genres entail and what kinds of fiction they enjoy, and people use those expectations to guide them to the narratives they seek out in everyday life (Fong, Mullin, & Mar, 2013). Once the genre of a novel or film becomes apparent, the reader or viewer anticipates a storyline consistent with knowledge of the genre (e.g., expecting tragedies to include major character deaths; Cutting, 2016; Leavitt & Christenfeld, 2011), allowing audiences to roughly understand the storyline of a film before they view it. Film and literature theorists suggest that the film industry produces narratives that, in part, conform to specific genres in order to coincide with audience expectancies, maximizing their enjoyment (Altman, 1984). In psychological terms, a movie's similarity to other films in its genre is bound to influence processing fluency—that is, the degree to which viewers find the narrative easy (familiar and expected) or difficult (unfamiliar and unexpected) to process (Reber, Winkielman, & Schwarz, 1998).

Processing fluency and disfluency Ease of processing, or processing fluency, has profound and complex effects on attitudes and decision-making. Early research on processing fluency established that individuals tend to like stimuli that are easier to process across multiple contexts, preferring familiar images (Winkielman, Schwarz, Fazendeiro, & Reber, 2008), images viewed over longer periods (Forster, Leder, & Ansorge, 2013; Reber et al., 1998), sans serif or Times fonts (Kaspar, Wehlitz, von Knobelsdorff, Wulf, & von Saldern, 2015; Reber, Wurtz, & Zimmermann, 2004b), legible handwriting (Greifeneder et al., 2010), faces with mathematically average or symmetrical features (Reber, Schwarz, & Winkielman, 2004a; Winkielman et al., 2008), prototypical animals (Halberstadt & Rhodes, 2000), and recently primed product labels (Labroo, Dhar, & Schwarz, 2007) over less fluent stimuli—effects that are related to but not wholly accounted for by

perceived familiarity (Halberstadt, 2006). Processing fluency influences reasoning, as well, and is likely the root of the availability heuristic, in which the likelihood of a statement is evaluated based on the ease with which a person can generate confirmatory examples or evidence (Schwarz et al., 1991; Tversky & Kahneman, 1973).

Recent research has suggested that there are benefits to disfluency in some contexts. For example, an experiment on fonts that differed in legibility revealed a quadratic relationship between legibility and both recall and transfer performance; that is, learning performance increased if the text was slightly illegible, but decreased at the highest level of illegibility (Seufert, Wagner, & Westphal, 2017). Along the same lines, difficult-to-process fonts increase individuals' ability to solve riddles, such as the Moses illusion ("How many animals of each species did Moses bring on the Ark?"), that rely on shallow reading of text (Song & Schwarz, 2008). In consumer behavior, as well, people seem to prefer products with more challenging or disfluent descriptions when seeking items for special occasions, theoretically due to individuals' beliefs that exclusive products should be difficult to obtain (Pocheptsova, Labroo, & Dhar, 2010). Altogether, the literature indicates that disfluent stimuli may force individuals to engage with the material at a deeper level, yielding greater learning and more persuasive advertising in cases where people are motivated to think critically or expect to be challenged.

Complementing research on the role of fluency in reading and persuasion, results have differed on how and under what circumstances more or less fluent visual art is preferred. Theoretically, art that challenges processing fluency may be favored because it encourages individuals to broaden how they perceive the piece and, perhaps as a result, their environments (Belke, Leder, & Carbon, 2015). Yet there is some evidence that the appeal of fluency carries over into visual art, at least for certain styles of paintings: For representational visual art (depicting real-world objects or events), processing fluency seems to increase liking, but enjoyment of abstract art is unrelated to processing fluency (Belke, Leder, Strobach, & Carbon, 2010). The way that individuals process visual art additionally influences whether people prefer challenging or easy-to-process pieces, suggesting that challenging paintings are appreciated only when people elaborate on and develop mastery over a particular painting or artist's style (Belke et al., 2015; Gerger, Forster, & Leder, 2017). Similarly, visually complex works of art are preferred over less complex pieces, but only when individuals are able to easily attribute meaning to the visually complex stimuli (i.e., low perceptual fluency paired with high conceptual fluency; Ball, Threadgold, Marsh, & Christensen, 2018). That is, conquering disfluency—or making sense of previously disfluent sensory experiences—may be one pathway through which difficult art is perceived as more rewarding than simpler art that requires less initial effort to enjoy (Ball et al., 2018; Graf & Landwehr, 2015).

Processing fluency has mixed associations with music preferences, as well. In a recent study that examined linguistic fluency through repeatedly exposing participants to music lyrics, a higher number of chorus repetitions as well as a higher Hirsch–Popescu point (indicative of repetitive words) predicted top-rated songs and songs achieving top status more quickly (Nunes, Ordanini, & Valsesia, 2015). However, songs with too many repetitive words were less successful (Nunes et al., 2015). Adding to these findings, by analyzing different genres of popular music separately, Berger and Packard (2018) found that songs in most genres—with the exception of pop and dance—are more popular, as indexed by digital download rates, when their lyrics are more mathematically distinct from the average lyrics in their respective genres (Berger & Packard, 2018).

Most germane to the present research, subjective experiences of processing fluency or disfluency likely inform individual preferences for film, too. Outside of psychology, literary and film critics seem to firmly support the argument that the best movies challenge viewers to broaden their perspectives. Film critics are particularly dismissive of formulaic or predictable movies, often categorically deriding any popular films or movies made by large Hollywood studios as “genre films” (Chandler, 1997, p. 6). Renowned film critic Roger Ebert argued that “the odds against making a truly good movie are discouraging in Hollywood, which uses formulas and deals and habit patterns to push even the most original projects into narrow channels” (Ebert, 2017, p. 83). Consistent with the finding that people like stimuli that are easier to process, films from less familiar genres tend to have lower box office revenues (unless the films included famous actors or actresses and were highly rated by film critics), whereas the box office revenues of films in more familiar genres were not affected by star power or critic ratings (Desai & Basuroy, 2005). People additionally prefer to learn central plot information in advance (or to read “spoilers”) in fantasy/thriller films but not comedies, perhaps because fantasy and thriller genres are expected to be more complex than comedies (Johnson & Rosenbaum, 2018). Finally, people are less likely to favor films with elements that correspond to multiple genres, potentially because films that are difficult to categorize are less easily processed (Hsu, 2006). Such results suggest that whether fluency increases or decreases viewers’ attitudes toward a film may depend on genre—specifically, on whether the genre itself is typically easy or difficult to process.

To summarize, although the positive effects of processing fluency remain robust across domains (Winkielman et al., 2008), some degree of disfluency appears to be helpful or appealing in cases where stimuli benefit from deeper processing (Graf & Landwehr, 2015)—either because they are meant to challenge conventions (e.g., rock or rap lyrics, in Berger & Packard, 2018) or because they require critical thinking (e.g., riddles, in Song & Schwarz, 2008). Therefore, the present study

extends previous research concerning processing fluency and art by examining how audience attitudes relate to genre-typicality in language categories identified as being central to the narrative arcs of fictional writing (Blackburn, 2015; Malin et al., 2014) and dramatic films (Nalabandian et al., 2018).

Critic versus audience film preferences

Although disfluency in the arts appears to predict positive evaluations of stimuli, alternative factors could moderate these effects. Individual differences in expertise, motivations, and roles influence the effect of processing fluency on attitudes. At the level of categorization, novices and experts differ in what they consider to be a basic-level category, with novices more commonly using relatively broad superordinate categories (e.g., horror, thriller) and experts more readily using finer-grained subordinate categories (e.g., Hammer horror or *giallo*; Grey, 2016; Tanaka & Taylor, 1991). Ease of processing—and, by connection, the relation between processing fluency and attitudes—likely differs as a function of expertise. For instance, critics may identify and become bored by narrative formulas in films at faster rates than do audiences, who may be less familiar with a given formula or trope.

