Decomposing the Fundamentals of Creepy Stories

Sakshi Goel, Haripriya Dharmala, and Yuchen Zhang
University of Southern California
{sakshigo, dharmala, yzhang71}@usc.edu

Keith Burghardt
USC Information Sciences Institute
keithab@isi.edu

Abstract
Fear is a universal concept; people crave it in urban legends, scary movies, and modern stories. Open questions remain, however, about why these stories are scary and more generally what scares people. In this study, we explore these questions by analyzing tens of thousands of scary stories on forums (known as subreddits) in a social media website, Reddit. We first explore how writing styles have evolved to keep these stories fresh before we analyze the stable core techniques writers use to make stories scary. We find that writers have changed the themes of their stories over years from haunted houses to school-related themes, body horror, and diseases. Yet some features remain stable; words associated with pseudo-human nouns, such as clown or devil are more common in scary stories than baselines. In addition, we collect a range of datasets that annotate sentences containing fear. We use these data to develop a high-accuracy fear detection neural network model, which is used to quantify where people express fear in scary stories. We find that sentences describing fear, and words most often seen in scary stories, spike at particular points in a story, possibly as a way to keep the readers on the edge of their seats until the story's conclusion. These results provide a new understanding of how authors cater to their readers, and how fear may manifest in stories.

1 Introduction
“The story had held us, round the fire, sufficiently breathless, but except the obvious remark that it was gruesome...I remember no comment uttered till somebody happened to say that it was the only case he had met in which such a visitation had fallen on a child.” This is the opening line to Henry James’s “The Turning Of The Screw” (James, 2008) and summarizes a key and counter-intuitive aspect of scary stories: they can be both revolting and captivating; they capture a fundamental emotion, fear (Adolphs, 2013), but in doing so, provide entertainment. The closest analogy might be how spicy foods cause pain yet are tasty. We understand what makes food spicy, and can tailor a meal’s spice to fit one’s palate, yet it is less clear how to generate fear in a story because we do not know the ingredients that make a story scary. This is a special case of a more general issue, in which we do not yet fully understand how stories elicit particular emotions, i.e., what emotion the reader should feel, in contrast to detecting the emotion a writer expresses (Alhuzali and Ananiadou, 2021; Strapparava and Mihalcea, 2008).

In this paper, begin to understand how stories make readers scared by quantifying patterns in scary stories. First, we analyze the topics associated with scary stories, and find writers are making fewer stories about houses and more stories about human bodies, school, and diseases, potentially to stay topical. We then develop a method to quantify more universal patterns in scary stories by measuring the percentage of times a word appears in scary and non-scary stories, a metric which we call Scary Story Token Prevalence (SSToP). Words with high SSToPs, such as creature or silhouette, are more likely to be human-like words, suggesting scary story writers intentionally use words that are semi-human. Finally, we analyze where writers use high SSToP words or words expressing fear within a story. We find that stories can be decomposed into modes where the high SSToP or fearful words appear in regular intervals, and combinations of these modes are used to develop scary story patterns. Overall, these results lend insight into how stories express fear, and patterns writers exploit to make scary stories.

Our contributions can be summarized as follows:

- we analyze dynamic patterns of scary stories to understand what themes have become popular or fell out of favor.
• we analyze static patterns, such as the relationship between scary story nouns like “human” to capture what words writers use and why they might use those words

• we capture how scary stories are written by quantifying the common scary story modes. We find some modes are particularly common, which offers clues about how writers make a scary story.

2 Related Work

2.1 Uncanny Valley

What makes something scary is most similar to previous work on the uncanny valley effect. Research on the uncanny valley began with Jentsch and Freud who coined the term *unheimlich* (literally, “unhomely”) when referencing human uncanniness and eeriness (Freud, 1919; Jentsch, 1906). The uncanny valley also came into prominence among human-computer interaction (HCI) and human-robot interaction (HRI) researchers with a paper by Mori (1970). Researchers were inspired by Mori to learn what can make a robot unsettling (Broadbent, 2017) because willingness to work with robots can depend on their likability (Destephe et al., 2015). An objective definition of uncanny (what makes something human-like) or valley (what is and is not unsettling) is still, however, a fiercely debated topic (Mathur and Reichling, 2016; Langer and König, 2018; Mathur et al., 2020; Palomaki et al., 2018; MacDorman and Ishiguro, 2006; Ho and MacDorman, 2017). Moreover, the qualitative shape of the uncanny valley varies significantly between studies (Mathur and Reichling, 2016; Langer and König, 2018; Mathur et al., 2020; Palomaki et al., 2018; MacDorman and Ishiguro, 2006; Ho and MacDorman, 2017). Finally, these works (with rare exception (McAndrew and Koehnke, 2016)) do not explore non-image media, such as text. Open questions, such as how to make less-scary text, or how make scarier ghost stories, have also been under-explored.

