Title page with all author information

The Textual Features of Fiction that Appeal to Readers: Emotion and Abstractness

Marta M. Maslej[1]
Raymond A. Mar[2]
Victor Kuperman[3]

Word count: 8,464

[1]Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4L8
[2]Department of Psychology, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3
[3]Department of Linguistics and Language, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4L8

Address correspondence to:
Marta M. Maslej
Department of Psychology, Neuroscience & Behaviour,
McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4L8
Email: maslejmm@mcmaster.ca

Abstract

In two studies we assessed whether simple surface-level textual features of fictional works predict their appeal to readers. Specifically, we focused on the potential role of emotional valence and arousal, the presence of affective themes (i.e., sex, religion, and death), and word concreteness and frequency. In Study 1, we compiled a corpus of 1,471 short stories posted to a creative writing website and along with ratings of these stories by readers. Using word lexicons we quantified the textual features of these stories, and then examined whether any of these features predicted ratings using correlation and mixed-effects regression. The presence of abstract, negatively-valenced, and commonly-used words predicted higher story ratings, with sexual content having a negative impact on ratings. In Study 2, we extended this investigation by examining how textual features relate to appeal for one central aspect of fiction: its characters. A sample of 207 undergraduates wrote brief character sketches, and another group of students rated the depicted characters based on interest, likeability, and complexity (n = 144). As revealed in correlations and mixed-effects regressions, characters described with abstract, negative, and arousing words tended to be rated as more interesting and complex. In contrast, characters described with positive and calm words tended to be more likeable. Our results succeed in tying specific surface-level textual features to the appeal of stories and characters, highlighting the importance of affect words and abstractness.

(230 words)

Keywords: short stories; characters; emotion; abstractness; corpus linguistics.

Introduction

Examining works of fiction has typically been the domain of literary criticism. However, over the past several decades, psychologists and linguists have begun to examine the nature of texts using quantitative and computational methods (Mahlberg, 2007). For example, computational advances now allow for the analysis of large corpora of texts, such as examining textual features to study the characteristics of a given genre, writer, or literary period (Semino & Short, 2004). One aspect of fiction is its appeal, which has been an important consideration in literary criticism. To date, however, the question of why some works of fiction are more appealing than others has not often been addressed with quantitative methods. There are likely many features of fiction that engage readers, and these also likely depend on contextual factors as well as individual factors (e.g., motivation, expectations). Despite the complexity of literary appeal, there may be observable consistencies in which textual features predict appeal, across many different texts and readers. The current set of studies examines whether surface-level textual features—such as emotional words and themes, word concreteness, and word frequency—predict the appeal of short stories and fictional characters.

Emotional Words & Themes

Emotion is central to the experience of narrative fiction (Frijda, 1989; Mar, Oatley, Djikic & Mullin, 2011; Tan, 1995), with researchers arguing that good stories must succeed in both arousing and resolving readers' emotions (Brewer & Lichtenstein, 1982). Accordingly, references to any objects or events that elicit primary emotions—such as surprise, fear, happiness, and disgust—are thought to be interesting (Schank, 1979). However, certain types of emotional responses may be more or less related to literary appeal, such as valence (i.e., positive or negative emotions) and arousal (i.e., intensity of emotion) (Lang, Greenwald, Bradley &

Hamm, 1993).

In terms of emotional valence, it seems intuitive that readers would enjoy stories that elicit positive responses, such as feeling happy or pleased (Bryant & Zillmann, 1984; Zillmann & Cantor, 1977). Individuals tend to seek out media content to manage their moods, listening to joyful music or watching comedies when they wish to feel better, for example (Dillman-Carpentier et al. 2008; Knobloch & Zillmann, 2002). Based on this research, stories that include references to positively-valenced content may be appealing because they can improve a bad mood or maintain a good one.

The relation between media and emotion is not so simple, however, as people can also enjoy media that elicits negative emotions. Iguarta (2010) and Oliver (1993) found that when participants elected to view a dramatic or sad film, the more negatively they felt while watching the film, the more they reported enjoying it. This has been observed for various forms of art and media, and this seeming paradox has generated a wealth of discussion among researchers (Eerola et al., 2018; Menninghaus et al., 2017; Oliver & Raney, 2011). Within literature, the notion that negative emotional content can be enjoyable or appealing seems to be reflected in the prototypical content of stories. Conflict (in either the domains of love or politics) is a universal property of narratives found across different cultures and geographies (Hogan, 1997). Novelists also tend to elaborate on negative rather than positive events, since stories that lack tension or conflict are considered boring and unsuccessful (Fiedler, 1982). Accordingly, literary writers appear to be preoccupied with negative emotions. A linguistic analysis of interviews with eminent writers found that they used more negative emotion words when describing their work, as compared to eminent physicists (Djikic, Oatley & Peterson, 2006). Negatively-valenced story content thus appears to be prevalent and may be experienced as enjoyable or appealing, when

readers are in the right mood for this type of content.

In addition to valence, the intensity or level of arousal for elicited emotions may influence a story's appeal. In stories, highly arousing experiences like violence or sex, or arousing topics like religion and death, are hypothesized to be particularly interesting (Kintsch, 1980). One study into this possibility measured emotional arousal by way of heart-rate variability (Wallentin et al., 2011). Participants rated the emotional valence and arousal of passages while reading the fairytale, *The Ugly Duckling*. Based on these ratings, the researchers identified the most emotionally intense parts of the story, both positive and negative in valence (e.g., the duckling being bullied, the duckling turning into a swan). These emotionally intense passages were associated with increased heart rate variability for participants who listened to the story, suggesting that both positive and negative events can be intense or arousing. Furthermore, there is some evidence that arousal relates to a story's appeal. Nell (1988) noted that when participants read story passages they reported enjoying, they exhibited increases in heart rate and respiration (although the effect was small). Importantly, arousal is frequently confounded with valence, since negative content tends to be more arousing than positive content (Anderson et al., 2003; Bradley, Codispoti, Cuthbert, & Lang, 2001). For this reason, it is important to examine the unique effects of valence and arousal when investigating the appeal of emotional content in works of narrative fiction.

Word-Level Features

In addition to emotional content, the appeal of fiction might depend on word-level features, such as word concreteness. To examine the characteristics of high-quality poetry, Kao and Jurafsky (2012) compared the textual features of poems written by professional poets with those by amateurs. The use of concrete words was among the most reliable predictors of poems

written by professionals. This finding is consistent with advice frequently disseminated to writers to "show not tell" (i.e., to demonstrate through action and example, rather than baldly stating a thesis or idea). It also reflects the degree to which Imagism has influenced poetry (i.e., using concrete images to convey deeper meanings; Kao & Jurafsky, 2012; Perkins, 1976; Pound, 1913). In fiction, the presence of concrete words could indicate that the narrative is being developed through the description of events, allowing readers to generate their own impressions of characters and interpretations of the plot. By way of contrast, the presence of abstract words suggests that generalizations and interpretations are provided for the reader. Since clichéd writing is thought to contain generalizations, too much abstract content might result in unoriginal or uninteresting stories (Burroway, 2007; Kao & Jurafsky, 2012). Furthermore, investigations into the effects of word concreteness on memory show that abstract words are harder to recall, perhaps because concrete words facilitate imagery, which in turn aids memorability (Paivio et al., 1966; Walker & Hulme, 1999). The imagery generated by concrete language might also be particularly pleasurable, helping readers to become transported into the story (Gerrig, 1993; Green & Brock, 2000; Green, Brock, & Kaufman, 2004).