Motivation additionally influences individuals’ attitudes, partly via changing expectancies (Freitas, Azizian, Travers, & Berry, 2005). Perhaps novice audiences’ enjoyment of movies will be bolstered by processing fluency not only due to their relative naïveté, but also due to the mindsets (goals and expectancies) they bring with them into a movie-viewing experience. Irrespective of genre, audiences’ and professional critics’ film preferences differ. Audiences favor entertaining blockbuster films that are theoretically easier to process, and critics more frequently prefer complex, artistic films (Austin, 1983; Holbrook, 1999; Simonton, 2009, 2011; Wanderer, 1970). Differences in audience and critic preferences may be explained by their locomotion versus assessment mindsets. A person with a locomotion mindset is focused on the movie-viewing experience itself (watching a movie to enjoy the film), whereas a person with an assessment mindset is focused on evaluating that experience (watching a movie to judge it; Avnet & Higgins, 2003; Finkel, Eastwick, Karney, Reis, & Sprecher, 2012; Gollwitzer & Bayer, 1999). If typical nonexpert audience members tend to have a locomotion mindset, they might prefer entertainment characterized by higher processing fluency (more prototypical, formulaic films). By the same logic, if professional film critics have an assessment mindset, then they would prefer stimuli that would challenge them (more atypical, less formulaic films).

Audiences’ and professional critics’ differing perceptions of movies—theoretically perpetuated by their locomotion (film as entertainment) and assessment (film as art) mindsets, respectively—may be magnified in genres that are typically

expected to be more entertaining or less complex. Expectations about whether a film will be entertaining or challenging differ across genres, with action movies stereotypically prioritizing entertainment and tragedies more often aiming to be artistic (Eliashberg, Hui, & Zhang, 2007). Therefore, in addition to predicting that audiences would prefer genre-typicality and critics would prefer genre-atypicality in films, we expected genre to moderate those effects. Specifically, we expected differences between audiences' and critics' preferences for genre-typicality or genre-atypicality, respectively, to be stronger for genres that aim to entertain (action/adventure, comedy, romance, family/kids) rather than to challenge (history/war, tragedy, science-fiction/fantasy, thriller/suspense; Eliashberg et al., 2007; Johnson & Rosenbaum, 2018).

Hypotheses

Despite masses of research on the cognitive architecture of categories and how category membership (especially with respect to averageness or typicality) influences attitudes, to our knowledge no research to date has tested how genre-typicality, as indexed by quantitative language patterns, relates to film preferences. Although there are many ways to operationally define films' prototypicality or averageness within genres, we focused on simple quantitative measures of semantic (the setting, emotions, and action of a narrative) and syntactic (function word usage within the narrative; Altman, 1984) narrative arc typicality within hand-coded basic-level genres. The use of automated text analysis in the behavioral and computer sciences has rapidly gained momentum in the last decade (Herrmann, van Dalen-Oskam, & Schöch, 2015; Iliev, Dehghani, & Sagi, 2015); however, only a few studies have analyzed narrative or fictional samples with the aim of testing psychological hypotheses (e.g., Berger & Packard, 2018; Danescu-Niculescu-Mizil & Lee, 2011; Iliev, Hoover, Dehghani, & Axelrod, 2016). In the present study, we first analyzed dramatic movie scripts using dictionary-based computerized text analysis and then used profile correlations to assess whether genre-typical arcs in several language dimensions from narrative arc theory (categorization, narrative action, and cognitive processing, as well as negative and positive emotional language; Blackburn, 2015; Malin et al., 2014; Nalbandian et al., 2018) differentially predict audience and critic ratings.

In summary, we did not have specific predictions about which language dimensions would show the strongest fluency effects or would be most clearly moderated by rater role; rather, we used Narrative Arc Theory to identify which specific language categories would be most central to individuals' understandings of filmic narratives. On the basis of processing fluency research and genre theory, we predicted that (H1) critics would prefer atypicality (i.e., the unfamiliar and

unexpected) and audiences would prefer typicality (i.e., the familiar and expected), with respect to each genre's average linguistic narrative arcs in the present sample, and (H2) the associations between genre-typicality and movie ratings would be stronger for more stereotypically formulaic genres, such as action/adventure, romance, comedy, and family/kids films. In other words, critics would prefer less genre-typical language and audiences would prefer more genre-typical language in the aforementioned genres. To test our hypotheses, several linear models (one per genre) regressed ratings of film quality on measures of linguistic genre-typicality, including rater role (audience or critic) as a moderator.

Method

Researchers focusing on the study of film and, in particular, individuals' perceptions of films have increasingly begun to utilize big-data approaches. Mostly, studies obtain massive samples of descriptive data on films by means of websites such as the Internet Movie Database (IMDb), Rotten Tomatoes, and MovieLens (Desai & Basuroy, 2005; Hsu, 2006; Hwang, Park, Hong, & Kim, 2016; Ramos, Calvão, & Anteneodo, 2015). The present study applies similar big-data practices, albeit on a more manageable scale that allowed for detailed hand-coding of films' genres, writer characteristics, and ratings. Here we focused primarily on temporal language patterns in scripts (referred to as *narrative arcs*), as well as audience and critic ratings.

The sample included 509 scripts from the Internet Movie Script Database's (IMSDb's) drama category, for films released between 1932 and 2014. IMSDb is a resource that allows one to freely download or view film scripts. The scripts available on IMSDb are not transcribed, edited, or annotated by fans; rather, they represent authentic film scripts written by expert screenwriters. Consistent with gender disparities in Hollywood (see Conor, Gill, & Taylor, 2015), all but a few scripts were written by individual men or all-male writing teams (87.4%), followed by individual women or all-female writing teams (7.1%), and mixed-sex writing teams (5.5%). Analyses processed all text from the scripts, including both screen directions and dialogue, on the basis that key narrative elements—particularly descriptions of characters' affect or thought processes (e.g., *angrily, thoughtful pause*)—are often included parenthetically, without being explicitly discussed in dialogue.

We excluded 25 scripts from the analysis, due to some scripts being formatted as images that were not amenable to optical character recognition ($n = 3$), having fewer than 20 user or critic reviews on Rotten Tomatoes and IMDb ($n = 21$), or having no dialogue to analyze ($n = 1$; i.e., *The Artist*, 2011).

Analytic strategies

Text analysis The Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2015; version 1.3.1) computer program was used to determine the rate of specific word categories for each film script. LIWC measures 125 different grammatical (e.g., pronoun, article), psychological (e.g., positive emotion, power), and topical (e.g., religion, death) language categories containing over 6,500 unique terms. Every language category includes a list of words and word stems, with the goal of representing commonly used words in that category. LIWC compares every word in a given text against those internal dictionaries and then outputs the percentage of the total words in each language category.

In this study, we focused on five specific language dimensions identified as central to narrative in past work on Narrative Arc Theory (Blackburn, 2015; Malin et al., 2014; Nalbandian et al., 2018): categorization words, narrative action words, cognitive-processing words, positive emotion words, and negative emotion words. Categorization and narrative action words are composite variables that incorporate various standardized language categories. The remaining language variables are each representative of one LIWC category. All 509 drama film scripts included in the analyses were split by word count into five equal segments (or acts), in order to assess the trajectory of each language category across the course of each script, as was recommended by both Blackburn (2015) and Malin et al. (2014). Other theories of narrative that focus on the visual elements of film (e.g., shot durations, scene transitions, music) have argued for the existence of four or six arcs in films (depending on whether a prologue and epilogue are included; Cutting, 2016). Because the present research is centered on linguistic analyses of film scripts and proposes to further extend Blackburn's (2015) and Malin et al.'s (2014) Narrative Arc Theory, in the present study we examined scripts using the five-arc format.

Language categories Table 1 displays examples of words that make up the aforementioned language categories. The categorization language composite was created by standardizing (z -scoring) and averaging the percentages for articles and prepositions (standardized Cronbach's $\alpha = .74$) within scripts. Similarly, the narrative action words composite was also computed by averaging the standardized impersonal pronouns, personal pronouns, adverbs, auxiliary verbs, conjunctions, and negations ($\alpha = .84$) in all scripts. Cognitive processing words consisted of the insight, causation, discrepancy, tentativeness, certainty, and differentiation language categories. Finally, positive and negative emotion words were standardized as distinct variables.