2.2 Emotion Detection

The fear detection model we create is related to previous work on emotion detection. Emotions are fundamental to humans across cultures (Plutchik, 1982), and are typically split into eight components: terror (fear), admiration, ecstacy, vigilance, rage, loathing, grief, and amazement. Research on detecting these emotions in text has a long history (Acheampong et al., 2020; Chen and He, 2022), but this research recently came to prominence with text embedding-based methods that can accurately detect emotion in sentences. Detection methods have recently exploited advances in Bi-LSTMs and attention for higher accuracy (c.f., (Cai and Hao, 2018; Alhuzali and Ananiadou, 2021)) and have proven useful for chatbots, for example, which are used in a range of applications (Chen and He, 2022). Emotion detection is closely related to sentiment analysis, such as VADER (Hutto and Gilbert, 2014), or the hedonometer, studied by Dodds et al., (Dodds et al., 2011; Reagan et al., 2016). Both methods aim to detect the positivity or negativity of words or sentences, and has been used to understand story arcs (Reagan et al., 2016) or tweet sentiment (Dodds et al., 2011; Hutto and Gilbert, 2014).

3 Data

3.1 Fear Training Data

We use multiple datasets to train a fear-detection model, where each dataset has sentences labeled as expressing fear or not. The datasets were gathered and converted into a binary classification problem with a label 1 or 0 representing if the sentence does or does not express fear, respectively. The datasets used for modeling are described in Table 1. This is one of the most complete collections of data to analyze fear in text. Alike to previous research (Bostan and Klinger, 2018), we collect a range of emotion data, but in contrast to that work, some of our data are unique and only available upon request from the authors (Aman and Szpakowicz, 2007). This diverse dataset is trained with a model we describe in the Metrics section. In total, there are 70K labeled sentences, with roughly 7K of these sentences labeled “fear.”

3.2 Reddit Data

We extract stories from subreddits using the Python wrapper for the Pushshift API (https://psaw.readthedocs.io). Stories are collected within two broad genres: scary stories and non-scary stories. For the scary stories, our data comes from the subreddits r/CreepyPasta, r/ShortScaryStories, and r/NoSleep, and for non-scary stories we selected the subreddits r/TIFU, r/Self, r/StoriesAboutKevin, r/Confession, and r/Confessions, which are common story-writing subreddits. All stories are collected from the start of the subreddit until June,
## Data Cleaning

To clean data for fear detection, we remove all characters except English alphabets and spaces. This ensures that the fear model’s vocabulary is only English words, and any numbers/HTML markup is removed. Next, we split phrases written in the camel case (a frequent occurrence in the Reddit and Tweet datasets due to hashtags) into its constituent words. (e.g. #EndThisTyranny split to ‘end this tyranny’). Finally, we convert all strings to lowercase and remove all English stopwords (collected from NLTK corpus), to ensure that the model doesn’t falsely learn fear from universally occurring stopwords.

## Metrics

### 3.4.1 Fear Confidence

To determine the confidence that a word or sentence expresses fear we train a model whose architecture

| Dataset | Description |
|---------|-------------|
| ISEARs (Scherer and Wallbott, 1994) | Obtained from cross-cultural studies in 37 countries and contains sentences annotated for joy, sadness, fear, anger, guilt, disgust and shame emotions |
| WASSA 2017 (Mohammad and Bravo-Marquez, 2017) | Constructed from tweets and annotated for joy, sadness, fear, and anger emotions |
| Affect Data (Alm, 2008) | Constructed from Tales and classified into angry, fearful, happy, sad, disgusted and surprised emotions |
| Emotion Stimulus No Cause Dataset (Ghazi et al., 2015) | Data developed from FrameNets’ annotated data for emotion lexical unit, each sentence is marked with an emotion tag. |
| CrowdFlower Twitter Sentiment Analysis Emotion Data (CrowdFlower, 2021) | Constructed from tweets and classified into 13 emotions (sadness, relief, anger, worry, enthusiasm etc.) |
| Emotion-Annotated Dataset (Aman and Szpakowicz, 2007) | Constructed from blog data sentences classified into happiness, sadness, fear, anger, surprise and disgust. |
| DailyDialog Dataset (Li et al., 2017) | Human-written text sentences tagged with emotions and communication intention |
| Emotion Hashtag Corpus (Mohammad and Kiritchenko, 2015) | Tweets annotated with emotions based on the emotion-word hashtag in the tweet text |

Table 1: Datasets used to construct fear model.

