The More Who Die, the Less We Care:

Evidence from Natural Language Analysis of Online News Articles and Social Media Posts

Sudeep Bhatia[a], Lukasz Walasek[b], Paul Slovic[c,d], Howard Kunreuther[a]

July 13, 2020

[a] University of Pennsylvania, PA, USA.

[b] University of Warwick, United Kingdom.

[c] Decision Research, OR, USA.

[d] University of Oregon, OR, USA.

Address correspondence to Sudeep Bhatia, Department of Psychology, Solomon Labs, 3720 Walnut St. Philadelphia, PA. 19104, USA. Email: bhatiasu@sas.upenn.edu.

Abstract

Considerable amount of laboratory and survey-based research finds that people show disproportional compassionate and affective response to the scope of human mortality risk. According to research on “psychic numbing”, it is often the case that the more who die, the less we care. In the present paper we examine the extent of this phenomenon in verbal behavior, using large corpora of natural language to quantify the affective reactions to loss of life. We analyze valence, arousal, and specific emotional content of over 100,000 mentions of death in news articles and social media posts, and find that language shows an increase in valence (i.e. decreased negative affect) and a decrease in arousal when describing mortality of larger numbers of people. These patterns are most clearly reflected in specific emotions of joy and (in a reverse fashion) of fear and anger. Our results showcase a novel methodology for studying affective decision making, and highlight the robustness and real-world relevance of psychic numbing. They also offer new insights regarding the psychological underpinnings of psychic numbing, as well as possible interventions for reducing psychic numbing and overcoming social and psychological barriers to action in the face of the world’s most serious threats.

Keywords: psychic numbing, big data, natural language, emotions, sentiment analysis

We are incoherent in the value we place on protecting human lives. We are motivated to expend great effort and money to protect specific individuals in distress, but become numbly indifferent to the fate of thousands of victims of genocides and other human crises. It took photographs of the body of three-year-old refugee Alan Kurdi, lying face down on a Turkish beach, to become front page news worldwide. This triggered a strong desire to help Syrian war victims that had not been aroused by almost five years of bloodshed resulting in hundreds of thousands of statistical deaths (Slovic, Västfjäll, Erlandsson, & Gregory, 2017). In just 24 hours, a group that rescues migrants at sea amassed £250,000 in donations (Goldberg, 2015). Yet, the subsequent drowning of 700 refugees a few months later (Harlan, 2016) and related events involving even greater numbers of casualties received little media attention and compassionate responses.

This inconsistency between our reverence for individual lives and our indifference to mass tragedies has been remarked upon by keen observers of the human condition, such as Mother Theresa who famously said, “If I look at the mass, I will never act; if I look at one I will”. Some, in particular, have sensed the discrepancy between the strong feelings aroused by an individual in distress and the lack of emotion when many are in danger. The Nobel prize winning biochemist, Albert Szent-Gyorgi, worried about the risk of nuclear war, mused: “I am deeply moved if I see one man suffering and would risk my life for him. Then I talk impersonally about the possible pulverization of our big cities, with a hundred million dead. I am unable to multiply one man’s suffering by a hundred million”. Similarly, statistics of disaster have been described as “human beings with the tears dried off”.

There are numerous laboratory studies providing empirical evidence in support of these observations. The importance of protecting individual lives has been labeled “the singularity effect" (Kogut & Ritov, 2005; Västfjäll, Slovic, Mayorga, & Peters, 2014) and the emotional indifference to mass suffering has been linked to a mental state known as "psychic numbing" (Slovic, 2007). In the present research, we offer new evidence of the scale of the insensitivity to the loss of human life by moving beyond laboratory demonstrations of these findings. Specifically, we take advantage of recent advances in computational linguistics, combined with the availability of large natural language data sets, to examine support for psychic numbing in news articles and in social media discussion forums. The sheer volume and richness of available data allow us to compare the affective valence and arousal associated with different types of deaths, as well as the exact emotions that are expressed in language pertaining to human death. Our hope is that by studying psychic numbing in natural settings we can better understand the robustness and real-world relevance of this phenomenon. This can give us novel insights regarding ways to mitigate psychological insensitivity to the large-scale loss of life, and lead to more effective policy measures to prepare for and respond to disastrous events.

Psychic Numbing

The combination of singularity and psychic numbing has been characterized as a nonrational arithmetic of compassion as it implies that the more who die, the less we care (Slovic, 2007). Numbing has been invoked to explain societal indifference and inaction in the face of catastrophic threats to humans and nature associated with genocide, war, climate change, and numerous other threats to populations from poverty, disease, natural disasters and the like (Slovic & Slovic, 2015).

Research on numbing is closely associated with the broad literature on the interplay between cognition and affect in judgment and decision-making. A large body of evidence shows that fast and automatic affective responses can motivate behavior, influence people's preferences, and attach meaning to information (Västfjäll et al., 2014). Models such as the affect heuristic (Slovic, Finucane, Peters, & MacGregor, 2007), risk as feeling (Loewenstein, Weber, Hsee, & Welch, 2001) and dual-system theories (Evans, 2008) attribute a central role to affective responses, which can contribute to, and at times, even override cognitive evaluations of risks. In the context of psychic numbing and the singularity effect, the emotional response to risk appears misaligned.

A large body of empirical evidence shows that affective reactions are fundamental to the value that people place on human life. For example, people choose to donate much more money to aid an identifiable victim (e.g. single named child or family) than to unidentifiable or statistical victim (Jenni & Loewenstein, 1997; Kogut, 2011; Kogut & Ritov, 2005; Small & Loewenstein, 2003; Smith, Faro, & Burson, 2013). In a series of studies, Västfjäll and colleagues (2014) showed that self-reported feelings of compassion and amounts donated begin to dissipate as soon as the number of victims is larger than one. Consistent with the idea that affect is central for explaining numbing and singularity effects, interventions that increase feelings of sympathy or compassion increase people's propensity to provide help when the number of victims or people in need is larger than one. For example, providing more detailed information about the charity and including vivid and affect-rich presentation formats of the victims increase donated amounts (Cryder, Loewenstein, & Scheines, 2013; Dickert, Kleber, Peters, & Slovic, 2011). Arousal also plays a central role in determining how people evaluate the scale of a tragedy. Participants who saw images rather than silhouettes of orphans in need donated more and showed heightened activation in the brain regions responsible for positive arousal (NAcc) (Genevsky, Västfjäll, Slovic, & Knutson, 2013).

Taken together, these studies show that people fail to respond appropriately to mass murder, atrocities and suffering of a large number of individuals. The common theme underpinning this research is that affective reactions are central to eliciting a response to such events, but that large atrocities do not generate proportional reactions to motivate action (Slovic, 2007). As such, our affective reactions can be misleading, and general insensitivity to the most severe disasters is a manifestation of this aspect of our psychology (Slovic & Västfjäll, 2010).