Film ratings Film ratings were collected from Rotten Tomatoes as well as IMDb. We chose those sites over other

options, such as Metacritic or Fandango, due to their simpler and more transparent rating algorithms. For instance, the ratings on Fandango appear to attribute at least three stars out of five for most films, perhaps to increase ticket revenue (Hickey, 2015). Metacritic weights critic ratings in nontransparent ways, primarily to give more influence to better-known critics. In contrast, IMDb utilizes weighted averages for audience ratings in order to limit errors (e.g., repeated ratings by the same user or extreme outliers). Rotten Tomatoes forgoes weighting and instead designates a score (ranging from 1 to 5 for audience ratings and 1 to 10 for critic ratings) to each review, reporting the proportion of total reviews that are positive—or “fresh”—as the percent rating (see <https://www.rottentomatoes.com/about/>).

Critic ratings Critic ratings were obtained from Rotten Tomatoes. Critics ranged from those associated with high-impact online and print news sources, such as *The Washington Post* and *Rolling Stone*, to more specialized movie review and entertainment news sites, such as Emanuel Levy or Common Sense Media. The percentage of positive critic ratings as well as the critic ratings out of ten were both standardized and averaged into a composite variable for critic ratings ($\alpha = .98$).

Audience ratings Audience ratings were also obtained from Rotten Tomatoes and IMDb. The ratings from Rotten Tomatoes (percentage of positive ratings and ratings out of five) and IMDb (a weighted average) were each standardized and subsequently averaged into a composite variable for audience ratings ($\alpha = .95$).

Genre Although the present corpus of films was collected from the IMDB's drama category, most of the films could be subcategorized into different genres, as well (e.g., *Knocked Up* as a comedy). Therefore, the sample of film scripts were coded into several subgenre categories by two research assistants who conducted in-depth searches on Rotten Tomatoes and IMDb for each film. Because both Rotten Tomatoes and IMDb identify multiple genres for films, the raters determined one genre—from the various genres specified on Rotten Tomatoes and IMDb—that best represented each film in the present study's sample. Subgenres for the majority of films were agreed upon by the two raters. A third rater acted as the tie-breaker for films in which the initial two raters reported discrepant subgenres. From the three raters' efforts, the films were all coded into eight subgenres: comedy ($n = 73$), romance ($n = 70$), thriller/suspense ($n = 134$), action/adventure ($n = 77$), science-fiction/fantasy ($n = 58$), history/war ($n = 68$), tragedy ($n = 18$), and family/kids ($n = 11$). The unequal sample sizes of films in the different subgenre categories are on account of the present sample originally being collected for the purposes of previous research examining the linguistic

arcs of narrative in film (Nalabandian et al., 2018). In the present study, we attempted to extend Nalabandian et al.'s (2018) work concerning which linguistic narrative arcs are preferred by audiences versus professional critics.

To facilitate the process of identifying one genre—from the multiple genres exhibited on Rotten Tomatoes and IMDB—for every film, raters were asked to select which subgenre was the most predominant descriptor of each film based on the following criteria illustrated by professional film critic Tim Dirks (2018a, 2018b, 2018c): (1) Films with a main storyline involving love and breakups were identified as romances, (2) films intended to be funny were identified as comedies, (3) films with fictional characters in past wars or other real historical settings (i.e., set more than a decade prior to the film's release date) were identified as historical fiction, (4) films featuring imagined future technology, magical/super powers, or mythical creatures were identified as science fiction or fantasy, (5) fast-paced films with violence—usually with car chases and lone or semi-lone heroes triumphing over a large number of bad guys—were identified as action and adventure, (6) films with mysteries unresolved until the end of the film, with surprise twists, or with an eerie atmosphere were identified as thriller and suspense, (7) films in which predominantly bad things happen to good characters or heroes, with little or no redemption at the end—often involving the deaths of most or all main characters—were identified as tragedies, and (8) films aimed at entertaining children were categorized as family and kids films.

Genre-typical language Genre-typicality was operationalized as the profile correlations calculated for each of the five narrative-arc-related language categories (categorization, narrative action, cognitive processing, negative emotion, and positive emotion) between each film's trajectory over the course of the five segments (or acts) and the average values for each segment within each genre. That is, after creating the composite language variables for categorization and narrative action, we calculated individual category-level profile correlations across the five acts of to films between each movie and the mean for that category in a movie's genre. Those category-level profile correlations were then transformed using Fisher's z . The Fisher (1921) z -transformation is used to convert correlations (which are bounded by -1 and 1) to a continuous scale and to stabilize the variance. The resulting number reflects whether the arc—or trajectory over the course of a film's five acts—for a given film's composite narrative language category matches its genre's average or typical arc for that same category.

Analytic strategy To test whether audiences and critics differed in their preferences for genre-typical language patterns, we regressed the film ratings on Fisher's z -transformed profile correlations and rater role (audience or critic) separately for

each genre: $\text{Ratings} \sim \text{Rater} \times \text{Genre-Typicality}$, where *Ratings* serves as the continuous variable referring to the standardized film ratings, *Rater* represents the categorical variable distinguishing the continuous *Ratings* variable as audience or critic ratings, and *Genre-Typicality* signifies Fisher's z -transformed profile correlations between films and their genre's typical pattern for the specific language category. All analyses were conducted in R (version 3.4.3; R Core Team, 2017), and mixed-effects models were executed with the lme4 package (version 1; Bates, Maechler, Bolker, & Walker, 2015).

Results

The interaction of rater role and language—testing the prediction that language typicality and film ratings would differ between audiences and critics—was significant only for genre-typical positive emotion language (Table 2). However, simple slope analyses were nonsignificant for both critic ($B = -.05$, $SE = .04$, $t(507) = -1.03$, $p = .304$, 95% CI $[-.13, .04]$) and audience ($B = .04$, $SE = .04$, $t(507) = 0.81$, $p = .417$, 95% CI $[-.05, .12]$) ratings. Subsequent analyses testing the hypothesis that genre would play a role in the audience and critic ratings were significant only for genre-typical positive emotion language in action/adventure films ($B = -.25$, $SE = .08$, $t(75) = -3.22$, $p = .002$, 95% CI $[-.41, -.10]$, all other $ps > .10$ see Table 3). Simple slope analyses revealed that critics, but not audiences ($B = .01$, $SE = .11$, $t(75) = .07$, $p = .948$, 95% CI $[-.22, .24]$), rated action/adventure films with less genre-typical positive emotion trajectories more highly ($B = -.25$, $SE = .12$, $t(75) = -2.14$, $p = .035$, 95% CI $[-.48, -.02]$, see Fig. 1). Such findings are consistent with the prediction that critics prefer less genre-typicality in films, and particularly for formulaic films such as action/adventure films.

Because our predictions regarding divergent preferences between audiences and critics for genre-typicality were corroborated solely for positive emotion language—not for categorization, narrative action, cognitive processing, or negative emotion language—we conducted additional analyses examining the main effects for genre-typicality independent of rating category (Table 4). Controlling for rater role, more genre-typical categorical language predicted significantly lower ratings of action/adventure films ($B = -.33$, $SE = .11$, $t(75) = -3.12$, $p = .003$, 95% CI $[-.54, -.12]$) and marginally lower ratings of romances ($B = -.22$, $SE = .11$, $t(68) = -1.95$, $p = .055$, 95% CI $[-.45, .01]$; see Fig. 2). Similarly, genre-typical positive emotion language predicted lower ratings of history/war films ($B = -.24$, $SE = .10$, $t(66) = -2.32$, $p = .023$, 95% CI $[-.45, -.03]$; see Fig. 3). Genre-typical narrative action language predicted marginally lower ratings for both action/adventure ($B = -.19$, $SE = .10$, $t(75) = -1.89$, $p = .063$, 95% CI $[-.40, .01]$) and history/war films. ($B = -.16$, $SE = .09$, $t(66) = -1.79$, $p = .078$, 95% CI $[-.34, .02]$, see Fig. 4). In

Table 2 Rater role-by-genre interaction effects predicting film ratings

Genre-Typical Language	B	SE	t	df	p	95% CI
Positive emotion	-.08	.03	-2.60*	507	.010	[-.14, -.02]
Negative emotion	.02	.03	0.70	507	.482	[-.04, .08]
Cognitive processes	.03	.03	1.04	507	.300	[-.03, .09]
Categorization	-.004	.03	-0.11	507	.909	[-.07, .06]
Narrative action	-.02	.03	-0.56	507	.579	[-.08, .04]

* $p < .05$

contrast, both genre-typical positive emotion ($B = .60$, $SE = .25$, $t(9) = 2.37$, $p = .042$, 95% CI [.03, 1.17], see Fig. 3) and negative emotion language ($B = .64$, $SE = .27$, $t(9) = 2.34$, $p = .044$, 95% CI [.02, 1.26], see Fig. 5) predicted higher ratings of family/kids films.