The use of unique or distinctive language might also contribute to a story's appeal. This type of language may be captured by word frequency, or how frequently a word is used within a language. In the case of poetry, the presence of original or uncommon words does not appear to be an important predictor of high-quality texts, since poems using low-frequency (less common) words are not any more likely to be written by professional poets than amateurs (Kao & Jurafsky, 2012). However, in a story, the presence of less frequent words could indicate that a writer is using language creatively, portraying the characters and plot events in a unique way. Indeed, the creative use of language in short stories seems to capture attention and impact

aesthetic responses, making the stories appear more striking and emotionally evocative (Miall & Kuiken, 1994). These effects might emerge because novel or distinctive stimuli (including words) tend to attract attention and elicit emotional arousal (Sokolov, 1963; Johnston, Hawley, Plewe, Elliott, & DeWitt, 1990). Given the hypothesized importance of emotion in fiction (e.g., Brewer & Lichtenstein, 1982), the presence of uncommon or less frequent words in narrative fiction might contribute to its appeal.

The Present Study

The current set of studies uses corpus linguistic methods to determine whether simple textual features that reflect some possible qualities of good fiction are related to the appeal of short stories and fictional characters (i.e., emotional valence and arousal, affective themes, and concrete or uncommon words). In Study 1, we collected a corpus of short stories submitted to an amateur writing website and used word lexicons to quantify the presence of these textual features in these short stories. Ideal for our purposes, users of this site also provided ratings for the submitted stories. To assess whether the textual features reflect content that readers find appealing, we examined their ability to predict story ratings.

In Study 2, we sought to replicate and extend our findings by focusing on characters, which are central to works of fiction (Gardner, 1964). Participants created brief character sketches, or descriptions of characters, which were then rated on three dimensions: liking, interest, and complexity. We examined whether the textual features of these character sketches could predict these ratings, focusing again on emotional valence and arousal, word concreteness, and word frequency.[1]

---

[1] The personality traits for these participants, and the relation between their characteristics (i.e., reading and writing habits, personality, and perspective-taking tendencies) and the ratings of the characters they created, have been reported elsewhere.

Study 1 Methods

Short Story Corpus

We constructed a corpus of short stories from the same writing website used by Kao and Jurafsky (2012) to construct their corpus of amateur poetry (http://www.amateur-writing.com/stories[2]). Because this study relied on analyzing archival data posted on a public website, we did not seek research ethics approval. At the time of our search, the website had 1471 standalone short stories available. We included any story from any genre that was between a pre-determined range of 1000–2600 words, had a rating available, and was read by at least 40 people. A sample of 172 stories met this selection criteria, containing a total of 96,872 words. The stories covered a variety of genres, including action, adventure, comedy, drama, erotic, fantasy, horror, mystery, romance, and science fiction. Stories had a mean length of 1648.40 words (*SD* = 413.04), and ranged from 1001–2600 words. Each story was read by a mean number of 128 people (*SD* = 160.66), with a range of 41–1070 readers. Stories were rated on a scale from 1–5 stars, with more stars indicating more favourable evaluations. The website did not provide instructions on how to rate stories, so we could not assess what criteria readers were using (e.g., compelling characters, interesting plot, or high-quality writing). However, we presume this rating reflects a global evaluation (e.g., akin to that for GoodReads, IMDB, or Yelp!). The mean rating was 4.01 (*SD* = 0.86, Range = 1–5). There were no restrictions on who could post, read, and rate stories, and so we could not guarantee that stories were only rated once by a reader.

Textual Feature Lexicons

We applied six lexicons to assess the textual features of the stories. Overall sentiment

---

[2] This website is no longer online, but an archive of it can be accessed and browsed at: https://web.archive.org/web/20150511092213/http://www.amateur-writing.com/stories/

was assessed using a database of 14,000 English words that were rated separately for valence and arousal using a 9-point scale (Warriner, Kuperman, & Brysbaert, 2013). Ratings were provided by online samples, crowdsourced from anywhere in the United States on Amazon's Mechanical Turk. To assess valence, participants rated a set of words on the degree to which the words made them feel positive emotions on one end (e.g., happy, pleased) and negative emotions on the other (e.g., unhappy, annoyed). Words rated highest on valence were "happiness" and "fantastic", with the lowest-rated words being "pedophile" and "rapist". For arousal, a different set of participants indicated the degree to which the words made them feel excited or uncontrolled on one end and calm or bored on the other. Words rated highest on arousal were "insanity" and "rampage", with the lowest-arousal words being "grain" and "dull".

To determine whether the concreteness of words in the stories contributed to their appeal, we used concreteness ratings compiled from over 37,000 words (Brysbaert, Warriner, & Kuperman, 2014). These ratings were also crowdsourced from Amazon's Mechanical Turk. Participants rated words on a 5-point scale, indicating the degree to which the words were concrete (i.e., experience-based, or directly experienced through the senses) or abstract (i.e., language-based, or having meanings that can be defined but not directly experienced). Words scoring highest on concreteness included "apple" and "daisy", and some of the lowest-scoring words were "essentialness" and "spirituality". To assess word frequency for evidence of lexical diversity, we used frequency scores for 51 million words compiled from an online corpus of film and television program scripts (Brysbaert & New, 2009). Lowest frequency words from this lexicon included "adages" and "ornerier", with some of the highest frequency words being "is" and "have".

Finally, to evaluate whether stories containing arousing events or controversial topics

would be more compelling, we used three categories from the Linguistic Inquiry and Word Count software (LIWC; Pennebaker et al., 2001). The LIWC is a textual analysis program that includes over 4500 words related to affective, social, and cognitive processes. For this study, we examined the categories for sex, religion, and death. Each word from each story received a binary score based on whether it belonged to each of these categories, as identified by the LIWC.

Statistical analyses

First, we examined correlations between each story's rating and its textual features. Because score distributions were nonparametric, we used Spearman rank-order correlations. To determine the unique effects of the textual features on ratings, we fitted a linear mixed-effects regression model. With ratings as our response variable, we examined valence, arousal, concreteness, frequency, and LIWC category (i.e., sexual, religion, or death) as predictors. In our model, we also included story length and number of readers, to control for their effects on ratings. To account for the same words having the same valence, arousal, concreteness, frequency, and LIWC membership scores, we specified word as a random effect in our model. After the initial model was fitted, we removed data points with scaled residuals that were more than 2.5 standard deviations away from the predicted values (approximately 2% of the data), to attenuate the influence of outliers. To gauge effect sizes, we plotted partial effects for each predictor (or fixed effect). These plots are provided in Supplementary Section A (which can be accessed at https://mfr.ca-1.osf.io/render?url=https://osf.io/kd8sq/?action=download %26mode= render). By way of effect-size, we report a measure of the difference in the value of the rating variable when it is estimated for the smallest and largest values of each predictor, which indicates how a maximal change in the predictor in a given word affects the rating. Analyses

were conducted using the lmer package (Bates, Maechler, Bolker, & Walker, 2015) in the statistical platform R (R Core Team, 2015).

Results

Stories had a mean valence rating of 5.77 (*SD* = 1.20; with higher scores indicating a more positive emotion), a mean arousal rating of 3.95 (*SD* = 0.95; with higher scores indicating greater arousal), mean concreteness rating of 3.15 (*SD* = 1.12), and a mean frequency score of 57,097.69 (*SD* = 97,390.38; the number of times a word appears in the corpus, with higher scores indicating a higher frequency). Across all stories, there were 1318 words related to sex, 537 words related to religion, and 492 words related to death.

The textual features were then correlated with one another (Table 1). Stories with more negatively-valenced words tended to contain more arousing and concrete words, and less frequently-occurring words. They also contained more words related to death. In turn, the presence of words about death in stories was related to the presence of religious words. Stories with more arousing words contained more concrete words, less frequently-occurring words, and more words related to affective themes (i.e., sex, religion, death). Finally, stories with more concrete words tended to have less frequently-occurring words. These correlations suggest that it is useful to examine the unique effects of the textual features on ratings, controlling for the influence of the other features.