From the perspective of public policy, one might argue that the size of response to a tragedy should be proportional to its magnitude. More specifically, the amount of effort devoted to saving 100 lives should be 10 times larger than the amount of effort used to save 10 lives. In line with the research on the role of affect in guiding people's evaluative judgments and choices, this relation appears nonlinear when estimates are based on people's self-reported emotions or choices (e.g. how much to donate). Consistent with the research on the judgments of psychophysical quantities and general research on judgment and decision-making, the value people assign to human life appears to be "inversely proportional to the magnitude of the threat" (Fetherstonhaugh, Slovic, Johnson, & Friedrich, 1997). Summers, Slovic, Hine, and Zuliani (1999) also estimated the coefficient of a power function to be 0.32 for the sensitivity to deaths from wars. This type of concavity in the number of deaths suggests a rather severe numbing whereby people become very insensitive to the variations in the number of deaths in the most catastrophic events. The same drop-off was observed for positive affect, donation amounts, and facial EMG measures (Västfjäll et al., 2014). In discussing implications for the theory of human value, Vastfjall et al. propose that fading of compassion may result in an inverted u-shaped value function for gains and a u-shaped value function for losses.

One of the limitations of past research that we seek to address in the present work, is that most of the evidence for numbing comes from either survey-based or laboratory studies, with either donation amounts or self-reported feelings as the key dependent variables. This research has led the way in defining the psychological processes underpinning insensitivity to disasters and suffering. Yet the use of laboratory and survey data is necessarily limiting. Explicitly asking individuals to rate how they feel in response to an experimenter-generated scenario may not be an accurate reflection of the emotions they experience when they consider real-world tragedies: Participants may be sensitive to experimenter demand effects and may have an imperfect assessment of their own emotions (Schwarz, 1999). For this reason, it is unclear whether psychic numbing persists in the types of natural environments that influence affective responses and guide actions in response to real disasters. Survey and laboratory experiments are also limited in the amount of data they collect. Typically, surveys involve a few hundred participants evaluating a handful of hypothetical scenarios. Thus, there is a lack of fine-grained data that could illustrate the full extent to which affective responses to life diminish with magnitude, and the specific emotions that are most sensitive to the large-scale loss of life.

We offer a solution to the above challenges in the present paper. Specifically, we utilize large corpora of natural language to measure affective reactions to loss of life of varying magnitudes. Our approach is novel in that we can use large volumes of unsolicited online text data to quantify the sentiment associated with real human deaths. In the subsequent section we summarize our approach, illustrating how it presents an opportunity to a) provide a naturalistic “field” assessment of the affective responses to deaths, and b) provide a fine-grained measure of how different emotions change with the magnitude of the event’s severity.

Affective Responses in News and Social Media

The rise of the digital era has radically altered the way in which we obtain information, and has transformed social interaction and behavior. We now learn about events in the world through news websites and social media forums, and by doing so leave behavioral footprints that, in real-time and on an unprecedented scale, reveal what we know and how we feel. These footprints open up new opportunities to study human behavior, and have paved a way for a new-type of "big data" research in the social and behavioral sciences, including in fields such as management (George, Osinga, Lavie, & Scott, 2016), public health (Hawn, 2009), cognitive science (Griffiths, 2015), marketing (Humphreys & Wang, 2017), and psychology (Harlow & Oswald, 2016; Kosinski & Behrend, 2017). In all of these fields, large amounts of digital data from social media, news media, and mobile devices, offers opportunities to test and extend existing theories, which have previously been confined to more traditional methods, such as small laboratory experiments or surveys.

One type of digital data that has transformed behavioral research is natural language data, i.e. text that is consumed (and generated) by individuals, and used to form beliefs, express emotions, communicate ideas, and influence behavior. Combined with recent advances in computational linguistics, text data from online news corpora and social media has allowed researchers to study a number of important topics in the behavioral sciences, including (in our own work) people's perceptions of mental health problems (McCaig, Bhatia, Elliott, Walasek, & Meyer, 2018), the associative underpinnings of people's risk perceptions (Bhatia, 2018), psychological responses to uncertainty (Bhatia, Mellers, & Walasek, 2019), psychological distance (Bhatia & Walasek, 2016), emotion and gender in organizations (Gallus & Bhatia, 2020), and the effect of inequality on consumer behavior (Walasek, Bhatia, & Brown, 2018).

Whereas there are many ways to use text data to study behavior, here we focus on methods that allow for the quantification of affect and emotion in text. Such methods are typically referred to as sentiment analysis (Pang & Lee, 2008; Liu, 2012), and have numerous practical applications in marketing, communication, human-computer interaction, and other fields. For the purposes of this paper we use sentiment analysis to measure affective reactions to texts that discuss the loss of some number of lives. Drawing on theories of affect and emotion in psychology, we analyze the valence and arousal of these texts. Valence and arousal are two dimensions that capture the majority of the variance in the structure of emotional experience (Russell, 1980). In the context of text analysis, valence corresponds to the overall positive or negative qualities of a text, and arousal corresponds to the degree to which a text connotes intensity and activation. We also examine the expression of six emotions in text: sadness, joy, anger, disgust, fear, surprise. These six emotions are considered to be “basic”, that is innate and universal (Ekman, 1992). Valence is our main variable of interest, as it most closely relates to the affective measures used in prior research (Västfjäll et al., 2014), and is the closest affective construct to the valuation measures used in the study of psychic numbing (e.g. Fetherstonhaugh et al., 1997). However, we include an exploratory analysis of arousal and emotion to provide additional nuance to our results and to showcase the value of natural language processing methods for the study of affect and behavior.

We apply sentiment analysis to text from three large corpora: a corpus of the New York Times articles (NYT), a corpus of a representative set of online news media articles (News on the Web – NOW), and a corpus of posts made on prominent social media forum, Reddit.com, for discussing news (RED). The first two of these datasets involve text that is written by professionals (journalists and writers), and is intended to communicate information about newsworthy events in the world. Out of these, the NYT dataset contains articles published between 1987 and 2007, whereas the NOW dataset contains articles published between 2010 and 2016. The third dataset involves text generated by the consumers of news media as they discuss the news events, written between 2007 and 2016.

Together these three datasets give us a large and diverse sample of digital data. The NYT and NOW datasets allow us to study how news about death events is discussed and disseminated in traditional media. The NYT dataset has news articles written over a large span of time, allowing us to include historical data in our analysis. The NOW dataset has news articles that are published online, allowing us to include more recent types of death events, as well as articles published by newer and less established news organizations. Finally, the RED dataset gives us a perspective onto how lay individuals discuss death events, and allows us to avoid biases in our analysis that could stem from journalistic practices or newspaper editorial guidelines.