The main effects for linguistic genre-typicality suggest that, regardless of whether the viewer is an audience member or a professional critic, most people prefer less genre-typical categorization words, narrative action words, and positive emotion words in action/adventure, romance, and history/war films. These results are inconsistent with the prediction that audiences would prefer genre-typicality within films in general. However, ratings of family/kids films suggest that both positive and negative emotion language are preferred to be more genre-typical. Violating expectancies by deviating from narrative formulas within this genre may be upsetting for children only beginning to develop how they view themselves and the world around them. However, given that there were only 11 family/kids films in our sample, any results for that genre should be treated with caution until the same pattern is replicated in larger and more representative samples.

Discussion

In the present research, we examined how genre-typical language—specifically, narrative arc language categories implicated in novels, short stories (Blackburn, 2015; Malin et al., 2014), and film (Nalbandian et al., 2018)—predicted film ratings. Partly consistent with our predictions, professional critics preferred less genre-typical positive emotion trajectories in action/adventure films. This finding coincides with research suggesting that critics may be drawn to more complex and artistic films (Austin, 1983; Holbrook, 1999; Simonton, 2009, 2011; Wanderer, 1970) that are likely to challenge the viewer with less typical or formulaic stimuli. Affectively challenging films elicit aversive or inconsistent emotions, whereas positive emotion in film is less challenging to process (Eden, Johnson, & Hartmann, 2018), perhaps explaining critics' preference for less genre-typical positive emotion in film,

particularly in action/adventure films known to follow a narrative formula. In other words, genres—or schematic information based on past experiences—guide the viewer in what to expect from a film, but when such information is inconsistent with one's schema, expectancies are subverted, and the viewer must work to further process the information. Being film experts, critics may also be more aware of—and thus perhaps more bored by—typical positive emotion trajectories within film genres, relative to more naïve audience members (Belke et al., 2010; Oppenheimer, 2008). Nevertheless, differences between audience and professional critic film preferences only yielded significant results for positive emotion language, leading us to suspect audience and critic preferences differ less than past research has suggested.

Exploratory analyses controlling for film rater role (audience vs. critic) revealed several significant main effects for genre-typical language. Contrary to our hypotheses and prior research indicating that individuals prefer stimuli that can be processed more fluently (Winkielman et al., 2008), both audiences and professional critics preferred *less* genre-typical (and theoretically harder-to-process) language patterns in most genres. Less typical categorical, narrative action, and positive emotion language patterns were preferred in romance, action/adventure, and history/war films, all genres that are sometimes criticized for being overly formulaic. Although we did not predict that genre-typicality would be more appealing for any one genre or set of genres, only family/kids films received higher ratings while adhering to genre-typical emotional tone. Since children are the intended audience of family/kids films, facilitating understanding and comprehension of the filmic narrative via genre-typical language use may be particularly important for this genre. Children may not have yet developed the ability to appreciate the challenge of interpreting more disfluent stimuli, which is partly supported by results demonstrating second graders' difficulty—and fifth graders' ease—in processing smaller font sizes (Katzir, Hershko, & Halamish, 2013).

Although much of the processing fluency literature supports the conclusion that people prefer stimuli that are easy to follow, understand, and comprehend, recent research (e.g.,

Table 3 Rater role-by-genre interaction effects predicting film ratings, subset by genre

Genre-Typical Language	Genre	B	SE	t	df	p	95% CI
<i>Positive Emotion</i>	Action/adv	-.25	.08	-3.22**	75	.002	[-.41, -.10]
	Romance	.04	.09	0.46	68	.647	[-.14, .23]
	Comedy	-.01	.08	-0.19	71	.846	[-.17, .14]
	Thriller/susp	-.01	.06	-0.13	132	.898	[-.13, .11]
	Sci-fi/fantasy	-.10	.08	-1.19	56	.239	[-.27, .07]
	Tragedy	-.18	.28	-0.64	16	.532	[-.78, .42]
	History/war	-.10	.07	-1.47	66	.146	[-.24, .04]
	Family/kids	-.30	.29	-1.01	9	.338	[-.95, .36]
<i>Negative Emotion</i>	Action/adv	-.08	.10	-0.79	75	.432	[-.27, .11]
	Romance	-.06	.07	-0.83	68	.410	[-.20, .08]
	Comedy	.02	.09	0.28	71	.781	[-.15, .20]
	Thriller/susp	.03	.06	0.52	132	.605	[-.09, .16]
	Sci-fi/fantasy	.01	.09	0.13	56	.897	[-.17, .19]
	Tragedy	-.01	.28	-0.05	16	.959	[-.61, .58]
	History/war	-.01	.07	-0.10	66	.922	[-.14, .13]
	Family/kids	.36	.31	1.15	9	.279	[-.34, 1.06]
<i>Cognitive Processes</i>	Action/adv	.12	.07	1.63	75	.110	[-.03, .26]
	Romance	.06	.10	0.59	68	.560	[-.14, .25]
	Comedy	-.03	.09	-0.29	71	.775	[-.20, .15]
	Thriller/susp	.004	.07	0.05	132	.958	[-.13, .14]
	Sci-fi/fantasy	-.002	.08	-0.02	56	.980	[-.16, .15]
	Tragedy	-.37	.26	-1.45	16	.167	[-.91, .17]
	History/war	.09	.07	1.44	66	.154	[-.04, .22]
	Family/kids	.02	.22	0.10	9	.924	[-.47, .51]
<i>Categorization</i>	Action/adv	.11	.09	1.32	75	.192	[-.06, .28]
	Romance	-.01	.09	-0.14	68	.888	[-.20, .17]
	Comedy	-.07	.09	-0.81	71	.422	[-.26, .11]
	Thriller/susp	.003	.07	0.04	132	.965	[-.13, .13]
	Sci-fi/fantasy	-.02	.08	-0.26	56	.793	[-.17, .13]
	Tragedy	-.20	.29	-0.68	16	.505	[-.80, .41]
	History/war	-.06	.06	-1.06	66	.292	[-.18, .05]
	Family/kids	.01	.19	0.07	9	.945	[-.41, .43]
<i>Narrative Action</i>	Action/adv	.08	.08	0.94	75	.348	[-.08, .24]
	Romance	-.08	.09	-0.93	68	.353	[-.26, .09]
	Comedy	-.04	.08	-0.53	71	.601	[-.21, .12]
	Thriller/susp	-.04	.07	-0.52	132	.602	[-.17, .10]
	Sci-fi/fantasy	-.11	.08	-1.37	56	.175	[-.27, .05]
	Tragedy	-.25	.32	-0.76	16	.449	[-.92, .43]
	History/war	.02	.06	0.30	66	.765	[-.10, .13]
	Family/kids	-.06	.26	-0.24	9	.813	[-.65, .53]

Action/adv = Action/adventure, Thriller/susp = Thriller/suspense, Sci-fi/fantasy = Science-fiction/fantasy. ** $p < .01$

Belke et al., 2015) suggests that such findings may not be generalizable to art, which is often celebrated for being novel or iconoclastic. Art that is more difficult to process allows the individual to encounter an unexpected, novel stimulus that elicits curiosity and provides an opportunity to learn from and attribute meaning to the aesthetic experience (Belke

et al., 2015). Psychological research on literature and film presents a more complex picture, however, suggesting that audiences have traditionally preferred narrative structures that are easy to process (Cutting, 2016; Thompson, 1999) but are progressively gaining appreciation for complex, more artistic films that deviate from popular norms