Correlations between Textual Features and Ratings

Correlations between story ratings and textual features are provided in the first row of Table 1. Stories that had higher ratings of quality or appeal tended to have less concrete words. There were also trends for higher-rated stories to be longer (p = .083) and to contain more negatively-valenced emotion words (p = .062), and less words related to sex (p = .058). The

number of readers, word frequency and arousal, and the presence of words about religion and death were not strongly related to story ratings.

Unique Effects of Textual Features on Ratings

In order to examine the unique contribution of these different textual features, we conducted a mixed-effects regression. Table 2 provides the estimates, standard errors, confidence intervals, *p*-values, and effect sizes for each predictor (fixed effect) included in our linear mixed-effects model. Stories that were more highly rated were longer ($\beta = .13$, $SE = .00$, 95% CI [.13, .14], $p < .001$) and they also tended to be read by fewer people, $\beta = -.02$, $SE = .00$, 95% CI [-.03, -.02], $p < .001$.

From our predictors of interest, several textual features were related to the story ratings, including valence, concreteness, frequency, and the presence of sexual words. Stories with more negatively-valenced words were given higher ratings, $\beta = -.02$, $SE = .00$, 95% CI [-.03, -.02], $p < .001$. Stories also had higher ratings if they contained less concrete words ($\beta = -.04$, $SE = .01$, 95% CI [-.05, -.03], $p < .001$) and more frequently-occurring words, $\beta = .03$, $SE = .01$, 95% CI [.01, .04], $p < .001$. Stories received lower ratings if they contained more sexual words, $\beta = -.01$, $SE = .01$, 95% CI [-.02, -.00], $p = .023$. However, the other affective themes (i.e., religion, death) and the arousal of words did not contribute to story ratings.

Study 1 Discussion

The results of Study 1 suggest that, in terms of emotion, readers find negatively-valenced content more appealing in a short story. This supports the theory that tension, conflict, and other negative events make for compelling stories (Fiedler, 1982; Hogan, 1997). It is also consistent with research showing that viewers can enjoy films that elicit negative emotions (Iguarta, 2010; Oliver, 1993). However, the zero-order correlation between negatively-valenced words and

ratings was weak and it did not pass our threshold for statistical significance, so it should be treated provisionally, especially given the number of tests. Additionally, the unique effect of negatively-valenced content on ratings was small. A maximum decrease in the valence of a given word uniquely contributed to just a 3% increase in a story's rating (Table 2). In other words, the story's rating would increase by only 0.15 of a point if we the valence of one of its words changed from 1 to 9. However, this effect is for a change regarding a single word, and effects would likely accumulate across several words and possibly interact within the context of a story's theme.

Surprisingly, the arousal level of words did not impact ratings. When it came to emotional themes, only the presence of words related to sex had an effect, resulting in lower ratings. To interpret this finding, it is likely important to consider the characteristics of the readers (e.g., age, cultural values), as well as their expectations. Some readers may be offended by sexual content or find it distasteful. Unfortunately, one limitation of this study is that ratings were provided anonymously, so there is no demographic information available about the readers or the criteria they used when evaluating the stories. Another possibility for this result is that poor writers might rely on salacious content to attract and hold reader interest, but this proves to be an unsuccessful strategy.

Another surprising finding is that less concrete word usage predicted higher ratings for stories, which was the opposite of what we predicted. In Kao and Jurafsky's (2012) study, professional poets tended to use concrete words more often compared to amateur poets. These researchers concluded that professional poets use concrete words to facilitate imagery, since these words evoke sensory representations; these images would show, rather than tell, the abstract conceptions being communicated (Kao & Jurafsky, 2012). In other words, skilled poets

resist the temptation to generalize or state their abstract ideas explicitly, which may set their poetry apart from amateur poetry. The discrepancy between these findings and our own could be explained in at least two separate ways. First, concrete detail may be especially valued in poetry relative to fiction, since poems are meant to provide brief snapshots of a subject or experience, rather than develop a narrative plot. In contrast, stories may better accommodate more than one approach to communicating ideas, including both demonstrations through concrete exemplars and abstract generalizations. The appeal of concrete language may therefore depend on genre: concreteness may be more appealing in a poem, but less appealing in a story. Alternatively, or in addition, the appeal of concrete language might depend on audience factors: the characteristics of the readers. In Kao and Jurafsky's (2012) study, the appeal of a poem was determined by the eminence of its writer and, consequently, based on critical acclaim from the literary community. In contrast, our short stories were rated by an online community of readers and writers. Concrete language might hold a different appeal for literary critics compared to a more amateur or lay audience. Perhaps people who read for leisure or entertainment (e.g., those visiting the website) prefer to be provided with abstractions, generalizations, and conclusions, which could involve less challenge or cognitive effort. A future study involving fictional and poetic texts, as well as ratings from literary experts and non-experts, might be useful to uncover whether either, both, or neither of these two explanations is true.

Consistent with the notion that less challenging stories are more appealing for this particular audience, the use of more frequent words uniquely predicted higher ratings. The presence of frequently-occurring words is a feature often used to gauge the difficulty or readability of a text (Breland, 1996), suggesting that readers in our study may have evaluated more easily readable stories more favourably. Alternatively, word frequency may be indicative

of a writer's skill, with better writers discarding unnecessary or superfluous language in favor of writing simply and concisely[3]. However, the effect of word frequency on ratings was small (with individual words having a maximum effect of less than 3%), and this association did not emerge when examining zero-order correlations.

Evidence of a positive association between word frequency and story ratings also seems inconsistent with our prediction that using language in unique or creative ways will contribute to a story's appeal. However, using language creatively may not necessarily involve using novel or uncommon words. Instead, it might mean using common words in uncommon ways, deviating from phonetic, grammatical, or semantic conventions (Mukařovský, 1976). The appeal of phonetic, grammatical, or semantic deviations were not captured in our current analysis. Although this topic has been examined empirically in various ways (e.g., Miall & Kuiken, 1994), it remains a future direction for corpus-analytic methods.

Some findings from Study 1 are difficult to interpret because we have no information about our readers, their motivations for reading a given story, and their rating criteria. Some visitors to this website may have been reading and rating stories not for leisure or entertainment, but simply to support or critique the work of friends or peers. Alternatively, they may have been amateur writers as well, reading stories to gain inspiration for their own work. Given these possibilities, the website ratings may not reflect what a general sample of readers finds appealing. These potential shortcomings were addressed in our Study 2 by collecting our own sample of writers and raters and establishing multiple rating criteria.

Study 2

[3] We thank an anonymous reviewer for pointing out that the process of becoming a good writer involves discarding unnecessary vocabulary and prose and that skilled writers may use more concise language.

In Study 2, we investigated whether textual features can predict a fictional work's appeal in a controlled laboratory setting. We restricted the time and resources of writers and the content of their writing, by designing a task that focuses on creating one important element of a story: a character. Arguably, the most important feature of a fictional narrative is its characters (Hogan, 2003; Miall, 1988). In Flaubert's five-stage theory of creative writing, the first step is creating an initial character sketch (Oatley & Djikic, 2008). This sketch, or outline or plan, is a foundational step in a character's depiction. In this study, we asked a sample of undergraduate students to create a character sketch by writing a description of a character, based on a portrait photograph of a man, in five minutes. After the character sketches were created, an entirely different group of students were recruited to rate the characters depicted in the sketches. This helps address concerns with Study 1, regarding unknown rater motivations. By recruiting known samples of writers and raters, it becomes unlikely that the sketches are being read and rated to provide support for peers or as a source of inspiration.