Overview of Analyses and Hypotheses

Our analysis also involves the development and application of an automated natural language processing algorithm for detecting whether or not a text (news article or social media post) mentions the deaths of individuals, and for quantifying the number of deaths referred to in the text. With these techniques we can rigorously measure affective reactions to the loss of life, and with the great amount of data available to us, we can establish how these affective reactions vary as a function of the number of deaths involved. Our main hypotheses, inspired by the findings of Västfjäll et al. (2014) and others, involves a U-shaped relationship between valence and the death count in the text, with the lowest valence for intermediate numbers of deaths (relative to no-deaths or extremely large numbers of deaths). Although our analysis of the six basic emotions is exploratory, we expect to observe a similar U-shaped relationship for positive emotions and the opposite (inverse-U) relationship for negative emotions. The magnitudes of these effects may differ between individual emotions (e.g. stronger effect on sadness vs. disgust – e.g. Smith & Ellsworth, 1985). We do not have precise predictions for arousal, however, if arousal is negatively correlated with valence in our data then we would expect to observe the opposite of the U-shaped relationship for valence. Valence and arousal are distinct independent dimensions (Russell, 1980), and this kind of negative correlation is not typically the case in neutral contexts in which high-valence, high-arousal expressions (such as those connoting excitement and jubilation), are as likely as low-valence, high-arousal expressions (such as those connoting anger and fear). However, given that our analysis involves only death contexts (in which expressions of jubilation and excitement are rare, and valence and arousal thus may display a negative relationship), we could obtain an inverse-U shaped pattern for arousal. Finally, we do not have predictions for differences between the text datasets, though we would not be surprised if we observed different patterns in the two news dataset (NYT and NOW, which are written under the constraints of editorial guidelines) relative to the social media dataset (RED, which involves informal discourse without editorial constraints). In contrast, we expect to observe similar patterns in the two news datasets.

Before proceeding it is useful to recognize some limitations of our proposed tests. Although the sentiment analysis methods are able to track affect in language, they are unable to directly capture either the emotion of the writer or the reason why the writer chose to discuss the event in consideration. Writers, who are journalists, restricted by editorial guidelines, and social media users, motivated by social concerns, are influenced by a number of factors that may not be at play in the general population (note that most social media users are also "lurkers", who don't directly contribute social media content). That said, since news and social media influences the opinions and attitudes of the general population, analyzing the emotionality of this kind of media offers a unique lens into lay psychology, on a very large scale. This is also why prior work has used sentiment analysis of news articles and social media posts to study wellbeing, emotion, as well as cultural and social attitudes (e.g. Michel et al., 2011; Hills et al., 2019).

Methods

Datasets

We implemented our analysis in three large and diverse natural language datasets:

1. The New York Times Annotated Corpus (NYT) – A dataset of over 1.8 million New York Times articles published from 1987 to 2007 (Sandhaus, 2008).
2. The News on the Web Corpus (NOW) – A dataset of over 5.5 million English language news articles from hundreds of different media sources, published on the internet between 2010 and 2016 (available at https://corpus.byu.edu/now/).
3. Reddit Politics and World News Comment Corpus (RED) – A dataset of over 73 million user comments made on the politics and world news discussion forums (www.reddit.com/r/politics and www.reddit.com/r/worldnews) on the social media website Reddit, between 2007 and 2016.

Extracting Expressions

In each dataset, automated text analysis techniques identified references in articles and posts to events in which people died. We describe two steps in which two automated algorithms were used to extract relevant phrases and then correct possible errors. Our approach was to first find all expressions of the form *N PHRASE*, where *N* is some number (e.g. *fifteen* or *15*), and *PHRASE* is some verbal expression of death (e.g. *people died*). We did so using the following algorithm:

1. Check if the text contains one of the 130 predetermined death phrases. If yes, move to step 2; else ignore text. Details regarding the death phrases are provided in the Appendix.
2. Check if the word preceding the death phrase is a number word (e.g. *fifteen*) or a number (e.g. *15*). If yes, move to step 3; else ignore text.
3. If the word preceding the death phrase is a number (and not a number word), then save it as the number of deaths associated with the phrase, and exit; if it is a number word move to step 4.
4. Go backwards sequentially through all the words preceding the death phrase until you reach a non-number word.
5. Extract all words between the non-number word and the death phrase, and convert words to a numerical format using the Python word2number module; save this as the number of deaths associated with the phrase, and exit. The word2number module is a Python program built to convert number words into numerical digits and is available at https://github.com/akshaynagpal/w2n.

We also tested for the context of the deaths mentioned in the text. Thus for each text containing a death phrase preceded by a number (i.e. each text passing step 2 above), we also checked whether that text contained words in the set {*murder*, *suicide*, *accident*, *crime*, *terrorism*, *genocide*, *war*, *holocaust*, *disease*, *famine*, *drought*, *flood*, *earthquake*, *hurricane*, *tornado*, *cyclone*}.

We also wished to compare the emotional content of texts extracted using the above method with the emotional content of texts in which there are no mentions of death. The emotionality of texts that do not mention deaths serves as a useful baseline measure for interpreting the magnitude of our results and for confirming that our algorithm works as intended. In order to perform this analysis, we needed a representative sample of texts that did not contain any of our 130 death phrases. We obtained this sample by randomly selecting 5% of the NYT and NOW articles and 1% of the RED comments that failed step 1 in the above algorithm.

The following additional steps were taken to correct phrases from our initial algorithm that were incorrectly identified.

1. Check if the two words preceding the death phrase are *no one*; if so, exclude text. The above algorithm incorrectly identifies *no one died* and related phrases as events involving one death, which is why death phrases preceded by *no one* must be ignored.
2. Check if the number preceding the death phrase is a year between 1800 and 2019; if so, exclude text. All instances of the form *YEAR PHRASE* that we have found in the data (e.g. *1987 murders*) refer to deaths happening in a particular year, rather than an event involving hundreds or thousands of deaths.
3. Check if the word preceding the number is a month or a month abbreviation; if so, exclude text. The numbers in such cases refer to dates (e.g. *June 20 murder*), rather than the number of deaths involved in the event.
4. Check if the words *age* or *aged* appear just before the number word; if so, exclude text. The numbers in such cases refer to ages (e.g.

aged 88 died last night) rather than the number of deaths involved in the event.

5. Check if the two words preceding the death word have the format *NUMBER billion*, *NUMBER million*, *NUMBER thousand*, or *NUMBER hundred*, where *NUMBER* is some number (e.g. in the case *2.1 million*). If so update the number of deaths associated with the phrase to *NUMBER*X*, where *X* corresponds to 1,000,000,000, 1,000,000, 1,000 or 100, depending on the number word (e.g. *2,100,000*). The initial algorithm allows for only numerical representations (e.g. *15*) or verbal representations (e.g. *fifteen*), and thus classifies phrases like *2.1 million* as referring only to a million deaths, which is why this had to be explicitly corrected.
6. Check if the words preceding the death word have the format *half a X*, *half X*, *quarter of a X*, or quarter *X*, where X is either *billion*, *million*, *thousand*, or *hundred.* If so, update the number of deaths associated with that phrase with to 0.5**X* or 0.25**X* depending on *X*. The initial algorithm cannot identify such verbal representations of fractions, which is why this had to be explicitly corrected.
7. Check if the words preceding the death word have the format *tens of X* or *hundreds of X*, where *X* is either *billions*, *millions*, *thousands*, or *hundreds*. If so, update the number of deaths associated with that phrase to 10**X* or *100*X* depending on *X*. The initial algorithm cannot identify such verbal representations, which is why this had to be explicitly corrected.