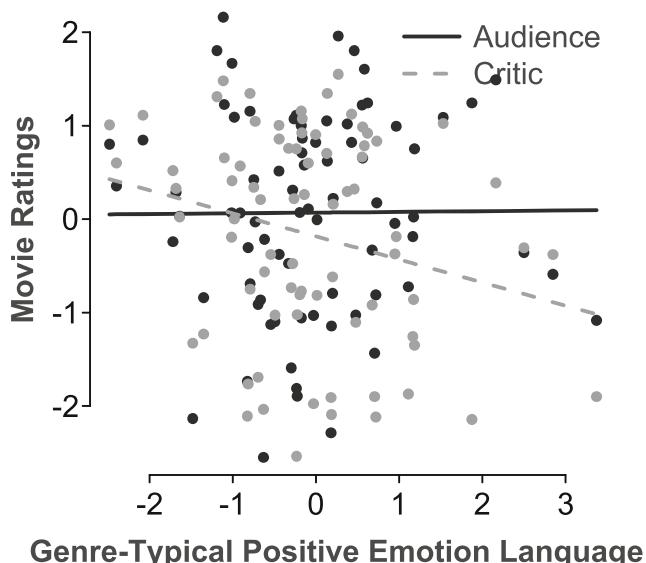


Fig. 1 Action/adventure movie ratings as a function of rater role and genre-typical positive emotion language

(Mittell, 2006). For example, television shows are increasingly abandoning stand alone episodes that require no previous knowledge of the show in favor of adding to season- or series-spanning narrative arcs with every episode—a change that is partly attributable to technological advances in digital recording and streaming that allow viewers to rewatch and minutely dissect complex scenes (similar to rereading a complex passage in a novel; Mittell, 2006). Perhaps critics' and audiences' shared preferences for less typical narrative arcs are related to these technology-driven cultural shifts in how people view films.

Implications of computerized text analysis

The present study represents a broad, theory-based snapshot of how genre-typicality relates to audiences' and critics' attitudes toward films. Although our results are informative and contribute to the literature on attitudes and processing fluency in art, there is clearly much work to be done on finer-grained—or more holistic—aspects of genre-typicality. For example, Berger and Packard (2018) took a more data-driven approach to analyzing creative work, operationalizing lyrical genre-typicality as the strength of the correlation between songs and genre means across several latent Dirichlet-allocation-derived topics (Blei, Ng, & Jordan, 2003) and LIWC categories. At the other end of the spectrum, it may be interesting to explore how the trajectories of single, psychologically telling words—such as subjective (*I*) versus objective (*me*) first-person singular pronouns—relate to audience enjoyment (Zimmermann, Brockmeyer, Hunn, Schauenburg, & Wolf, 2017). Bringing this paradigm into the laboratory could be informative, as well, in terms of assessing how personality, intelligence, demographics, or

other individual differences relate to individuals' enjoyment of typical or atypical narrative arcs. Even our analysis of genre itself was relatively simple and can potentially be bolstered by more data-driven methods of determining which genre or genres a movie belongs to.

A broad aim of this project was to spark other behavioral scientists' interest in using naturalistic data and computational linguistic tools to study psychological phenomena. Movies are an especially valuable, and often overlooked, source of real-life archival data. Not everyone reads print publications, but nearly every person alive today—of all ages, and across developed and developing nations—regularly consumes television or movies of some kind, with an estimated 119.6 million homes in the United States alone currently having televisions (Nielsen, 2017). Since film is a highly profitable and globalized industry, copious and well-curated records for film expenditures, theatrical releases, and sales are often freely available to the public (e.g., from Box Office Mojo, <http://www.boxofficemojo.com>, operated by IMDb).

The sanctity of naturalistic behavioral data in the social and behavioral sciences cannot be exaggerated. The field of social psychology, in particular, rose out of researchers' desire to understand how common people, without the excuse of psychopathy or other mitigating factors, could contribute to the unthinkable horrors of the early 20th century, including the World Wars, the Holocaust, and the Armenian genocide (Perry, 2018; Russell, 2011). Theories that work well in the laboratory are worth very little if they do not help predict, understand, and, when necessary, influence real-life human behavior (Allport, 1919). Despite that widely accepted tenet, most research on psychological phenomena today relies on laboratory or online surveys and experiments, with an occasional field study to validate results demonstrated dozens of times in tightly controlled lab settings. Although several counterexamples to these generalizations have generated considerable excitement (e.g., Brown, Blake, & Sherman, 2017; Mehl, 2017; Pennebaker et al., 2014; Rocklage & Fazio, 2015; Youyou, Kosinski, & Stillwell, 2015), studying psychological phenomena in the real world remains relatively rare in psychology. Analyzing natural language use—particularly fictional language use—is much rarer in psychology departments, despite fictional texts' psychometric value as a source of abundant quantifiable behavioral data. In this respect, researchers from linguistics (e.g., Davies, 2008) and computer science (e.g., Danescu-Niculescu-Mizil & Lee, 2011; Pakhomov, Chacon, Wicklund, & Gundel, 2011) have been more adventurous, although computational linguists are traditionally trained to prioritize prediction and classification over psychological insights (for counterexamples, see computational social scientists; e.g., Boghrati, Hoover, Johnson, Garten, & Dehghani, 2018; Lazer et al., 2009; Park et al., 2015).

Table 4 Main effects of genre-typicality on film ratings after controlling for rater role, subset by genre

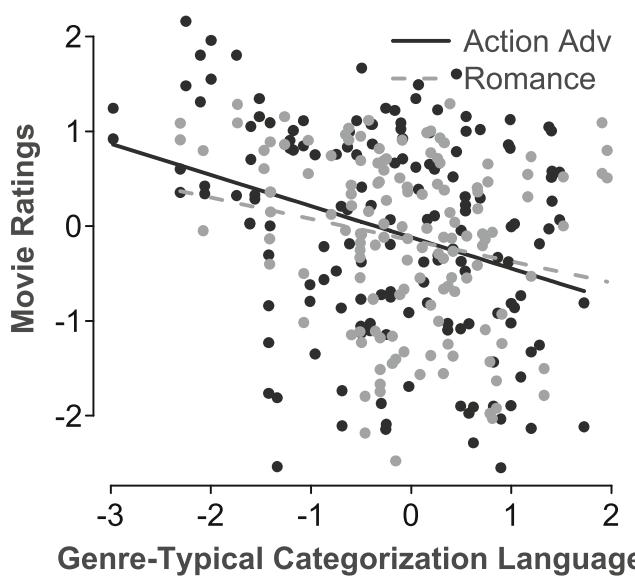
Genre-Typical Language	Genre	B	SE	t	df	p	95% CI
<i>Positive Emotion</i>	Action/adv	-.12	.11	-1.11	75	.271	[-.33, .10]
	Romance	.14	.11	1.26	68	.214	[-.08, .37]
	Comedy	.09	.10	0.87	71	.386	[-.11, .28]
	Thriller/susp	.02	.09	0.26	132	.799	[-.15, .19]
	Sci-fi/fantasy	-.08	.12	-0.67	56	.505	[-.32, .16]
	Tragedy	.14	.22	0.65	16	.525	[-.32, .60]
	History/war	-.24	.10	-2.32*	66	.023	[-.45, -.03]
	Family/kids	.60	.25	2.37*	9	.042	[.03, 1.17]
<i>Negative Emotion</i>	Action/adv	.10	.12	0.85	75	.400	[-.14, .35]
	Romance	-.04	.09	-0.44	68	.663	[-.21, .14]
	Comedy	-.12	.11	-1.06	71	.291	[-.34, .10]
	Thriller/susp	-.09	.09	-0.98	132	.330	[-.26, .09]
	Sci-fi/fantasy	-.06	.12	-0.51	56	.611	[-.31, .18]
	Tragedy	.31	.20	1.59	16	.131	[-.10, .73]
	History/war	-.17	.10	-1.62	66	.111	[-.37, .04]
	Family/kids	.64	.27	2.34*	9	.044	[.02, 1.26]
<i>Cognitive Processes</i>	Action/adv	-.15	.09	-1.65	75	.103	[-.34, .03]
	Romance	-.06	.12	-0.49	68	.627	[-.30, .18]
	Comedy	.01	.11	0.07	71	.944	[-.22, .23]
	Thriller/susp	-.03	.10	-0.32	132	.752	[-.23, .16]
	Sci-fi/fantasy	-.05	.11	-0.43	56	.670	[-.26, .17]
	Tragedy	-.15	.2	-0.74	16	.472	[-.58, .28]
	History/war	.02	.10	0.19	66	.851	[-.19, .22]
	Family/kids	.09	.22	0.41	9	.695	[-.42, .60]
<i>Categorization</i>	Action/adv	-.33	.11	-3.12**	75	.003	[-.54, -.12]
	Romance	-.22	.11	-1.95†	68	.055	[-.45, .005]
	Comedy	.05	.12	0.39	71	.699	[-.19, .28]
	Thriller/susp	.05	.09	0.48	132	.629	[-.14, .23]
	Sci-fi/fantasy	.01	.11	0.13	56	.894	[-.20, .23]
	Tragedy	-.07	.22	-0.33	16	.742	[-.54, .39]
	History/war	-.05	.09	-0.55	66	.585	[-.23, .13]
	Family/kids	.15	.19	0.78	9	.456	[-.28, .57]
<i>Narrative Action</i>	Action/adv	-.19	.10	-1.89†	75	.063	[-.40, .01]
	Romance	-.09	.11	-0.78	68	.440	[-.31, .14]
	Comedy	.07	.11	0.65	71	.520	[-.14, .28]
	Thriller/susp	.06	.10	0.66	132	.513	[-.13, .26]
	Sci-fi/fantasy	.13	.11	1.13	56	.264	[-.10, .35]
	Tragedy	-.34	.23	-1.47	16	.161	[-.83, .15]
	History/war	-.16	.09	-1.79†	66	0.078	[-.34, .02]
	Family/kids	.16	.27	0.58	9	0.575	[-.45, .76]