In this study, we also improved on Study 1 by employing pre-determined rating criteria, both decomposing appeal into separate components and providing descriptions of how these components should be rated. Students rated the sketches on three dimensions: interest (the character's ability to capture attention), likeability (feeling warmth toward the character), and complexity (whether the character is multidimensional and realistic). We included these various rating dimensions to capture different aspects of characters that could appeal to readers. Even if characters were not likeable, they could still elicit interest or be compelling, which is often the case for villains or anti-heroes (Ruiz, 2015). Interest in these types of characters might stem from different sources, including a desire to feel better about one's own flaws (Mares & Cantor, 1992). Villains or anti-heroes could also be interesting or compelling because they explore dark

or forbidden aspects of our own personalities, or of human nature. Gaining insight into all of human nature, even its dark facets, may be one motivation for engaging with narratives (Nabi, Finnerty, Domschke & Hull, 2006; Oliver & Raney, 2011; Zillmann, 2000).

To replicate and extend Study 1, we examined whether simple textual features of these character sketches predicted the depicted characters' interest, likeability, and/or complexity. Since characters are central to works of fiction (Gardner, 1964), we anticipated that some of the predictors of story ratings uncovered in Study 1 might also predict character ratings here in Study 2. Specifically, we expected that characters described with more negative words would be more interesting, but perhaps less likeable. This finding would provide evidence consistent with Study 1 that negative content is appealing to readers, but clarify that this appeal is related to interest rather than some other aspect of appeal.

Given that characters are just one element of fictional works however, we anticipated possible differences in findings between Studies 1 and 2. In terms of concreteness, the development of characters in a short sketch might depend on generalizations or abstractions of the character's personality and behaviour. On the other hand, in a sketch, characters may be better conveyed using concrete language. Sketches could resemble poetry in that they provide brief snapshots of a character, rather than developing a plot or narrative. The depicted characters might therefore appear more interesting, likeable, or complex if the sketches refer to concrete objects or behaviours, describing where the character lives and what he does for example. Due to these competing possibilities, we did not make a directional prediction about how concrete language would affect character ratings.

Methods

Study 2 had two stages. In the first stage, one set of participants created written sketches

of characters. In the second stage of the study, a different set of participants rated the depicted characters with respect to how interesting, likeable, and complex they were.

Participants. Participants in both stages of the study were undergraduate students enrolled in an introductory psychology course at a university in Ontario. A sample of 207 students completed character sketches (142 females). Their mean age was 20.68 (*SD* = 4.18). In the second stage of the study, a different sample of 247 participants (recruited from the same introductory course, but two years later) rated a subset of the depicted characters in the second stage of the study. Before any data analysis was performed, we removed data from 103 raters because they were identified as inattentive responders. To determine inattentive responding, we included two fake character sketches with specific rating instructions embedded within the text (e.g., "respond with a '3' for the ratings corresponding to this description", akin to the Conscientious Responders Scale; Marjanovic, Struthers, Cribbie, & Greenglass, 2014). When raters did not follow the instructions for either of these fake sketches, we inferred that they were not reading the sketches carefully and we excluded their data (for more details, see Maslej, Oatley & Mar, 2017). With these participants excluded, our final sample for the second stage included 144 raters (114 female) with a mean age of 19.64 (*SD* = 4.08). Each character description was rated by a mean number of 20.6 individuals (*SD* = 2.73), ranging from 13–28 raters.

Materials

Character creation task. Participants in the first stage of the study were given a head-and-shoulders photograph of a man. Based on this photograph, they were instructed to create an imaginary person in the same way a writer creates a fictional character. They were asked not to copy an existing character (e.g., that they have seen on television or read about in a book). Each

participant had five minutes to type out a description of this character in a text box. To begin, 205 participants were asked to complete this task. However, character descriptions were missing for nine participants due to computer error or because they did not complete the task, leaving a final sample of 196 character sketches.

Perceptions of characters. Participants in the second stage of our study rated the character sketches produced in the first stage on interest, likeability, and complexity. Each rater read a random subset of 28 character depictions from the 196 eligible sketches generated by first-stage participants. Characters were rated on the three dimensions, each measured with a single question (e.g., "How interested are you in this character?"). Each question was accompanied by three sub-questions placed in parentheses, to elaborate on what participants should consider when rating that dimension (e.g., for the interest item: "Would you be interested in reading a book or watching a movie that had this character in it?"). Briefly, the interest dimension included sub-questions about whether raters wanted to know or read more about the character, and whether the character captured their attention. The likeability sub-questions asked raters whether they felt warmth toward the character and wanted to spend time with him. The complexity sub-questions asked about how easily raters could imagine the character, and whether the character was detailed and multifaceted. Raters provided one response to each dimension on a scale from 1 (*Not at all*) to 7 (*Very much*).

Word Lexicons. To quantify the textual features of these character sketches, we used the same word lexicons from Study 1. We did not assess LIWC categories related to religion, death, and sex, since brief character sketches were unlikely to contain these themes the way that a complete story would.

Procedure

Both stages of the study were approved by the university's research ethics board. We conducted the first stage of our study in the laboratory. Participants read and agreed to a consent form outlining the nature of the tasks and questionnaires, and then they completed the character creation task and other questionnaires about their traits (unrelated to the present study[1]). After completing the study, participants were debriefed and given course credit for participating.

Participants in the second stage read and agreed to a consent form, and then completed the character ratings in a randomized order, using an online survey. After completing the ratings, they read a debriefing form and received course credit.

Statistical analyses

We examined associations between the textual features of character sketches and the three rating dimensions using nonparametric Spearman rank-order correlations to account for skewness in the data. To examine any unique effects of textual features on the character dimensions, we constructed three linear mixed-effects regression models with interest, likeability, and complexity ratings as responses, and valence, arousal, concreteness, and frequency as predictors. In each model, we controlled for the length of sketches by including the number of words as predictors. We specified word as a random-effect in each model and removed data points with scaled residuals smaller than 2.5 (approximately 1% of the data). Plots showing the partial effects of each predictor on the three rating dimensions are provided in Supplementary Section B (https://mfr.ca-1.osf.io/render?url=https://osf.io/8n75g/?action=download%26mode=render).

Results

The character sketches had a mean length of 129.49 words (*SD* = 47.85, Range 19–270). Their mean interest rating was 3.96 (*SD* = 1.01), average likeability was 3.69 (*SD* = 1.10), and

average complexity was 4.45 (*SD* = 0.79). Words in the sketches had a mean valence score of 5.85 (*SD* = 1.21), a mean arousal of 3.96 (*SD* = 0.91), a mean concreteness of 3.03 (*SD* = 1.08), and a mean frequency of 77,458.40 (*SD* = 459,663.00). Character sketches rating highest on each dimension, as well as those scoring highest on each textual feature, are provided in Supplementary Section C (https://mfr.ca-1.osf.io/render?url=https://osf.io/a2fqd/?action=download%26mode=render).

Correlations among the textual features of the character sketches were consistent with those reported for Study 1 (correlations are provided in Table 3). Sketches with negatively-valenced words were more likely to contain arousing words, and sketches with frequently-occurring words were less likely to contain arousing and concrete words.

Correlations between textual features and character dimensions

Correlations between the textual features of character sketches and the rating dimensions are presented in the first three rows of Table 3. Longer sketches were rated as more interesting, more likeable, and more complex. None of the rating dimensions were related to word frequency.

Character sketches that were rated as more interesting tended to have more arousing words, but fewer concrete words. Consistent with Study 1, these sketches also had more negatively-valenced words, but this effect just failed to pass threshold for statistical significance (p = .053). More likeable characters were depicted with more positively-valenced words, but likeability was not related to arousal or concreteness. Finally, characters rated as more complex tended to be described using more arousing words and less concrete words. The emotional valence of words in the sketches was not related to complexity, however.