Finally, note that both algorithms are relatively conservative when classifying the number of deaths in a given text. Particularly, in cases of ambiguity in the exact number of deaths, they implicitly classify the text based on the lower bound on range of deaths. For example, these algorithms classify texts referring to *millions of deaths* as involving 1,000,000 deaths. Likewise, these algorithms classify texts with statements such as *at least fifteen people died* as involving 15 deaths.

Sentiment Analysis

We analyzed the emotional content of the texts extracted with the methods described in the previous section. Our approach performed automated sentiment analysis on these texts, using affective ratings released by Warriner, Kuperman, and Brysbaert (2013). In this dataset, valence is defined as the overall positive or negative quality of a word, whereas arousal is defined as the degree to which the word connotes excitement, intensity, and activation. The Warriner et al. dataset has valence and arousal ratings for nearly 14,000 English words, and is the largest such lexicon currently in existence. These ratings were collected by asking US participants to evaluate the words on a scale of 1 to 9 (with 1 indicating words with the lowest valence or arousal, and 9 indicating words with the highest valence or arousal). Ratings were averaged across participants to generate a single valence or arousal measure for each word. The highest valence words in the Warriner et al. dataset are *vacation* and *happiness* (average ratings of 8.53 and 8.48, respectively), the lowest valence words are *pedophile* and *rapist* (average ratings of 1.26 and 1.30, respectively), the highest arousal words are *insanity* and *gun* (average ratings of 7.79 and 7.74, respectively), and the lowest arousal words are *grain* and *dull* (average ratings of 1.60 and 1.67, respectively).

Our primary analysis uses the valence and arousal dimensions, as these have been argued to capture the majority of the variance in the structure of emotional experience (Russell, 1980). Although there are other datasets that could be used instead (e.g. Bradley & Lang, 1999), the

Warriner et al. dataset is the largest valence and arousal lexicon currently in existence. Unlike other lexicons, it has also been compiled by psychologists using human survey data. For this reason, this dataset is widely used in psychological research on emotion, language, memory, and decision making.

In analyzing the sentiment of each extracted text, we converted the text to lowercase, split it into individual words based on white spaces, and removed any numbers and death phrases from the text. For each word in the remaining text, we queried the Warriner dataset for the valence and arousal ratings for the word. If a particular word was not present in the Warriner dataset, the word was ignored. Average valence and arousal scores were obtained by summing valence and arousal ratings for the words in the text, and then dividing those by the total number of words in the text with ratings. For example, the calculated valence of text *i* using this method was: $V_i = \sum_{j=1}^{N} n_{ij} \cdot v_j / \sum_{j=1}^{N} n_{ij}$, with j = 1, 2, … N = 13,915 indexing words in the Warriner et al. data, n_{ij} capturing frequency of occurrence of word j in text i (removing number words and death phrases from the text), and v_j indicating the valence rating of word j in the Warriner et al. data (1 to 9, with higher values for more positively valenced words). Average arousal was calculated using the same formula.

In addition to examining valence and arousal in the Warriner dataset, we also performed automated content analysis examining how much of each of six basic emotions (sadness, joy, anger, disgust, fear, surprise) were expressed in each text. We did this using the NRC word-emotion lexicon (Mohammed & Turney, 2010), which contains binary ratings of over 14,000 words on these six basic emotions (as well as four other emotions, not analyzed in this paper). Specifically, each word in this lexicon is coded as 1 on the emotion if it was judged by participants as conveying that emotion, and coded as 0 otherwise. Using pre-processed text data split into individual words (as in the previous paragraph), we queried the Mohammed and Turney dataset to determine whether the word in a text is associated with each of the six basic emotions (sadness, joy, anger, disgust, fear, surprise). For each word with a match, we summed up the number of words in the text associated with each emotion, and then divided these scores by the total number of words in the text, to obtain the average sadness, joy, anger, disgust, fear, and surprise of the text. For example, the calculated sadness of text *i* using this method was: $S_i = \sum_{j=1}^{N} n_{ij} \cdot s_j / \sum_{j=1}^{N} n_{ij}$, with $j = 1, 2, \ldots N = 14{,}182$ indexing words in the Mohammed and Turney dataset, n_{ij} capturing the frequency of occurrence of word j in text i (removing number words and death phrases from the text), and s_j indicating the binary rating of word j on sadness in the Mohammed and Turney dataset (1 if it connoted sadness; 0 otherwise). Average joy, anger, disgust, fear, and surprise were calculated using an identical formula.

This approach, which averages the affective qualities of the constituent words in a text, has considerable precedent in prior work, and has been shown be a simple and effective way of measuring the affective qualities of the text. Additionally, the Warriner et al. dataset is a much larger version of the emotion and valence lexicons that are common in similar natural language applications (e.g. sentiment analysis). However, unlike other lexicons, the Warriner et al. dataset has established psychometric properties which makes it better suited for psychological applications such as ours. By asking human participants to rate individual words it is also better able to approximate the emotionality of texts involving these words. This is one reason why this is a popular dataset for psychological applications such as ours (see e.g. Gallus & Bhatia, 2020 and Hills et al., 2019 for recent work using both this averaging approach and the Warriner dataset).

Finally, although our primary analysis involved analyzing the emotional context of the entire text (news article or user comment) mentioning the death phrase, we also applied the above algorithms to words in the immediate vicinity of the death phrase. In particular, as a robustness test, we specified a 750-character radius around the death phrase in each text and measured the emotional content of only the words in this window. As proximity in text is a measure of relatedness, restricting our analysis in this manner reduces the likelihood that we incorrectly analyzed language referring to events other than the focal deaths in each text.

Human Coding

In order to ensure that our automated algorithms for identifying the number of deaths in a given article or comment were accurate, we selected a subset of our data for human coding. Specifically, we first divided our texts into four categories based on the numbers of deaths they mentioned: 1-9 deaths, 10-99 deaths, 100-999 deaths, and 1,000+ deaths. We then randomly selected 200 texts from each of these four categories from each of the three datasets (equal to a total of 2,400 texts) to give to two human coders. Our coders verified whether or not the texts actually referred to people dying, and whether or not the specific number of deaths identified by our algorithms were correct. We used the results of the coding to update our flagging algorithm (see preregistration section in our Appendix for a summary of changes resulting from the human coding). We also used the coding results to measure the accuracy of our algorithm's classification. Note that our use of the four death categories here was to ensure that we obtained texts refencing both large and small numbers of deaths (without this measure we would have had relatively few texts with a large number of deaths). Below we also use these death categories to discretize our data and understand the patterns in the data in a more intuitive manner.