Figure 1: The fear model framework. (a) The neural network used to predict fear in sentences (5,840,850 parameters total). (b) The receiver operative characteristic for the neural network model.
is described in Fig. 1a. A Bidirectional GRU (Bi-GRU) is used because it is capable of incorporating contextual information in the forward and reverse direction in contrast to a standard GRU or LSTM. A Global Average Pooling layer was used after the Bi-GRU layer and a dropout layer of 50% was used to regularize the dense layer just before the two-node prediction layer. A batch size of 64 was used with categorical cross-entropy loss and an Adam optimizer.

To train this model, we split the meta-dataset into 3 parts (where each datapoint is assigned at random): 75% for training, 10% for validation and 10% for testing. The original training dataset contained 55.8K sentences, of which 5.5K were sentences expressing fear. To reduce imbalance, we up-sample the fear sentences to create 100.5K sentences for training. Validation and test sets are separate sets of 7.0K sentences without up-sampling or down-sampling (roughly 700 sentences in each set were labeled “fear”). Tensorflow Keras is used for the tasks of tokenization, sequencing and padding of the three datasets. The tokenizer is only fit on the training data. The vocabulary size was 56.5K. Each sentence is then converted into a sequence and zero-padded to a length of 100. The model ROC curve is shown in Fig. 1b, which shows high micro- and macro-averaged AUC (0.96 and 0.85, respectively).

We compare this model to logistic regression models with L1 or L2 regularization. Text is tokenized and fed through FastText (Joulin et al., 2016) embeddings trained on the scary and non-scary story corpora. All embedded words in a story are then averaged to create a document embedding. When training the logistic regression models, the training data is down-sampled so that each class is of equal size. The comparisons against the baselines are shown in Table 2. We see significant improvements over the baseline models.

### 3.4.2 Scary Story Token Prevalence

SSToP is defined as the odds ratio between a word’s frequency in scary and non-scary stories:

$$c(word) = \frac{f_{\text{scary}}(word)}{f_{\text{normal}}(word)}$$  \hspace{1cm} (1)

Where $f_X(word)$ is the word frequency across all posts and subreddits within the genre $X$. To check the robustness of this method, we re-calculate $c(word)$ such that non-scary word frequencies were calculated from each of the non-scary subreddits (creating five separate metrics) and results are qualitatively the same. To calculate this metric for each word, we first select all stories which contain at least 500 words. This helps ensure that the lengths of the stories are approximately uniform and allows us to calculate story modes described later. The stories are preprocessed by tokenizing and removing punctuation. We utilize a dictionary to keep track of the number of times each word occurs. This is calculated for both scary and non-scary stories and Eq. 1 is applied to determine a SSToP value for each word. Since Reddit stories are not checked for correct grammar or spelling, there are several unique names or words which may be misspelt. Because the occurrence of each of these words does not span broadly throughout the entire subreddit, SSToP cannot accurately assign a value to these unique terms. Therefore, we only consider words which have occurred at least 100 times in the entire corpora of scary and non-scary stories.

### 3.5 Generating Story Modes

The main motivation behind creating story modes is to identify patterns writers exploit to make scarier stories. Some stories may, for example, get scary towards the end while other stories may start out scary and end scary, but have a less-scary segment in the middle. Therefore, while we generate a model to detect whether fear is expressed in sentences, it is the pattern of fear, the modes, that we study to better understand how readers might be provoked to experience fear.