Unique Effects of Textual Features on Character Dimensions

We examined the unique effects of the textual features on character dimensions in three separate linear mixed-effects models. Estimates, standard errors, confidence intervals, *p*-values, and effect sizes corresponding to each predictor for each rating dimension are listed in Table 4.

Word frequency was not a predictor of interest, but longer character sketches tended to be judged as more interesting, $\beta = .32$, $SE = .01$, 95% CI [.30, .34], $p < .001$. In contrast, valence was a negative predictor of interest, so characters that were described with more negatively-valenced words were perceived as more interesting, $\beta = -.06$, $SE = .01$, 95% CI [-.08, -.03], $p < .001$. Arousal was also a positive predictor, with character sketches containing more arousing words being more interesting, $\beta = .03$, $SE = .01$, 95% CI [.00, .06], $p = .015$. Lastly, concreteness was a negative predictor, with character sketches containing less concrete words being more interesting, $\beta = -.04$, $SE = .02$, 95% CI [-.07, -.01], $p = .011$.

As with interest, longer texts depicted more likeable characters, $\beta = .21$, $SE = .01$, 95% CI [.18, .23], $p<.001$. Characters were rated as more likeable when they were described with more positively-valenced words ($\beta = .11$, $SE = .01$, 95% CI [.08, .14], $p < .001$) and less arousing words, $\beta = -.04$, $SE = .02$, 95% CI [-.07, -.01], $p = .019$. Word concreteness and frequency were not predictors of likeability.

Longer texts depicted more complex characters, $\beta = .41$, $SE = .01$, 95% CI [.40, .42], $p < .001$). As with interest, more complex characters were described with more negatively-valenced words, $\beta = -.03$, $SE = .01$, 95% CI [-.04, -.02], $p < .001$. Arousal was not a predictor of complexity, but concreteness was a negative predictor, such that the presence of less concrete words in the sketch was associated with higher ratings of complexity, $\beta = -.02$, $SE= .01$, 95% CI [-.04, -.01], $p = .009$. Word frequency did not predict the complexity of the depicted characters.

Discussion

The findings from Study 2 were consistent with Study 1, while also demonstrating the value of decomposing appeal into various dimensions. When it comes to emotional content in character sketches, textual features appear to have different effects depending on the particular aspect of appeal in question. Characters rated as more likeable tended to be described using positive words, with likeable characters also described with less arousing words (controlling for the other textual features). Since likeability was based on feeling warmth toward the character and wanting to spend time with him, it seems reasonable that likeable characters should contain positive, and perhaps calming, attributes. Readers might also like characters (and by extension stories) that are positive or pleasant, perhaps because these characteristics elicit a happy or pleasant mood (Forgas, 2003; Josephson, Singer, & Palovey, 1996).

In contrast, interesting characters were described using negative and arousing words. It therefore seems that readers are drawn toward characters that elicit negative emotions or possess arousing characteristics, or perhaps ones that do negative or exciting things. Although readers may not necessarily like these characters, they can still be compelled or intrigued by such characters. An attraction to negative characters has been explored in research on parasocial interaction, which refers to a tendency to interact with fictional characters as though they were real people (Tian & Hoffner, 2010). These interactions can occur with disliked or fearful characters (Dibble & Rosaen, 2011; Tian & Hoffner, 2010), perhaps because of the interest elicited by their negative and intense or arousing features.

The textual features that predicted interest were also related to complexity, with complex characters described with arousing words, and negatively-valenced content uniquely predicting complexity. These findings suggest that believable characters are multi-faceted, possessing negative or emotionally-intense attributes in addition to more typical or positive characterisitcs.

Complex characters also tended to be described using abstract words, perhaps because abstractions and generalizations convey more information about a character than concrete details, at least in a brief sketch.

Overall, the findings of Study 2 suggest that negative and abstract content is appealing in fiction, by way of eliciting interest or exhibiting complexity. This discovery was only possible after decomposing appeal into different dimensions (unlike in Study 1), proving that these dimensions are unique and dissociable. These findings are also consistent with Study 1, which is rather remarkable given the differences between the two studies. In Study 2, we focused on just one element of fiction (characters), and the texts were generated by a different population in a completely different context. The fact that we observe such consistency regardless of whether the texts involve entire stories or just characters, and regardless of whether they are generated and rated spontaneously in a naturalistic context for individual satisfaction or as part of a laboratory task, speaks to the robustness of this phenomenon. Moreover, this consistency in what predicts literary appeal also emerged despite a wide variety of different types of stories, characters, and raters.

At the same time, some of the divergent findings between our two studies offer intriguing directions for future work. For example, emotional arousal did not predict the appeal of short stories, but it was differentially related to the interest, complexity, and likeability of characters. Future research should examine why arousal plays a role in the appeal of fictional characters, but not stories in general.

General Discussion

In the current set of studies, we used a corpus linguistics approach to address the question of what makes some works of fiction more appealing than others. Specifically, we examined

whether simple, surface-level textual features of fictional works could predict their appeal to readers. In Study 1, we used an ecologically-valid sample of spontaneously generated stories from a public website, and found that negatively-valenced, abstract, and frequently-occurring words uniquely predicted greater appeal for short stories, but the arousal of words or the presence of affective themes did not. The one exception to the latter finding was the presence of sexual content, which actually resulted in more negative evaluations for stories. In Study 2, we moved to a controlled laboratory setting, focusing on one important aspect of fiction: characters. We also expanded our measurement of appeal, decomposing this into ratings of interest, likeability, and complexity. Consistent with Study 1, characters described with negative, arousing, and abstract words tended to be rated as more interesting and complex. In contrast, characters described with positive and calm words were viewed as more likeable.

In both of our studies, we found evidence of a positive relation between negatively-valenced content and more appeal for short stories and character sketches, validating past theorizing (Fiedler, 1982). In Study 1, this association emerged even though individual readers may have evaluated the stories based on different criteria, and likely had different literary preferences, motivations for reading, and expectations for these stories. The results of Study 2 additionally suggest that these positive evaluations might be based on characters being seen as more interesting or complex, but not necessarily more likeable.

There are many reasons why readers might find negative content in fiction to be interesting. This interest may be part of a more global, and robust, attentional bias toward negative stimuli (Baumeister et al., 2001), which is thought to have evolved so individuals can quickly react to aversive or threatening circumstances in the environment (Ohman, Flykt & Esteves, 2001). In addition, readers be attracted to stories that depict difficult or challenging

circumstances, because these depictions provide them with information on how best to cope with such circumstances (Nabi, Finnerty, Domschke & Hull, 2006; Zillmann, 2000). Alternatively, observing a protagonist who is in a bad situation may make readers feel better about their own circumstances (Mares & Cantor, 1992). In exploring the topic of the appeal of negative media, Oliver and Raney (2011) argue that individuals have an intrinsic need to gain insight into human nature by deriving meaning, truth, or purpose from the world around them, which includes narratives. Characters that grapple with negative experiences—such as failure, frailty, or mortality—tend to pose questions about human nature and perhaps facilitate insight into questions regarding our own existence. These types of stories may not be pleasant or enjoyable, necessarily, but they can certainly be insightful.