Overall, we found that our algorithm correctly identified the number of deaths in 98.15% of texts in the NYT dataset, 97.46% of texts in the NOW dataset, and 93.69% of texts in the RED dataset. The lower accuracy for the RED dataset is due to the fact that user comments on Reddit often involve colloquial references that are not easily disambiguated by our algorithms. Besides this, a cursory examination of our coding data revealed that most of the errors in our classification involve texts that refer to animals or plants dying (rather than humans dying), or texts that refer to murderers or suicide bombers (rather than their victims). Automated disambiguation in these settings is difficult, however, as a robustness check, we repeated our analysis for the subset of death phrases that explicitly mention that a man, woman, child, or person died. Also, as a robustness test, we repeated our analysis with only the texts identified by coders as involving the correct classification of number of deaths.

Note that the above tests only examine whether or not texts identified as referring to some positive number of deaths actually refer to those deaths. These tests do not examine whether or not texts classified by our algorithm as involving no deaths actually contain some deaths. In other words, the above tests check for Type 1 errors (incorrectly classifying no deaths as some deaths) but not Type 2 errors (incorrectly classifying some deaths as no deaths). To evaluate these Type 2 errors, we further selected a set of 200 texts from each of the three datasets that were considered by algorithm as not involving any deaths (i.e. texts that did not contain any of the death phrases). These 600 texts were given to two human coders to examine whether or not the texts referred to some number of people dying. Overall, we found that human coders agreed with our algorithm for 93.5% of texts in the NYT dataset, 97.5% of texts in the NOW dataset, and 99% of texts in the RED dataset, indicating very low Type 2 error rates.

Preregistration

The death phrases, datasets, and algorithms were all preregistered at https://osf.io/4uhyz/. We closely followed our registration, with some small exceptions. These were mainly motivated by the difficulties of accurately extracting death phrases from the corpora, which we could not anticipate prior to collecting our data. We summarize all deviations from the pre-registration in the Appendix.

Results

Summary of Data

There were a total of 16,152 NYT articles, 80,912 NOW articles, and 21,214 RED comments that mention one or more death (and passed our exclusion criteria discussed above). The distribution of the natural log deaths for these articles is shown in Figure 1. Note that there is a spike in frequency around one million (Log(number) = 14), especially in the RED dataset. This is due to a large number of hyperbolic references to "million deaths" in this dataset (e.g. *millions of people will die if …*). As a robustness test, we repeated our analysis with all texts mentioning one million deaths excluded.

The distribution of number of deaths in our texts is also given in Table 1. This table also shows the distribution of texts that mention different contexts of death. Here, for example, *accident* refers to texts that refer to one or more deaths, and mention *accident* at least once. Table 1 indicates that the distributions of deaths and death contexts are roughly similar across the NYT and NOW datasets, both of which involve news articles. We do find some differences between these two datasets and the RED dataset, which involves user comments on social media news discussion forums. The RED dataset includes more frequent references to events resulting in

1,000+ deaths. Correspondingly, we also see more frequent references to genocides and the holocaust in RED relative to NYT and NOW. RED also has fewer references to natural disasters (e.g. floods, cyclones, hurricanes, earthquakes, and droughts).

Emotionality and Number of Deaths

As a first step to studying how the emotional content of texts varies with the number of deaths and the context of the death, we again divided our datasets into a set of discrete categories. These categories were: 1. Texts that do not mention any deaths; 2. Texts that mention between 1-9 deaths; 3. Texts that mention between 10-99 deaths; 4. Texts that mention between 100-999 deaths; and 5. Texts that mention 1,000+ deaths. We then analyzed the average emotionality of texts in each of these categories based on our emotional content scores. The output of this analysis for each of the three datasets, is shown in Figure 2 for valence and arousal, and Figure 3 for the six basic emotions. Error bars indicate +/- one standard error (though in places these standard errors are too small to be visible), and red lines indicate maximum or minimum emotionality in the dataset. In the appendix we provide additional plots for death categories 2 to 5.

As can be seen in Figure 2, texts that do not mention any deaths are higher in valence (i.e. more positive) than texts that involve mention of some deaths ($p < 10^{-15}$ for all pairwise comparisons with the four death categories for all datasets). This is a useful sanity check on our methods and confirms that algorithm is behaving as expected. More interestingly, however, there is a non-monotonic relationship between the valence of the text and the number of deaths referred to in the text. Although the highest valence texts are those that do not mention any deaths, the lowest valence texts are not those that mention 1,000+ deaths. Rather for both the

NYT and NOW datasets we see that valence eventually increases (i.e. language becomes more positive) as the number of deaths increases, with the lowest valence when there are between 10 and 99 deaths in the text ($p < 10^{-4}$ for all pairwise comparisons between average valence of this death category and average valence of all other death categories). This trend leads to articles mentioning 1,000+ deaths as having a much higher valence than articles mentioning 1-9 deaths, 10-99 deaths, or 100-999 deaths ($p < 10^{-13}$ for all pairwise comparisons between average valence of this death category and average valence of all other death categories). This is consistent with our hypotheses and with prior research on psychic numbing.

Such a relationship does not exist in the RED dataset. Here valence is the lowest when the text mentions 10-99 deaths, but this is not significantly different from the valence of texts mentioning 1,000+ deaths. In general, although valence drops as we move from texts that do not mention deaths to texts that mention some deaths, it appears that valence does not change in a systematic manner as the number of deaths increases in the RED dataset.

Figure 2 also shows arousal as a function of the number of deaths in each text. Here we can see significantly higher arousal in texts that mention deaths compared to texts that do not mention deaths ($p < 10^{-10}$ for all pairwise comparisons with the four death categories for all datasets). We also observe a non-monotonic relationship between arousal and the number of deaths mentioned in the texts, with arousal peaking at either 1-9 deaths or 10-99 deaths, depending on the dataset. Although the changes in arousal across the death categories are relatively smaller, we observe significantly lower arousal for texts mentioning 1,000+ deaths relative to texts mentioning 1-9 deaths or 10-99 deaths for all three datasets ($p < 10^{-6}$ for all pairwise comparisons). These results confirm our initial suspicions that arousal and valence may have opposite effects in the datasets.

Figure 3 shows similar patterns regarding expressions of different emotions, in all three of our datasets. Visibly, fear and anger are all expressed significantly less in texts that do not mention any deaths relative to texts that mention some deaths for all datasets ($p < 10^{-15}$ for all pairwise comparisons). Figure 3 also shows that the non-monotonicities documented above for valence and arousal emerge systematically with these emotions in all three datasets, with fear and anger being less frequently expressed in texts involving 1,000+ deaths compared to texts involving 1-9 deaths ($p < 10^{-3}$ for all pairwise comparisons).