Action/adv = Action/adventure, Thriller/susp = Thriller/suspense, Sci-fi/fantasy = Science-fiction/fantasy. The rater role categories were audience and critic. [†] $p < .1$, ^{*} $p < .05$, ^{**} $p < .01$

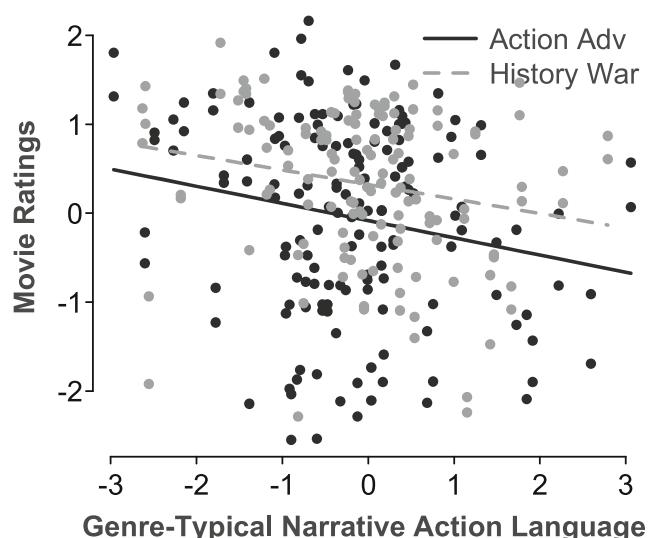
Conclusion

The present study provides evidence for the utility of a novel method of measuring genre-typicality within narrative arcs using dictionary-based text analysis. The computerized text

analysis and profile correlation methods used in the study are face-valid, transparent, and relevant to theories spanning a number of fields, ranging from social and cognitive psychology to literary and film criticism; as a result, the methods developed in this project have the potential for wide use across



diverse behavioral science fields. A more abstract aim of this work is to encourage researchers in psychology and related fields to experiment with similar methods (dictionary-based computerized text analysis, particularly of fictional language use) in their own research. Through the computerized analysis of fictional dialogue, researchers interested in perspective-taking can explore how various groups of authors have imagined the minds of people from different social and demographic groups (Ireland, Davis, Schumacher, & Pennebaker, 2018). Cognitive and affective scientists can explore how fiction relates to empathy and social simulation (Mar & Oatley, 2008). Experts on culture and sociology can test how art drives and reflects social change (Michel et al., 2011). Creativity



researchers can study the semantic signature of creative or atypical narratives (Gray, 2018). People who care about attitudes and personality can assess what language topics or narrative arcs are most appealing to various groups of people, as we aimed to do here. The questions that can be explored by computerized analyses of fictional text samples are virtually unlimited (see Pennebaker & Ireland, 2011).

Dictionary methods—relying on top-down word lists that were handmade and validated by psychologists—are blunt instruments that only indicate how broad topics or patterns of behavior relate to the outcomes we care about. However, because dictionary methods are simple and face valid, they have a low entry point—it costs little to experiment with them, and the potential gains are significant. By incorporating these

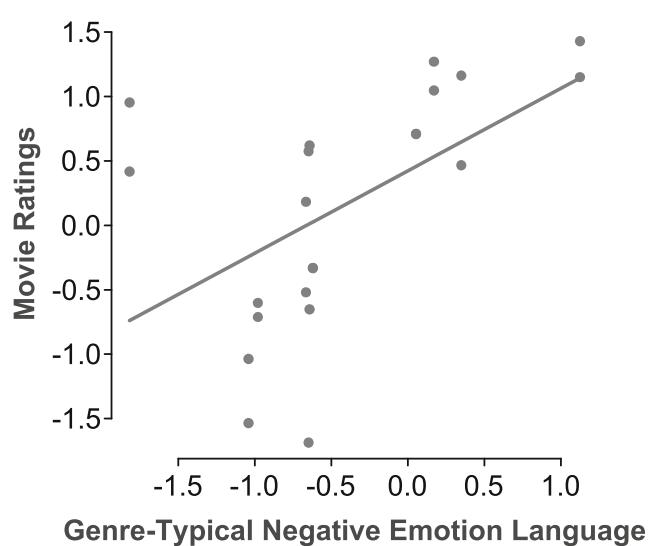
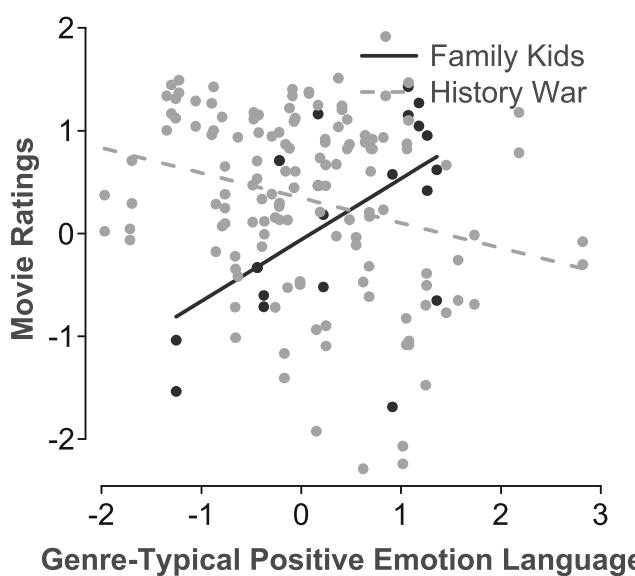


Fig. 2 Main effects of genre-typical categorization language for action/adventure and romance films

Fig. 4 Main effects of genre-typical narrative action language for action/adventure and history/war films

methods in existing areas of research, it is possible to quickly and rigorously add quantitative behavioral data. In combination with self-report data and experimental outcomes, dictionary-based analyses of verbal behavior have the potential to help triangulate the psychological mechanisms at the core of many behavioral phenomena that are of interest to researchers across the social, behavioral, and computer sciences.