Future research into the appeal of negative and abstract content should examine the context and meaning of these surface-level features, perhaps in a qualitative analysis. The negatively-valenced character sketches involve a variety of negative depictions: characters that have negative traits (are boring or mean), do negative things (steal cars), and/or find themselves in negative situations (falsely accused of murder) (Supplementary Section C). A corpus linguistic approach such as ours cannot differentiate between these different depictions, and so a qualitative analysis is required to determine if they may differentially affect appeal. Any effect of the context of these negative depictions could be further complicated by the motivations of readers. Readers might find some kinds of negative content more or less appealing depending on whether they are reading stories to feel better about themselves, or to extract lessons and gain insights (Mares & Cantor, 1992; Nabi, Finnerty, Domschke & Hull, 2006; Zillmann, 2000).

Although the effect of text length was not a primary focus of our studies, it emerged as a robust predictor of story and character ratings, with readers consistently giving longer texts

higher ratings. In this case, text length might be a proxy for effort. Individuals who wrote longer stories or character sketches may have dedicated more effort to the task, developing the narrative more fully or generating a more complete representation of the character and his life. Notably, all of our regressions control for length, so these results cannot be attributed to length or effort only. Alternatively, the length of a story or character sketch might help facilitate appeal by way of increased exposure.

One strength of these two studies is that they investigate the impact of the same textual features in two very different contexts: in a naturalistic setting out in the real world and in a controlled laboratory setting. In Study 1, we compiled existing stories and ratings from a public website, so the short stories we analyzed reflect literature that is produced in a natural and realistic manner (by an online community of novice writers). Website users also provided their ratings in an entirely ecologically-valid context, choosing the stories they wished to read and rate. Furthermore, using these ratings allowed us to examine the responses of everyday readers, who tend to be underrepresented in corpus-based studies of literature. Corpus linguistic studies often examine works from eminent, critically-acclaimed, writers (McIntyre, 2012; Morini, 2016; Murphy, 2015), under the assumption that a text valued by the literary community is of high quality and should be appealing to readers (Kao & Jurafsky, 2012). When it comes to assessing literature however, the opinions of readers with and without literary training seem to diverge (Peskin, 1998). Examining the appeal of fiction therefore requires input from different types of readers, including those without literary training.

Although Study 1 possessed high ecological validity, it lacked control. We do not know, and could not control, how these writers wrote their stories, the time they allocated to their writing, where they found their inspiration, and so forth. We also cannot know how readers

decided to evaluate these stories, whether they provided more than one rating for a story, or if they were biased in some way (e.g., giving favourable ratings to friends, or only rating stories they enjoyed). Readers and raters self-selected themselves to participate on this website, and so the population they represent might be rather circumscribed (i.e., aspiring writers who participate in online forums). We were able to address some of these issues in Study 2. By conducting this study in the lab, we knew more about who our writers were, limited the time they were given to write, and provided the same single photograph to all writers to serve as inspiration. Moreover, we controlled the rating process, ensuring that all raters only provided one set of ratings per sketch and used explicit instructions for the criteria on which they were to base their ratings. Which sketches each participant rated was also completely randomized, and the samples for both writers and raters in this study were rather diverse and not as narrow as in Study 1. These differences between Studies 1 and 2 means that we balanced the strengths and weaknesses of the two broad approaches (naturalistic versus laboratory), suggesting that the converging results we observed across the two studies are rather robust. Integrating our findings across Studies 1 and 2 may elucidate what criteria are being used to provide the generic ratings in Study 1. The textual features of stories that predicted these ratings were the same as the textual features that predicted how interesting or complex the characters were in Study 2 (i.e., negative valence and abstractness). This similarity in findings suggests that the website users used interest and/or complexity as their rating criteria. Given that generic ratings are often used to provide recommendations for books, films, and other media (e.g., on GoodReads or IMDB), it is useful to conduct research into what these ratings reflect and how rating criteria might vary depending on genre or other factors.

Lastly, in both studies, the combined textual features we measured accounted for 10-30%

of the variance in ratings (see Tables 1 and 2). However, it is important to note that in many cases the size of the individual effects we observed were small. This is perhaps not surprising, because stories and characters are complex constructs, and we are looking at very simple surface-level characteristics. Additionally, our regressions capture the effects of individual words, not sentences or entire stories or character sketches. When combined however, these textual features were most successful at predicting character complexity (accounting for 34% of the variance), compared to the other dimensions of appeal. This may be because the length of the character sketches was an important predictor of complexity (maximally affecting this rating by almost 30%). Developing a multifaceted character that is easy for a reader to imagine may benefit from including more detail, resulting in a lengthier character sketch. Including more detail may be less important for establishing a character's interest or likeability, which may be tied to other textual features that have less predictive power, such as emotional valence. Our textual features accounted for the least amount of variance in the ratings of entire stories in Study 1 (10%). Stories are more complex constructs than brief character sketches and explaining their appeal may require a larger set of textual features, as well as consideration of deeper levels below surface characteristics. There are likely other influential textual features that our analyses did not capture, and these remain to be investigated in future work.

Thus, our results offer a useful starting point for future corpus linguistic investigations into the appeal of fiction. Here we have examined short stories and character sketches. Future work should examine other elements of fiction, such as setting and plot. Another important direction for future work is examining longer texts, such as entire novels, as well as different narrative modalities, such as television shows and film. Many novels and film scripts are publicly available and are associated with ratings (on sites like GoodReads and IMDB), making

this research highly feasible. Applying linguistic approaches to longer form texts also allows for the ability to capture more complex features, such as changes in emotional valence or abstractness over the course of a narrative.

In closing, our findings point to some interesting nuances in how readers respond to short stories and fictional characters. It seems that negative content draws interest: a story without conflict, turmoil, or a problem is boring and unsuccessful (Fiedler, 1982). At the same time, positive or calming content may be important for developing fondness toward a character. Although different readers undoubtedly find various aspects of fiction appealing for many different reasons, our research demonstrate that some observable consistencies emerge across various types of stories, characters, and readers. In this way, our findings provide a valuable starting point for understanding which of these aspects are important and how they might come together to appeal to readers.

References

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., Ghahremani, D. G., Glover, G., ... & Sobel, N. (2003). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neuroscience, 6*(2), 196-202.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., & Singmann, H. (2015). lme4: Linear mixed-effects models using Eigen and S4, 2014. *R Package Version, 1*(4).

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323-370.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation: Defensive and appetitive reactions in picture processing. *Emotion, 1*(3), 276-298.

Burroway, J. (2007). *Imaginative Writing: The Elements of Craft. 2 Ed.* London: Pearson.

Breland, H. M. (1996). Word frequency and word difficulty: A comparison of counts in four corpora. *Psychological Science, 7*(2), 96-99.

Brewer, W. F., & Lichtenstein, E. H. (1982). Stories are to entertain: A structural-affect theory of stories. *Journal of Pragmatics, 6*(5-6), 473-486.

Bryant, J., & Zillmann, D. (1984). Using television to alleviate boredom and stress: Selective exposure as a function of induced excitational states. *Journal of Broadcasting & Electronic Media, 28*(1), 1-20.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977-990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*(3), 904-911.

De Biasi, P. M. (2002). Flaubert: The labor of writing. In A.-M. Christin (Ed.), *A history of writing* (pp. 340-341). Paris: Flammarion.

Dibble, J. L., & Rosaen, S. F. (2011). Parasocial interaction as more than friendship: Evidence for parasocial interactions with disliked media figures. *Journal of Media Psychology: Theories, Methods, and Applications, 23*(3), 122-132.

Djikic, M., Oatley, K., & Peterson, J. B. (2006). The bitter-sweet labor of emoting: The linguistic comparison of writers and physicists. *Creativity Research Journal, 18*(2), 191-197.

Dillman Carpentier, F. R., Brown, J. D., Bertocci, M., Silk, J. S., Forbes, E. E., & Dahl, R. E. (2008). Sad kids, sad media? Applying mood management theory to depressed adolescents' use of media. *Media Psychology, 11*(1), 143-166.