We do observe some of the above trends for sadness, though these results are much less robust than the effects of fear and anger. Particularly, we do find that sadness is expressed significantly less in texts that do not mention any deaths relative to texts that mention some deaths, for all datasets ($p < 10^{-07}$ for all pairwise comparisons). However, we do not find non-monotonicities in all datasets. For example, sadness appears to be roughly stable for the NYT dataset amongst the texts that mention at least one death. There is a drop in sadness in the NOW and RED datasets as the number of deaths increases, though the effect in the RED dataset doesn’t appear to be systematic.

The relationships documented for valence in Figure 2 persist for the positive emotion: joy. Joy is expressed significantly more in texts that do not mention any deaths relative to texts that mention some deaths for all datasets ($p < 10^{-16}$ for all pairwise comparisons). Additionally, for the NYT and NOW datasets, joy displays a non-monotonic relationship with the number of deaths, with higher joy for 1,000+ deaths relative to 1-9, 10-99, or 100-999 deaths ($p < 10^{-16}$ for all pairwise comparisons). As in Figure 2, this relationship does not emerge in the RED dataset, with joy remaining roughly stable for texts involving 10 or more deaths.

Finally, we do not observe systematic trends for disgust or surprise. Disgust appears to increase mildly as the number of deaths increases for the NYT and NOW datasets, but not for the RED dataset. Likewise, surprise remains roughly stable as the number of deaths increases. These trends are likely due to the fact that both surprise and disgust are expressed relatively infrequently in our datasets (and are not expressed much more or much less frequently in texts mentioning some deaths vs. texts mentioning no deaths).

The results shown in Figures 2 and 3 are based on the automated coding of language, and involve the calculation of word frequencies for thousands of words across different texts in the three datasets. Valence, arousal, and emotion ratings are obtained by computing the aggregate emotionality of these words. In Figure 4 we hope to provide a more intuitive illustration of the features of our data responsible for the results shown in these figures. Thus, to generate Figure 4, we examined only twelve emotion words, corresponding to the six emotions used in Figure 3 (*anger, disgust, fear, joy, sadness,* and *surprise*) as well as their adjectives (*angry, disgusted, afraid, happy, sad,* and *surprised*). For each of the twelve words and each of our three corpora, we calculated the relative probabilities of occurrence of the words in each of our death categories, and displayed the death category in which the word had the highest occurrence probability. Thus, for example, we calculated the number of times the word *anger* occurs in NYT texts that mention 1-9 deaths, and divided this by the total number of words in NYT texts that mention 1-9 deaths, to determine the occurrence probability of the word *anger* in this death category for the NYT corpus. We did the same for NYT texts that mention 10-99 deaths, 100-999 deaths, and 1000+ deaths. As can be seen in Figure 4, the occurrence probability of *anger* is in fact the highest in the first death category relative to the remaining death categories. This is why anger is listed under the 1-9 deaths category for the NYT corpus in Figure 4. Overall, Figure

4 shows that the vast majority of negative emotion words are most likely to occur in death categories representing small and moderate amounts of deaths, and that only a few negative emotion words occur when the highest number of people have died. Note that the trends shown here are likely quite noisy (due to the fact that the analysis involves only individual words), and thus we cannot make robust statistical claims using this figure. Nonetheless, this figure provides an intuitive visual demonstration of the patterns discussed earlier on in this section.

Emotionality and Context of Death

We also examined expressions of valence, arousal, and the various emotions, as a function of the context of death. This is illustrated in Figure 5, which shows the average valence, arousal, anger, disgust, fear, joy, sadness, and surprise, in texts that mention each of the death context words (along with some number of deaths), in each dataset. Here we can see that the lowest valence and highest arousal in all three datasets emerges for death contexts such as terrorism, war, murder, suicide, and genocide. In contrast, death contexts involving natural disasters are typically associated with higher valence and lower arousal.

A similar pattern emerges for the six distinct emotions. For example, fear and anger emerge most strongly in death contexts involving terrorism, and most weakly in death contexts involving natural disasters (note that cyclone is associated with high fear in the RED dataset, but this is likely not statistically reliable --- there are only 9 comments in RED that mention cyclone). Sadness, in contrast, appears to be most pronounced in the context of natural disasters (as well as the holocaust). There doesn't seem to be a systematic relationship between joy/disgust/surprise and the various types of death contexts.

Regression Analysis

Although Figures 2 and 3 do show systematic effects of the number of deaths on expressions of emotion in our datasets, the analysis they present did not control for the death context. The results in the preceding section indicate that the context of the death does have an effect on emotional expression, suggesting that the results of Figures 2 and 3 may be confounding the effects of the number of deaths in the text with the context of death in the text. To avoid this issue, we ran a series of linear regressions in which we examined the relationship between the log-number of deaths (the independent variable) and the expression of emotions (dependent variables), controlling for the context of death. We ran one such regression for each emotion in each dataset. Each regression had fixed effects for each of the sixteen death contexts, as well as random effects for the 130 different death phrases that we used to capture mentions of death (thus allowing us to control not only for the context of the death but also the specific language used to describe the death). Importantly, we ran these regressions only on texts that mentioned at least one death. Thus, the outputs of these regressions capture the incremental effects a one-unit increase in the log-number of deaths has on the emotional content of the texts, given that the texts involve some mention of death. Also note that these regressions do not involve discretizing our data into a small set of categories (as in Figures 2 and 3) and thus allow for a more rigorous analysis of the relationship between the (continuous) number of deaths and the expressions of various emotions.

The outputs of these regressions are displayed in Figure 6. Here we can see the magnitude of the coefficients on the log-number of deaths (along with 95% confidence intervals) for each emotion in each dataset. Figure 6 shows that most of the results documented in the above sections persist even after we control for the context of the death and the specific death phrase involved. Particularly, for the NYT and NOW datasets, we see that the effect of log-number of deaths on valence is significantly positive ($p < 10^{-4}$ for both datasets). As in Figure 2, this pattern does not emerge for the RED dataset ($p = 0.37$), in which we observe a non-significant relationship between the log-number of deaths and the valence of the text.

Figure 6 also shows that the log-number of deaths has a systematically negative effect on arousal, with texts about events in which a large number of people die being less arousing than texts about events in which only a small number of people die. This pattern emerges for all three of our datasets ($p < 10^{-4}$ for all datasets), and supports the qualitative pattern illustrated in Figure 2.