Author note We are grateful to Sage Maliepaard for her thoughtful comments on multiple drafts, and Micah Iserman for helping us compile the script sample.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Allport, F. H. (1919). Behavior and experiment in social psychology. *Journal of Abnormal Psychology, 14*, 297–306.
- Altman, R. (1984). A semantic/syntactic approach to film genre. *Cinema Journal, 23*, 6–18. <https://doi.org/10.2307/1225093>
- Aristotle. (1961). *Aristotle's poetics* (S. H. Butcher, Trans.). New York, NY: Hill & Wang. (Original work published c. 335 BCE)
- Austin, B. (1983). Critics' and consumers' evaluations of motion pictures: A longitudinal test of the taste culture and elitist hypotheses. *Journal of Popular Film and Television, 10*, 156–167.
- Avnet, T., & Higgins, E. T. (2003). Locomotion, assessment, and regulatory fit: Value transfer from "how" to "what". *Journal of Experimental Social Psychology, 39*, 525–530. [https://doi.org/10.1016/s0022-1031\(03\)00027-1](https://doi.org/10.1016/s0022-1031(03)00027-1)
- Ball, L. J., Threadgold, E., Marsh, J. E., & Christensen, B. T. (2018). The effects of stimulus complexity and conceptual fluency on aesthetic judgments of abstract art: Evidence for a default–interventionist account. *Metaphor and Symbol, 33*, 235–252. <https://doi.org/10.1080/10926488.2018.1481255>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Belke, B., Leder, H., & Carbon, C. C. (2015). When challenging art gets liked: Evidence for a dual preference formation process for fluent and non-fluent portraits. *PLoS ONE, 10*, e131796. <https://doi.org/10.1371/journal.pone.0131796>
- Belke, B., Leder, H., Strobach, T., & Carbon, C. C. (2010). Cognitive fluency: High-level processing dynamics in art appreciation. *Psychology of Aesthetics, Creativity, and the Arts, 4*, 214–222. <https://doi.org/10.1037/a0019648>
- Berger, J., & Packard, G. (2018). Are atypical things more popular? *Psychological Science, 29*, 1178–1184. <https://doi.org/10.1177/0956797616678187>
- Blackburn, K. G. (2015). *The narrative arc: Exploring the linguistic structure of the narrative* (Unpublished doctoral dissertation). University of Texas at Austin, Austin, Texas.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.
- Boghrati, R., Hoover, J., Johnson, K. M., Garten, J., & Dehghani, M. (2018). Conversation level syntax similarity metric. *Behavior Research Methods, 50*, 1055–1073. <https://doi.org/10.3758/s13428-017-0926-2>
- Brown, N. A., Blake, A. B., & Sherman, R. A. (2017). A snapshot of the life as lived: Wearable cameras in social and personality psychological science. *Social Psychological and Personality Science, 8*, 592–600.
- Bushman, B. J., & Anderson, C. A. (2001). Media violence and the American public: Scientific facts versus media misinformation. *American Psychologist, 56*, 477–489. <https://doi.org/10.1037/0003-066X.56.6-7.477>
- Chandler, D. (1997). *An introduction to genre theory* (Web document). Accessed October 1, 2018, at https://www.researchgate.net/profile/Daniel_Chandler4/publication/242253420_An_Introduction_to_Genre_Theory
- Conor, B., Gill, R., & Taylor, S. (2015). Gender and creative labour. *Sociological Review, 63*(Supp), 1–22. <https://doi.org/10.1111/1467-954X.12237>
- Cutting, J. E. (2016). Narrative theory and the dynamics of popular movies. *Psychonomic Bulletin & Review, 23*, 1713–1743. <https://doi.org/10.3758/s13423-016-1051-4>
- Danescu-Niculescu-Mizil, C., & Lee, L. (2011). Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics* (pp. 76–87). Stroudsburg, PA: Association for Computational Linguistics.
- Davies, M. (2008). The corpus of contemporary American English: 425 million words, 1990–present (Database). Available at <http://corpus.bu.edu/coca/>. Accessed 04 Oct 2018
- Desai, K. K., & Basuroy, S. (2005). Interactive influence of genre familiarity, star power, and critics reviews in the cultural goods industry: The case of motion pictures. *Psychology and Marketing, 22*, 203–223. <https://doi.org/10.1002/mar.20055>
- Dirks, T. (2018a). Film sub-genres. Retrieved from <https://www.filmsite.org/subgenres.html>. Accessed 19 Sept 2018
- Dirks, T. (2018b). Main film genres. Retrieved from <https://www.filmsite.org/genres.html>. Accessed 19 Sept 2018
- Dirks, T. (2018c). Other major film categories. Retrieved from <https://www.filmsite.org/genres2.html>. Accessed 19 Sept 2018
- Ebert, R. (2017). *Awake in the dark: The best of Roger Ebert*. Chicago: University of Chicago Press.
- Eden, A., Johnson, B. K., & Hartmann, T. (2018). Entertainment as a creature comfort: Self-control and selection of challenging media. *Media Psychology, 21*, 352–376. <https://doi.org/10.1080/15213269.2017.1345640>
- Eliashberg, J., Hui, S. K., & Zhang, Z. J. (2007). From story line to box office: A new approach for green-lighting movie scripts. *Management Science, 53*, 881–893.
- Finkel, E. J., Eastwick, P. W., Karney, B. R., Reis, H. T., & Sprecher, S. (2012). Online dating: A critical analysis from the perspective of psychological science. *Association for Psychological Science, 13*, 3–66. <https://doi.org/10.1177/1529100612436522>
- Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron, 1*, 3–32.
- Fong, K., Mullin, J. B., & Mar, R. A. (2013). What you read matters: The role of fiction genre in predicting interpersonal sensitivity. *Psychology of Aesthetics, Creativity, and the Arts, 7*, 270–376. <https://doi.org/10.1037/a0034084>
- Forster, M., Leder, H., & Ansorge, U. (2013). It felt fluent, and I liked it: Subjective feeling of fluency rather than objective fluency determines liking. *Emotion, 13*, 280–289. <https://doi.org/10.1037/a0030115>
- Freitas, A. L., Azizian, A., Travers, S., & Berry, S. A. (2005). The evaluative connotation of processing fluency: Inherently positive or moderated by motivational context? *Journal of Experimental Social Psychology, 41*, 636–644. <https://doi.org/10.1016/j.jesp.2004.10.006>
- Freytag, G. (1894). *Technique of the drama*. Chicago: S. C. Griggs.
- Frow, J. (2014). *Genre* (2nd ed.). New York: Routledge.
- Gerger, G., Forster, M., & Leder, M. (2017). It felt fluent but I did not like it: Fluency effects in faces versus patterns. *Quarterly Journal of*