Eerolaa, T., Vuoskoskib, J. K., Peltolab, H-R., Putkinenb, V., Schäferb, K. (2018). An integrative review of the enjoyment of sadness associated with music. *Physics of Life Reviews, 25*, 100–121.

Fiedler, L.A. (1982) *Love and death in the American Novel.* New York: Stein & Daly.

Forgas, J. P. (2003). Affective influences on attitudes and judgments. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Series in affective science. Handbook of affective sciences* (pp. 596-618). New York: Oxford University Press.

Frijda, N. H. (1989). Aesthetic emotions and reality. *American Psychologist, 44*(12), 1546-1547.

Gardner, J. (1964). *The art of fiction*. New York: Knopf, reissued by Random House, 1991.

Gerrig, R. (1993). *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. New Haven: Yale University Press.

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.

Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory, 14*(4), 311-327.

Hassabis, D., Spreng, R. N., Rusu, A. A., Robbins, C. A., Mar, R. A., & Schachter, D. L. (2014). Imagine all the people: How the brain creates and uses personality models to predict behavior. *Cerebral Cortex, 24*, 1979-1987.

Hogan, P. C. 1997. Literary universals. *Poetics Today 18*(2): 223-249.

Hogan, P. C. 2003. *The mind and its stories: Narrative universals and human emotion*, United Kingdom: Cambridge University Press.

Holbrook, M. B., & Addis, M. (2007). Taste versus the market: An extension of research on the consumption of popular culture. *Journal of Consumer Research, 34*(3), 415-424.

Iguarta, J. J. (2010). Identification with characters and narrative persuasion through fictional feature films. *Communications, 35*, 347-373.

Jockers, M. L., & Mimno, D. (2013). Significant themes in 19th-century literature. *Poetics, 41*(6), 750-769.

Johnston, W. A., Hawley, K. J., Plewe, S. H., Elliott, J. M., & DeWitt, M. J. (1990). Attention capture by novel stimuli. *Journal of Experimental Psychology: General, 119*(4), 397-411.

Josephson, B. R., Singer, J. A., & Salovey, P. (1996). Mood regulation and memory: Repairing sad moods with happy memories. *Cognition & Emotion*, *10*(4), 437-444.

Kao, J., & Jurafsky, D. (2012, June). A Computational Analysis of Style, Affect, and Imagery in Contemporary Poetry. In the *Workshop on Computational Linguistics for Literature* (pp. 8-17).

Kintsch, W. (1980). Learning from text, levels of comprehension, or: Why anyone would read a story anyway. *Poetics*, *9*(1-3), 87-98.

Knobloch, S., & Zillmann, D. (2002). Mood management via the digital jukebox. *Journal of Communication*, *52*(2), 351-366.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261-273.

Mahlberg, M. (2007). Clusters, key clusters and local textual functions in Dickens. *Corpora*, *2*(1), 1-31.

Maslej, M. M., Rain, M., Fong, K., Oatley, K., & Mar, R. A. (2014). The hierarchical personality structure of aspiring creative writers. *Creativity Research Journal*, *26*(2), 192-202.

Maslej, M. M., Oatley, K., & Mar, R. A. (2017). Creating fictional characters: The role of experience, personality, and social processes. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(4), 487-499.

Mar, R. A., Oatley, K., Djikic, M., & Mullin, J. (2011). Emotion and narrative fiction: Interactive influences before, during, and after reading. *Cognition & Emotion*, *25*(5), 818-833.

Mares, M. L., & Cantor, J. (1992). Elderly viewers' responses to televised portrayals of old age: Empathy and mood management versus social comparison. *Communication Research, 19*(4), 459-478.

Marjanovic, Z., Struthers, C. W., Cribbie, R. A., & Greenglass, E. R. (2014). The Conscientious Responders Scale: A new tool for discriminating between conscientious and random responders. *SAGE Open, 4*, 1–10.

McIntyre, D. (2012). Prototypical characteristics of blockbuster movie dialogue: A corpus stylistic analysis. *Texas Studies in Literature and Language, 54*(3), 402-425.

Menninghaus, W., Wagner, V., Hanich, J., Wassiliwizky E., Jacobsen, T., & Koelsch, S. (2017). The distancing–embracing model of the enjoyment of negative emotions in art reception. *Behavioral and Brain Sciences, 40,* 1–58.

Miall, D. S. (1988). Affect and narrative: A model of response to stories. *Poetics 17*(3), 259-272.

Miall. D. S. & Kuiken, D. (1994). Foregrounding, defamiliarization, and affect. Response to literary stories. *Poetics, 22,* 389-407.

Mohr, J. W., & Bogdanov, P. (2013). Introduction-Topic models: What they are and why they matter. *Poetics, 41*(6), 545-569.

Morini, M. (2016). *Jane Austen's narrative techniques: A stylistic and pragmatic analysis*. Abingdon: Routledge.

Mukařovský, J. (1976). *On Poetic Language*. Lisse: De Ridder Press.

Murphy, S. (2015). I will proclaim myself what I am: Corpus stylistics and the language of Shakespeare's soliloquies. *Language and Literature, 24*(4), 338-354.

Nabi, R. L., Finnerty, K., Domschke, T., & Hull, S. (2006). Does misery love company? Exploring the therapeutic effects of TV viewing on regretted experiences. *Journal of Communication 56*(4), 689-706.

Nell, V. (1988). The psychology of reading for pleasure: Needs and gratifications. *Reading Research Quarterly* 23(1), 6-50.

Oatley, K., & Djikic, M. (2008). Writing as thinking. *Review of General Psychology, 12*(1), 9-27.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: a threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80*(3), 381-396.

Oliver, M. B. (1993). Exploring the paradox of the enjoyment of sad films. *Human Communication Research 19*(3), 315-342.

Oliver, M. B., & Raney, A. A. (2011). Entertainment as pleasurable and meaningful: Identifying hedonic and eudaimonic motivations for entertainment consumption. *Journal of Communication 61*(5), 984-1004.

Paivio, A. (1966). Latency of verbal associations and imagery to noun stimuli as a function of abstractness and generality. *Canadian Journal of Psychology, 20*(4), 378-387.

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology, 1*(1), 33-60.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Mahway: Lawrence Erlbaum Associates.

Perkins, D. (1976). The Imagist Movement. In *A history of modern poetry: from the 1890s to the high modernist mode* (pp. 330-332). Cambridge, MA: Harvard University Press.

Peskin, J. (1998). Constructing meaning when reading poetry: An expert-novice study. *Cognition and Instruction, 16*(3), 235-263.

Pound, E. (1913). A Few Don'ts by an Imagiste. *Poetry, 1*(6), 200-206.

Ruiz, J. (2015). Dark Matters: Vince Gilligan's Breaking Bad, Suburban Crime Dramas, and Latinidad in the Golden Age of Cable Television. *Aztlan: A Jounal of Chicano Studies, 40*(1), 37-62.

Schank, R. C. (1978). *Interestingness: Controlling inferences*. Yale University: Department of Computer Science.

Semino, E., & Short, M. (2004). *Corpus stylistics: Speech, writing and thought presentation in a corpus of English writing*. Abingdon: Routledge.

Sokolov, E. N. (1963). Higher nervous functions: The orienting reflex. *Annual Review of Physiology, 25*(1), 545-580.

Tan, E. S. H. (1995). Film-induced affect as a witness emotion. *Poetics 23*(1-2), 7-32.

Tangherlini, T. R., & Leonard, P. (2013). Trawling in the Sea of the Great Unread: Sub-corpus topic modeling and Humanities research. *Poetics, 41*(6), 725-749.