Most of the results regarding the six discrete emotions, shown in Figure 3, also emerge in Figure 6. Particularly, we find that anger and fear are negatively related to the log-number of deaths in the three datasets, though the effects of anger are weaker in the NOW dataset ($p < 0.01$) compared to the NYT and RED datasets ($p < 10^{-4}$), and the effects of fear are weaker (and non-significant) in the NYT dataset ($p = 0.38$) compared to the NOW and RED datasets ($ps < 10^{-5}$). Again, this shows that fear and anger are expressed somewhat less frequently in texts that mention a large number of deaths compared to texts that mention a small number of deaths.

As in Figure 3, we also observe a less robust effect of sadness, with the NOW dataset showing a strong drop in sadness as the log-number of deaths increases ($p < 10^{-15}$), the RED dataset showing a mild but significant drop in sadness as the log-number of deaths increases ($p < 0.001$), and the NYT dataset showing a mild but significant increase in sadness as the log-number of deaths increases ($p < 0.001$). Likewise, as in Figure 3, we find that joy increases with the log-number of deaths in the NYT and NOW dataset ($p < 0.01$ and $p < 10^{-15}$ respectively), but that it decreases with the log-number of deaths in the RED dataset ($p < 0.001$). The effects of disgust are also similar to Figure 3, with log-number of deaths having a strong positive effect on disgust in the NOW and RED datasets ($p < 10^{-4}$), and a non-significant positive effect on disgust in the NYT dataset ($p = 0.14$). Unlike Figure 3, we observe a consistent negative effect of log-number of deaths on surprise for all three datasets ($p < 0.01$ for NYT and $p < 10^{-15}$ for NOW and RED). Note that the sizes of the coefficients for valence/arousal and the six emotions are not comparable, and the fact that the coefficients for the emotions are smaller does not necessarily indicate that log-number of deaths has a smaller effect on the emotions than valance and arousal. Rather these coefficient size differences are likely due to the fact that valence and arousal are rated on a 9-point scale, and thus the dependent variables in the valence and arousal regressions have a large scale and range. In contrast, the emotions are given a binary rating (1 or 0) and the dependent variables in the emotion regressions have a smaller scale and range.

In Table 2 we display the above coefficients, along with the coefficients of the sixteen death contexts, for each of the emotions. Consistent with Figure 5, we again see that different death contexts have different effects on valence, arousal, and our six emotions. Thus, for example, an article that mentions terrorism has lower valence and joy, and higher arousal, anger, disgust, fear, sadness and surprise. In Table 3 we provide additional outputs of our regression, including specific p-values and confidence intervals for the coefficient on the log-number of deaths, as well as R^2 and F-tests for each regression. These outputs again show that the coefficients are often highly significant (with $p < 0.001$ for most coefficients) and that our regressions fit the data well (with $p < 0.001$ for all F-tests). In the Appendix we show that these results persist when we introduce a quadratic component to the regression analysis to accommodate non-linear effects of number of deaths on the emotionality of the text.

Robustness Tests

We tested the robustness of the above results by implementing six different restrictions on our data. Our first robustness test involved analyzing only the words in the texts occurring in a 750-character radius around the death phrase. Although such a restriction omits a significant amount of potentially relevant text data from the analysis, it also reduces the likelihood that the analysis incorrectly includes text referring to events other than the focal deaths in each text. As can be seen in the left panels of Figure 7 this restriction has almost no effect on our results. The only coefficient that changes in sign is the coefficient for valence in the RED dataset: This moves from being non-significantly negative to being non-significantly positive (in line with the sign of the coefficients for valence in NYT and NOW).

Our second robustness test involved considering only the death phrases that explicitly refer to people dying. Instead of examining all 130 of the death phrases outlined above, we examined only the subset of the death phrases that mentioned men, women, children, or people. Again, this restriction omits a large amount of potentially relevant data from the analysis. However, it is also prevents the incorrect inclusion of references to deaths of animals or plants (rather than deaths of humans), which were the main causes of incorrect text classification identified by our human coders. As can be seen in the middle panels of Figure 7, such a restriction does not alter the main results of our analysis. In fact, the most relevant change again involves a reversal of the effect on valence in the RED dataset: Now, instead of having a non-significant negative effect, log-number has a significant positive effect on valence ($p < 10^{-3}$), in line with the effects observed for the NYT and NOW datasets. The other notable change involves the effect on sadness in the RED dataset, which reverses from being mildly negative to mildly positive ($p < 0.01$).

Our third robustness test involved an analysis of only the texts verified by human coders to be correctly classified. Although restricting our analysis in this manner greatly weakens our statistical power (as our sample size is reduced to less than 800 texts in each dataset – roughly 5% of the NYT data, 1% of the NOW data, and 4% of the RED data), it also ensures that there is perfect accuracy in our estimates of the number of deaths associated with each text. As shown in the right panels of Figure 7, this robustness test does not substantially alter the sign of our estimated coefficients (though our confidence intervals and p-values are much larger, due to the greatly diminished sample size). The only coefficients to change in sign are those pertaining to sadness and anger in the RED dataset and surprise in the NYT dataset (these go from being significantly negative to non-significantly positive).

In addition to the three robustness tests outlined above, we conducted three additional tests which excluded either texts mentioning only one death, texts mentioning a million deaths, or texts mentioning deaths with the use of the word *if*. The first of these robustness tests is necessary as events in which only one person dies are often described differently to events in which multiple people die. These descriptions often refer to the individual's name (e.g. *Joe Smith died*) rather than the total number of deaths in the event (e.g. *one person died*). Texts involving only one death are also more likely to be obituaries relative to texts involving multiple deaths. There are also minor grammatical differences involved in describing events referring to singular vs. plural deaths. These issues lead to the possibility that texts identified as involving only one death by our algorithms could be drawn from a systematically different sample of events relative to texts identified as involving more than one death. To avoid such a confound, we thus performed our regression analysis only on texts referring to two or more deaths.

The second robustness test is necessary as our data involves a large number of colloquial (and potentially hyperbolic) references to a million people dying. For example, more than 25% of the texts in the RED dataset involve exactly a million deaths. Although such references do fall within the bounds of our analysis (which involves all linguistic references to real or hypothetical events in which some number of people die), we did not want such references to skew our results. Thus, we ran our regression analysis again after excluding texts referring to a million deaths. The third robustness test also helps avoid the incorrect inclusion of hypothetical and counterfactual deaths. Such deaths are often mentioned with the word *if* and can be flagged by testing if *if* occurs within a 50 character radius of the death phrase.

The results of these three robustness tests are shown in Figure 8. As can be seen in this figure, excluding texts that refer to only one death and texts that refer to a million deaths makes only one small difference in our results, involving the effect of log-number on valence in the RED dataset. After these exclusions, this effect becomes (non-significantly) positive, in line with the valence effects observed in the NYT and NOW datasets. There is no qualitative change in our results after excluding texts that involve the use of the word *if* within a 50 character radius of the death phrase.