To construct these story modes, we first segment each story into ten points based on the percentage of story covered (0%, 10%,...,90%). We then calculate the fear confidence of a sentence located at that point in the story or average the SSToP score for 50 words after that point in the story. After the metrics are sampled, we construct a matrix, $M$, such that columns represent the 10 sampled points, and rows represent each story. We then apply singular value decomposition to calculate the story modes (Reagan et al., 2016), such that

$$M = V \Sigma U = WU$$

The matrix $U$ represents the modes while $W$ represents the coefficients of each mode used to reconstruct each story, alike to the coefficients of a Fourier transform. The largest element in each row of $W$ represents the mode that best aligns with each story, but in general several
| Model          | Accuracy | F1 (Macro Avg) | ROC-AUC |
|---------------|----------|----------------|---------|
| Neural Net    | 0.92     | 0.76           | 0.85    |
| Logistic (L1) | 0.71     | 0.60           | 0.84    |
| Logistic (L2) | 0.74     | 0.61           | 0.84    |
| Majority      | 0.90     | 0.57           | 0.50    |

Table 2: Model performance for neural network-based fear model and logistic regression baselines. Bottom row are metrics for majority class null model, in which we guess all test data is not scary (the majority class).

| High SSToP Words (Common in Scary Stories) | SSToP | Low SSToP Words (Uncommon in Scary Stories) | SSToP |
|--------------------------------------------|-------|---------------------------------------------|-------|
| eyeless                                    | 100   | rslef                                       | 0.0018|
| scarecrows                                 | 108   | fwb                                        | 0.0021|
| cryptids                                   | 110   | superjail                                   | 0.0022|
| skinwalkers                                | 130   | 19f                                         | 0.0023|
| headstones                                 | 137   | tldr                                        | 0.0025|

Table 3: Examples of words with high and low SSToP scores. Skin-walkers, which have a high SSToP score, are a harmful witch in Navajo culture.

modes are needed to reconstruct a story.

3.6 Word similarity to human

To determine how similar a noun is to human, we first embed nouns using FastText. Human synonyms, namely human, mortal, person, soul, Homo Sapien, earthing, higher animal, and living person (common synonyms of human that are gender neutral) are averaged together into one human vector to avoid noise associated with the embedding of a single word. This follows similar work on gender bias in word embeddings, where the first principle component of a collection of words are used to construct a he-she vector, in order to reduce noise associated with each individual pair of words (Bolukbasi et al., 2016). Next we determine what words in the word embeddings are nouns using Python’s TextBlob library (https://textblob.readthedocs.io). If NN, PN, or NP are in the TextBlob tags for a word, the word is categorized as a noun. This method helps us avoid finding the unrealistic similarity between, e.g., a verb and a noun. Finally, we determine distance between words using Euclidean distance. Other distance metrics, such as Manhattan distance (Aggarwal et al., 2001), show similar results.

4 Results

We show data statistics in Fig. 2. This figure demonstrates when writers post their stories and what they discuss. First, Fig. 2a shows the frequency of stories over time, which rapidly increases to over 20,000 stories a year before dropping slightly. In the inset, we see that users are least likely to post their stories at 12 p.m. GMT, or around 5 a.m. EST, perhaps because most writers might be from America and do not post when they just wake up. Next, we use Latent Derichlet Allocation (with lemmatized words and all stop words removed) to find ten story topics, the largest of which are shown in Fig. 2b. We notice that the house topic (with words like “door” or “window”) drops off in frequency while topics related to bodies (with words like “bodies” or “blood”) and school (with words like “phone” or “school”) become increasingly popular. These results suggest long-term trends in what writers consider topical for their stories, possibly because writers focus on different topics as they age or they believe readers find school or body horror more interesting. Finally, we show the frequency of disease-like stories, in which “lockdown,” “infect,” “virus,” or “disease” lemmas are within the story, increase dramatically in 2020 (likely due to COVID-19 being in the news), but this follows a long-term trend, in which disease-like stories have become increasingly popular. We also notice a small peak in their prevalence in mid-2014, when the disease Ebola became topical.

Next, we show in Fig. 3 that a word’s SSToP score varies with its similarity to human. Words seen in scary stories tend to be more similar to hu-
Figure 2: Scary story statistics. (a) The number of stories each year (inset: the number of stories written each hour in GMT time). (b) LDA topics over time, which shows a significant drop in the house-like topic but an increase in body-related and school-related topics in recent years. (c) The proportion of disease-related stories over time.

Figure 3: Scary story writers use more human-like words. Mean SSToP score (fraction of words in scary stories) in red versus noun similarity to “human.” Error bars are standard errors.

man, which is reminiscent of the uncanny valley, in which people find semi-human objects more unsettling. A number of plausible reasons for these patterns exist. For example, scary story authors use semi-human words (such as “devil” or “clown”) in an attempt to unsettle readers, or they are discussing human-related topics, or they avoid words that are static and boring, such as cars, which are far from “human” in latent embedding space. Whatever the reason, this pattern gives new insights into how writers make scary stories.