Tian, Q., & Hoffner, C. A. (2010). Parasocial interaction with liked, neutral, and disliked characters on a popular TV series. *Mass Communication and Society*, *13*(3), 250-269.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1256-1271.

Wallentin, M., Nielsen, A. H., Vuust, P., Dohn, A., Roepstorff, A., & Lund, T. E. (2011). Amygdala and heart rate variability responses from listening to emotionally intense parts of a story. *Neuroimage 58*(3), 963-973.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191-1207.

Zillmann, D. (2000). Mood management in the context of selective exposure theory. *Annals of the International Communication Association*, *23*(1), 103-123.

Zillman, D., & Cantor, J. R. (1977). Affective responses to the emotions of a protagonist. *Journal of Experimental Social Psychology*, *13*(2), 155-165.

Table 1

Correlations between textual features and story ratings in Study 1

		Words	Readers	Frequency	Valence	Arousal	Concrete	Sexual	Religion	Death
Rating	r_s	.13	-.01	.07	-.14	-.01	**-.15**	**-.14**	.02	.09
	(95% CI)	(-.02, .28)	(-.15, .13)	(-.09, .23)	(-.28, 0)	(-.14, .11)	(-.29, -.02)	(-.27, -.01)	(-.12, .18)	(-.05, .24)
	p	.083	.930	.349	.062	.909	.045	.058	.752	.250
Words	*rs*	-	.04	.08	-.01	-.04	-.07	-.02	.11	-.03
	(95% CI)		(-.10, .18)	(-.08, .24)	(-.18, .17)	(-.17, .1)	(-.21, .09)	(-.17, .13)	(-.06, .27)	(-.17, .12)
	p		.565	.278	.894	.618	.397	.785	.169	.699
Readers	r_s		-	-.13	.04	**.26**	**.20**	**.43**	-.09	**-.18**
	(95% CI)			(-.27, .01)	(-.11, .18)	(.12, .40)	(.04, .37)	(.29, .56)	(-.23, .06)	(-.32, -.05)
	p			.091	.625	<.001	.009	<.001	.230	.017
Frequency	r_s			-	**.39**	**-.28**	**-.63**	-.11	-.07	-.16
	(95% CI)				(.24, .52)	(-.42, -.14)	(-.72, -.51)	(-.25, .03)	(-.22, .1)	(-.31, 0)
	p				<.001	<.001	<.001	.146	.394	.034
Valence	r_s				-	**-.35**	**-.25**	.12	-.14	**-.45**
	(95% CI)					(-.50, -.19)	(-.39, -.11)	(-.02, .25)	(-.28, .02)	(-.55, -.32)
	p					<.001	.001	.118	.072	<.001
Arousal	r_s					-	**.26**	**.45**	**.16**	**.26**
	(95% CI)						(.11, .42)	(.32, .58)	(.02, .3)	(.12, .39)
	p						<.001	.004	.032	.001
Concrete	r_s						-	**.22**	**-.16**	0
	(95% CI)							(.07, .37)	(-.3, -.02)	(-.17, .16)
	p							.004	.032	.964
Sexual	r_s							-	.02	**-.18**
	(95% CI)								(-.16, .19)	(-.31, -.04)
	p								.821	.020
Religion	r_s								-	**.20**
	(95% CI)									(.04, .35)
	p									.009

Note. Words = Number of words in the story. All confidence intervals are bootstrapped based on 1000 resamples, bias-corrected and accelerated. Correlations in which confidence intervals do not include zero are in bold.

Table 2

Unique effects of textual features on story ratings in Study 1

	Story ratings				
Feature	β	*SE*	95% CI	*p*	ES
Words	.13	.00	.13, .14	<.001	0.51
Readers	-.02	.00	-.03, -.02	<.001	0.14
Frequency	.03	.01	.01, .04	<.001	0.14
Valence	-.02	.00	-.03, -.02	<.001	0.15
Arousal	.00	.01	-.01, .01	.998	0.00
Concrete	-.04	.01	-.05, -.03	<.001	0.12
Sex	-.01	.01	-.02, -.00	.023	0.10
Religion	-.00	.00	-.01, .00	.362	0.05
Death	.00	.00	-.01, .01	.399	0.05

Note. Words = Number of words in the text; ES = Effect size indicating the difference in the value of a rating when there is a maximal change in a given predictor. *SD* (random effect) = 0.21; *SD* (residual) = 0.74; R^2 = 0.10.

Table 3

Correlations between textual features and character ratings in Study 2

		Likeable	Complex	Words	Frequency	Valence	Arousal	Concrete
Interesting	r_s	**.33**	**.78**	**.43**	-.04	-.14	**.23**	**-.19**
	(95% CI)	(.19, .46)	(.31, .53)	(.31, .53)	(-.18, .11)	(0, -.28)	(.09, .37)	(-.04, 32)
	p	<.001	<.001	<.001	.547	.053	.001	.009
Likeable	*rs*	-	**.30**	**.28**	-.11	**.46**	-.06	.02
	(95% CI)		(.17, .43)	(.15, .40)	(-.25, .05)	(.33, .57)	(-.19, .08)	(-.11, .16)
	p		<.001	<.001	.122	<.001	.442	.735
Complex	r_s		-	**.68**	-.01	-.08	**.18**	**-.19**
	(95% CI)			(.59, .74)	(-.15, .14)	(-.22, .05)	(.04, .31)	(-.32, -.06)
	p			<.001	.933	.277	.013	.007
Words	r_s			-	.06	.09	.06	**-.22**
	(95% CI)				(-.10, .21)	(-.05, .22)	(-.10, .19)	(-.35, -.07)
	p				.420	.203	.441	.002
Frequency	r_s				-	.07	**-.24**	**-.26**
	(95% CI)					(-.07, .20)	(-.38, -.10)	(-.39, -.13)
	p					.351	<.001	<.001
Valence	r_s					-	**-.15**	.04
	(95% CI)						(-.28, -.01)	(-.11, .17)
	p						.035	.605
Arousal	r_s						-	.02
	(95% CI)							(-.12, .17)
	p							.733

Note. Words = Number of words in the story. All confidence intervals are bootstrapped based on 1000 resamples, bias-corrected and accelerated. Correlations in which confidence intervals do not include zero are in bold.

Table 4

Unique effects of linguistic features on character ratings in Study 2

	Interest					Likeability					Complexity				
Feature	β	*SE*	95% CI	*p*	ES	β	*SE*	95% CI	*p*	ES	β	*SE*	95% CI	*p*	ES
Frequency	.02	.02	-.02, .05	.293	0.08	.00	.02	-.03, .04	.966	0.00	-.00	.01	-.02, .01	.751	0.01
Words	.32	.01	.30, .34	<.001	1.64	.21	.01	.18, .23	<.001	1.04	.41	.01	.40, .42	<.001	2.09
Valence	-.06	.01	-.08, -.03	<.001	0.36	.11	.01	.08, .14	<.001	0.69	-.03	.01	-.04, -.02	<.001	0.19
Arousal	.03	.01	.00, .06	.015	0.22	-.04	.02	-.07, -.01	.019	0.24	.01	.01	-.00, .03	.137	0.08
Concrete	-.04	.02	-.07, -.01	.011	0.14	-.02	.02	-.05, .02	.361	0.05	-.02	.01	-.04, -.01	.009	0.08

Note. Words = Number of words in the text. ES = Effect size indicating the difference in the value of a rating when there is a maximal change in a given predictor.
Interest: *SD* (random effect) = 0.24; *SD* (residual) = 0.85; R^2 = 0.21
Likeability: *SD* (random effect) = 0.25; *SD* (residual) = 0.10; R^2 = 0.14
Complexity: *SD* (random effect) = 0.09; *SD* (residual) = 0.53; R^2 = 0.34