- Experimental Psychology*, 70, 637–648. <https://doi.org/10.1080/17470218.2016.1145705>
- Gollwitzer, P. M., & Bayer, U. (1999). Deliberative versus implemental mindsets in the control of action. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 403–422). New York: Guilford Press.
- Graf, L. K., & Landwehr, J. R. (2015). A dual-process perspective on fluency-based aesthetics: The pleasure-interest model of aesthetic liking. *Personality and Social Psychology Review*, 19, 395–410. <https://doi.org/10.1177/1088868315574978>
- Gray, K. (2018). Quantifying forward motion in naturalistic thought and its link to creativity. Paper presented at the annual meeting of the Society for Personality and Social Psychology, Atlanta.
- Greifeneder, R., Alt, A., Bottenberg, K., Seele, T., Zelt, S., & Wagener, D. (2010). Processing fluency systematically biases evaluations of handwritten material. *Social Psychological and Personality Science*, 1, 230–237. <https://doi.org/10.1177/1948550610368434>
- Grey, O. (2016). *Monsters from the vault*. Vancouver: Innsmouth Free Press.
- Halberstadt, J. (2006). The generality and ultimate origins of the attractiveness of prototypes. *Personality and Social Psychology Review*, 10, 166–183.
- Halberstadt, J., & Rhodes, G. (2000). The attractiveness of nonface averages: Implications for an evolutionary explanation of the attractiveness of average faces. *Psychological Science*, 11, 285–289.
- Herrmann, J. B., van Dalen-Oskam, K., & Schöch, C. (2015). Revisiting style, a key concept in literary studies. *Journal of Literary Theory*, 9, 25–52. <https://doi.org/10.1515/jlt-2015-0003>
- Hickey, W. (2015). Be suspicious of online movie ratings, especially Fandango's. *FiveThirtyEight*. Retrieved from <https://fivethirtyeight.com/features/fandango-movies-ratings/>. Accessed 21 Mar 2018
- Holbrook, M. B. (1999). Popular appeal versus expert judgments of motion pictures. *Journal of Consumer Research*, 26, 144–155.
- Hsu, G. (2006). Jacks of all trades and masters of none: audiences' reactions to spanning genres in feature film production. *Administrative Science Quarterly*, 51, 420–450. <https://doi.org/10.2189/asqu.51.3.420>
- Hwang, T. G., Park, C. S., Hong, J. H., & Kim, S. K. (2016). An algorithm for movie classification and recommendation using genre correlation. *Multimedia Tools and Applications*, 75, 12843–12858. <https://doi.org/10.1007/s11042-016-3526-8>
- Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and Cognition*, 7, 265–290. <https://doi.org/10.1017/langcog.2014.30>
- Iliev, R., Hoover, J., Dehghani, M., & Axelrod, R. (2016). Linguistic positivity in historical texts reflects dynamic environmental and psychological factors. *Proceedings of the National Academy of Sciences*, 113, E7871–E7879.
- Ireland, M. E., Davis, T., Schumacher, J., & Pennebaker, J. W. (2018). Experts' and naïve participants' fictional dialog reveals individual differences in perspective-taking. Manuscript submitted for publication.
- Johnson, B. K., & Rosenbaum, J. E. (2018). (Don't) Tell me how it ends: Spoilers, enjoyment, and involvement in television and film. *Media Psychology*, 21, 582–612. <https://doi.org/10.1080/15213269.2017.1338964>
- Kaspar, K., Wehlitz, T., von Knobelsdorff, S., Wulf, T., & von Saldern, M. A. O. (2015). A matter of font type: The effect of serifs on the evaluation of scientific abstracts. *International Journal of Psychology*, 50, 272–278.
- Katzir, T., Hershko, S., & Halamish, V. (2013). The effect of font size on reading comprehension on second and fifth grade children: Bigger is not always better. *PLoS ONE*, 8, e74061. <https://doi.org/10.1371/journal.pone.0074061>
- Labroo, A. A., Dhar, R., & Schwarz, N. (2007). Of frog wines and frowning watches: Semantic priming, perceptual fluency, and brand evaluation. *Journal of Consumer Research*, 34, 819–831.
- Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... Jebara, T. (2009). Life in the network: The coming age of computational social science. *Science*, 323, 721.
- Leavitt, J. D., & Christenfeld, N. J. S. (2011). Story spoilers don't spoil stories. *Psychological Science*, 22, 1152–1154. <https://doi.org/10.1177/0956797611417007>
- MacWhinney, B. (2015). Language development. In L. Liben & U. Müller (Eds.), *Handbook of child psychology and developmental science: Vol. 2. Cognitive processes* (pp. 296–338). New York, NY: Wiley.
- Malin, J. J., Vine, V. J., Stanton, A., Cannava, K., Bodie, G., & Pennebaker, J. W. (2014). The arc of narrative: Using language markers to identify stories. Poster presented at the 15th annual meeting of the Society for Personality and Social Psychology, Austin.
- Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, 3, 173–192. <https://doi.org/10.1111/j.1745-6924.2008.00073.x>
- Mehl, M. R. (2017). The Electronically Activated Recorder (EAR): A method for the naturalistic observation of daily social behavior. *Current Directions in Psychological Science*, 26, 184–190.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... Pinker, S. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331, 176–182. <https://doi.org/10.1126/science.1199644>
- Mittell, J. (2006). Narrative complexity in contemporary American television. *Velvet Light Trap*, 58, 29–40.
- Mixer, L. (2018). "And then they boned": An analysis of fanfiction and its influence on sexual development (Unpublished master's thesis). Humboldt State University, Arcata, California.
- Nalbandian, T., Iserman, M., & Ireland, M. E. (2018). *The narrative arc of film scripts and their relation to audience and critic preferences*. Manuscript submitted for publication.
- Nielsen (2017). *Nielsen estimates 119.6 million TV homes in the U.S. for the 2017–18 TV season*. Retrieved from <http://www.nielsen.com/us/en/insights/news/2017/nielsen-estimates-119-6-million-us-tv-homes-2017-2018-tv-season.html>. Accessed 4 Oct 2018
- Nunes, J. C., Ordanini, A., & Valsesia, F. (2015). The power of repetition: Repetitive lyrics in a song increase processing fluency and drive market success. *Journal of Consumer Psychology*, 25, 187–199. <https://doi.org/10.1016/j.jcps.2014.12.004>
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Science*, 12, 237–241.
- Pakhomov, S., Chacon, D., Wicklund, M., & Gundel, J. (2011). Computerized assessment of syntactic complexity in Alzheimer's disease: A case study of Iris Murdoch's writing. *Behavior Research Methods*, 43, 136–144. <https://doi.org/10.3758/s13428-010-0037-9>
- Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108, 934–952. <https://doi.org/10.1037/pspp0000020>
- Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, 9, e115844. <https://doi.org/10.1371/journal.pone.0115844>
- Pennebaker, J. W., & Ireland, M. E. (2011). Using literature to understand authors: The case for computerized text analysis. *Scientific Study of Literature*, 1, 34–48. <https://doi.org/10.1075/ssol.1.1.04pen>

- Perry, G. (2018). Real-life lord of the flies. *New Scientist*, 237, 41–43.
- Pocheptsova, A., Labroo, A. A., & Dhar, R. (2010). Making products feel special: When metacognitive difficulty enhances evaluation. *Journal of Marketing Research*, 47, 1059–1069.
- Ramos, M., Calvão, A. M., & Anteneodo, C. (2015). Statistical patterns in movie rating behavior. *PLoS ONE*, 10, e136083. <https://doi.org/10.1371/journal.pone.0136083>
- R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Retrieved from <https://r-project.org>
- Reber, R., Schwarz, N., & Winkielman, P. (2004a). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8, 364–382.
- Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, 9, 45–48. <https://doi.org/10.1111/1467-9280.00008>
- Reber, R., Wurtz, P., & Zimmermann, T. D. (2004b). Exploring "fringe" consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition*, 13, 47–60. [https://doi.org/10.1016/s1053-8100\(03\)00049-7](https://doi.org/10.1016/s1053-8100(03)00049-7)
- Rocklage, M. D., & Fazio, R. H. (2015). The Evaluative Lexicon: Adjective use as a means of assessing and distinguishing attitude valence, extremity, and emotionality. *Journal of Experimental Social Psychology*, 56, 214–227.
- Rosenberg, R. S. (2013). *Superhero origins: What makes people tick and why we care*. n.p.: CreateSpace.
- Russell, N. J. C. (2011). Milgram's obedience to authority experiments: Origins and early evolution. *British Journal of Social Psychology*, 50, 140–162.
- Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, 61, 195–202. <https://doi.org/10.1037/0022-3514.61.2.195>
- Seufert, T., Wagner, F., & Westphal, J. (2017). The effects of different levels of disfluency on learning outcomes and cognitive load. *Instructional Science*, 45, 221–238. <https://doi.org/10.1007/s11251-016-9387-8>
- Shevy, M. (2008). Music genre as cognitive schema: Extramusical associations with country and hip-hop music. *Psychology of Music*, 36, 477–498. <https://doi.org/10.1177/0305735608089384>
- Simonton, D. K. (2009). Cinematic success criteria and their predictors: The art and business of the film industry. *Psychology and Marketing*, 26, 400–420.
- Simonton, D. K. (2011). *Great flicks: Scientific studies of cinematic creativity and aesthetics*. New York: Oxford University Press.
- Song, H., & Schwarz, N. (2008). Fluency and the detection of misleading questions: Low processing fluency attenuates the Moses illusion. *Social Cognition*, 26, 791–799.
- Stuart-Smith, J., Pryce, G., Timmins, C., & Gunter, B. (2013). Television can also be a factor in language change: Evidence from an urban dialect. *Language*, 89, 501–536. <https://doi.org/10.1353/lan.2013.0041>
- Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, 23, 457–482.
- Thompson, K. (1999). *Storytelling in the new Hollywood*. Cambridge: Harvard.
- Trujillo, L. T., Jankowitsch, J. M., & Langlois, J. H. (2014). Beauty is in the ease of the beholding: A neurophysiological test of the averageness theory of facial attractiveness. *Cognitive, Affective, & Behavioral Neuroscience*, 14, 1061–1076. <https://doi.org/10.3758/s13415-013-0230-2>
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207–232. [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9)
- Wanderer, J. J. (1970). In defense of popular taste: Film ratings among professionals and lay audiences. *American Journal of Sociology*, 76, 262–272.
- Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R., (2008). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 195–223). Mahwah: Erlbaum.
- Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112, 1036–1040.
- Zimmermann, J., Brockmeyer, T., Hunn, M., Schauenburg, H., & Wolf, M. (2017). First-person pronoun use in spoken language as a predictor of future depressive symptoms: Preliminary evidence from a clinical sample of depressed patients. *Clinical Psychology and Psychotherapy*, 24, 384–391.