Finally, we analyze how scary story authors develop story modes. In Fig. 4, we show story modes of fear confidence and SSToP score in the r/NoSleep subreddit (14K stories in total). Blue lines in Fig. 4a correspond to the three lowest modes for fear confidences, red lines in Fig. 4b are the three lowest modes for SSToP scores. Gray lines correspond to 20 stories which have the highest coefficients in the corresponding W matrix for a given mode. We normalize the story modes by dividing by the largest W coefficient in each row. These figures show that stories use high SSToP words, such as “shadow” or “creature” at very particular points in the story, presumably at story climaxes or key scenes. Stories also express fear at particular points in the story as well, presumably to keep the audience thrilled or afraid until the end of the story.

In Fig. 4c, we show the frequency of the largest mode associated with each story. We see that many stories are in the lowest SSToP mode. Scary story authors therefore tend to use scary words in the center of the story, perhaps to set up scenes. Surprisingly, however, there is a variety of modes used for fear, and a combination of these modes often appear in stories. Furthermore, the fear and SSToP modes rarely align. Namely, the largest SSToP modes for a given story have a low correlation with the top fear modes (Spearman correlation = 0.06, p-value < 10\(^{-10}\)). Finally mode 9 is rarely the largest mode because it corresponds to a story without any change (the mode is a flat line).

5 Conclusions

In conclusion, we analyze text patterns to better understand how writers make scary stories. The results suggest that writers tend to be topical or change the story topic to fit the mood of readers. Moreover, scary story authors tend to use human-like words, which is reminiscent of the uncanny valley in which people find human-like images more unsettling. In addition, authors create stories whose
Figure 4: Story modes in r/NoSleep. (a) 3 lowest story modes for fear (modes are blue lines), (b) 3 lowest story modes for SSToP scores (modes are red lines), and (c) histograms of each story mode for the respective metrics. Gray lines in (a-b) are example stories that best match those respective modes (they have the largest W coefficients for those modes).

modes may help scare readers. Overall, our work uncovers patterns associated with text that elicits fear, but we believe more work will be needed to understand how these patterns may cause fear to be elicited.

6 Limitations

There are a number of limitations in our work. While we uncover patterns associated with scary stories, more work is needed to understand whether or how these patterns create fear in readers. This task will require analyzing reader reactions in a controlled study. In addition, the relationship between nouns and words like “human” have not been tested on alternative text embedding models, such as GLoVe (Pennington et al., 2014), or attention-based embeddings, such as BERT (Devlin et al., 2019). Finally story motifs we uncover should be compared against alternative methods (Reagan et al., 2016).

7 Ethical Considerations

All results come from publicly available and anonymized data, therefore the study poses minimal ethical risk.

8 Acknowledgements

This project was developed at the USC Center for Knowledge-Powered Interdisciplinary Data Science (CKIDS) DataFest 2020. We also thank the USC Graduates Rising in Information and Data Science (GRIDS) for their help. Code to reproduce some of our major results are shown here: https://github.com/KeithBurghardt/Creepypasta.

References

Francisca Adoma Acheampong, Chen Wenyu, and Henry Nunoo-Mensah. 2020. Text-based emotion detection: Advances, challenges, and opportunities. Engineering Reports, 2(7):e12189.

Ralph Adolphs. 2013. The biology of fear. Current Biology, 23(2):R79–R93.

C.C. Aggarwal, A. Hinneburg, and D.A. Keim. 2001. On the surprising behavior of distance metrics in high dimensional space. In Database Theory – ICDT 2001. Lecture Notes in Computer Science, volume 1973, Berlin, Heidelberg. Springer.

Hassan Alhuzali and Sophia Ananiadou. 2021. Spanemo: Casting multi-label emotion classification as span-prediction. arXiv preprint arXiv:2101.10038.

Ebba Cecilia Ovesdotter Alm. 2008. Affect in text and speech.

Saima Aman and Stan Szpakowicz. 2007. Identifying expressions of emotion in text. In Text, Speech and Dialogue, pages 196–205, Berlin, Heidelberg. Springer Berlin Heidelberg.

Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS’16, page 4356–4364, Red Hook, NY, USA. Curran Associates Inc.

Laura-Ana-Maria Bostan and Roman Klinger. 2018. An analysis of annotated corpora for emotion classification in text. In Proceedings of the 27th International Conference on Computational Linguistics, pages 2104–2119, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Elizabeth Broadbent. 2017. Interactions with robots: The truths we reveal about ourselves. Annual
Xiaofeng Cai and Zhifeng Hao. 2018. Multi-view and attention-based bi-lstm for weibo emotion recognition. In Proceedings of the 2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018), pages 772–779. Atlantis Press.

Y. Chen and J. He. 2022. Deep learning-based emotion detection. Journal of Computer and Communications, 10:57–71.

CrowdFlower. 2021. The emotion in text data set.

Matthieu Destephe, Martim Brandao, Tatsuhiro Kishi, Massimiliano Zecca, Kenji Hashimoto, and Atsuo Takanishi. 2015. Walking in the uncanny valley: importance of the attractiveness on the acceptance of a robot as a working partner. Frontiers in psychology, 6:204–204.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186. Minneapolis, Minnesota. Association for Computational Linguistics.

Peter Sheridan Dodds, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth. 2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. PLOS ONE, 6(12):1–1.

S. Freud. 1919. Unheimlich. Imago.

Diman Ghazi, Diana Inkpen, and Stan Szpakowicz. 2015. Detecting emotion stimuli in emotion-bearing sentences. In Computational Linguistics and Intelligent Text Processing, pages 152–165. Cham. Springer International Publishing.

Chin-Chang Ho and Karl F. MacDorman. 2017. Measuring the uncanny valley effect. International Journal of Social Robotics, 9(1):129–139.

C. Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1).

Henry James. 2008. The Turn of the Screw. Project Gutenberg.

Ernst Jentsch. 1906. Zur psychologie des unheimlichen. Psychiatrisch-Neurologische Wochenschrift, 8(22–23):195–198, 203–205.

Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hervé Jégou, and Tomas Mikolov. 2016. Fasttext.zip: Compressing text classification models. arXiv preprint arXiv:1612.03651.

Markus Langer and Cornelius J. König. 2018. Introducing and testing the creepiness of situation scale (cross). Frontiers in Psychology, 9:2220.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. DailyDialog: A manually labelled multi-turn dialogue dataset. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 986–995. Taipei, Taiwan. Asian Federation of Natural Language Processing.

Karl F. MacDorman and Hiroshi Ishiguro. 2006. The uncanny advantage of using androids in cognitive and social science research. Interaction Studies, 7(3):297–337.

Maya B. Mathur and David B. Reichling. 2016. Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. Cognition, 146:22–32.

Maya B. Mathur, David B. Reichling, Francesca Lunardini, Alice Geminiani, Alberto Antonietti, Peter A.M. Ruijtjen, Carmel A. Levitan, Gideon Nave, Dylan Manfredi, Brandy Bessette-Symons, Attília Szuts, and Balazs Aczel. 2020. Uncanny but not confusing: Multisite study of perceptual category confusion in the uncanny valley. Computers in Human Behavior, 103:21–30.

Francis T. McAndrew and Sara S. Koehneke. 2016. On the nature of creepiness. New Ideas in Psychology, 43:10–15.

Saif Mohammad and Felipe Bravo-Marquez. 2017. WASSA-2017 shared task on emotion intensity. In Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 34–49. Copenhagen, Denmark. Association for Computational Linguistics.

Saif M. Mohammad and Svetlana Kiritchenko. 2015. Using hashtags to capture fine emotion categories from tweets. Computational Intelligence, 31(2):301–326.

M. Mori. 1970. The uncanny valley. Energy, 7(4):33–35.

Jussi Palomaki, Anton Kunnari, Marianna Drosinou, Mika Koverola, Noora Lehtonen, Juho Halonen, Marko Repo, and Michael Laakasuo. 2018. Evaluating the replicability of the uncanny valley effect. Heliyon, 4(11):e00939.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543.

Robert Plutchik. 1982. A psychoevolutionary theory of emotions. Social Science Information, 21(4-5):529–553.
Andrew J. Reagan, Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, and Peter Sheridan Dodds. 2016. The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Science*, 5(1):31.

K. R. Scherer and H. G. Wallbott. 1994. Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 67(1):55.

Carlo Strapparava and Rada Mihalcea. 2008. Learning to identify emotions in text. In *Proceedings of the 2008 ACM Symposium on Applied Computing*, SAC ’08, page 1556–1560, New York, NY, USA. Association for Computing Machinery.