General Discussion

Why might people be so insensitive to the suffering of others in some cases, and yet be able to show great deal of compassion in other situations? According to the research on psychic numbing and the singularity effect, we care less when more people die, and this effect is presumed to be driven by the lessened affective reactions to tragedies that influence many. Here we tested these predictions using a novel method – by analyzing the sentiment of more than 100,000 texts mentioning the loss of human life in news articles and posts on social media forums. Our findings support and further elucidate the relation between the magnitude of a tragedy and people's affective responses to it.

First, there is a non-monotonic relation between overall valence and arousal of posts in our three text corpora. As one would expect, we found that, relative to texts not mentioning death, texts talking about the loss of life are lower in valence but higher in arousal. However, consistent with the psychic numbing hypothesis, valence does not become more negative and arousal does not increase as the number of deaths increases. Rather, the lowest valence and highest arousal are often observed for events involving a small number of deaths (1-9 or 10-99 deaths), rather than a large number (1,000+) deaths. These patterns are also clearly reflected in specific emotions of joy, fear and anger, which are most likely to occur in texts mentioning smaller rather than larger numbers of deaths.

Variation in emotional content is not independent of the type of event involving death, as the lowest valence and highest arousal are predominantly visible among texts discussing terrorism, war, murder, suicide, and genocide. This could be due to the fact that these deaths (unlike those from hurricanes, cyclones, earthquakes, droughts, and floods) reflect human agency, which has a strong effect on emotion (e.g. Lazarus, 1982; Smith & Ellsworth, 1985). Nonetheless, our statistical tests, which controlled for these death contexts, found a primarily positive association between valence and the number of deaths and a negative association between arousal and the number of deaths. Similar results emerged for the specific emotions of joy, anger and fear. Additionally, these results persisted when the data being analyzed was restricted to be in a 750-character radius around the mention of the death, when the death phrase used explicitly mentioned humans dying, when the data was restricted to the texts determined to be correctly classified by human coders, and when texts referring to only one death or to exactly a million deaths were excluded, suggesting these results are highly robust. Overall, consistent with the laboratory and survey-based findings on psychic numbing (Slovic, 2007; Västfjäll et al., 2014), our tests indicate that affective responses are paradoxically the most negative when a small number of people die. In other words, our findings are in line with the argument that the more who die, the less we care.

Our findings were rather consistent across the three databases, but there were some differences between the two news corpora (NOW and NYT) and the social media data from Reddit (RED). For example, we found a very large spike in occurrences of one million deaths on the user-generated content in the Reddit data. Also, in the Reddit data we did not find a robust positive association between the number of deaths and valence, with the valence being rather stable across the board. These findings highlight the unique nature of social media discourse. On one hand, social media forums are less constrained than heavily edited and moderated news articles. As such they may offer a better insight into the way in which general population converses about death. On the other hand, language posted in social media forums is likely to involve more hyperbolic expressions and colloquialisms that introduce a lot of noise. Indeed, accuracy of our extraction algorithm was the lowest in the case of Reddit (although it was still respectably high). Taken together, whereas our results are consistent with psychic numbing, more research is needed to understand how people talk about death and tragic disasters online.

Although news articles are influenced by journalistic practices and norms, and may not always reflect people's emotional responses to disasters, we believe that such datasets are suitable for our analysis for two reasons. Firstly, people's responses are often informed by what they hear on the news. This makes news articles a reasonable proxy for people's attitudes and beliefs. Indeed, prior work on risk perception has shown that perceived frequencies of causes of death and the societal impact of all manner of risk events are strongly influenced by reporting in newspapers (Burns et al., 1993; Combs & Slovic, 1979). Secondly, news articles provide a large and relatively controlled linguistic dataset for natural language analysis. For example, the NYT and NOW corpora do not have the types of typos and the colloquial references to death that we observed in our Reddit dataset. For this reason, news articles are important benchmark datasets in numerous natural language processing applications, and provide a useful starting point for the type of analysis performed in this paper. Nonetheless future work should attempt to replicate our tests with other types of natural language datasets that more directly reflect people's attitudes and opinions

Our results also open up a number of novel questions regarding the boundary conditions and nuances of psychic numbing. For example, in many of our tests we found that texts with a moderate number of deaths (e.g. 10-99 deaths) had the lowest affect. It is not immediately clear to us why this would be the case. We speculate that such events may involve singular groups (e.g. in a school shooting) and thus are more emotionally compelling than the deaths of singular individuals or large numbers of distinct individuals and groups. It is also not obvious why the specific emotions of joy, fear, and anger, should have the most pronounced psychic numbing effects. This could be because these emotions (unlike, for example, sadness) are especially high in arousal, and (unlike, for example, surprise) have unambiguously positive or negative valence. We also found that the vast majority of texts involved discussions of small numbers of deaths. For example, almost half of all texts referred to events with less than ten deaths. We speculate that this may also be a type of psychic numbing phenomenon, according to which events with only a few deaths are more salient and thus receive more attention in news and social media discourse. However, it is also possible that events with small numbers of deaths are more common than those with large numbers of deaths, and that the events frequencies in our texts are representative event frequencies in the world (e.g. Olivola & Sagara, 2009). Overall, these open questions are a product of the richness of the data analyzed in this paper and highlight the power of our methodology. We look forward to future work that uses our methodology (in conjunction, perhaps, with laboratory or survey data) to answer these questions and shed light on the complex causes and consequences of the psychic numbing phenomenon.

Despite these open questions, our results contribute to a more detailed understanding of psychic numbing. We provide independent evidence, outside of laboratory and survey-based methods, for the paradoxical relation between magnitude of a tragedy and affect. In the news and in social media discussions, the emotional response seems to be most pronounced for tragedies affecting fewer people, mirroring results of various studies and a great variety of real-world observations. Our results also align with the observations that value of life appears to follow a power law with a very low exponent and may even be captured by u-shaped function with the negative value actually decreasing for very negative events (Fetherstonhaugh et al., 1997; Vastfjall et al, 2014). The present evidence emerges from the way in which news and social media users talk about events involving death. Our findings do not rely on explicit reports or predictions of affect but are rather evident in the sentiment of the language people use.

The methodological approach taken in the present study illustrates the usefulness of investigating properties of natural language in relation to theories and effects established in behavioral sciences. As with commonly used empirical approaches, this method has its limitations. Thus, it is a complement, and not a replacement, for standard experimental protocols. Perhaps the most important caveat to our findings is that the validity of our analysis relies on the accuracy of the algorithm which we used to extract relevant texts. We revised our steps and complemented them with human coding, but it is possible that not every text considered here relates to human death. Still, the sheer volume of data and robustness checks that we performed suggest that our findings are indeed capturing emotional content of discourse pertaining to the loss of human life.

We consider our analysis part of a cumulative process (Jones, 2016) whereby our results contribute to the evidence established using experiments, surveys and also qualitative analysis. Sentiment analysis (Pang & Lee, 2008; Liu, 2012) is very powerful in characterizing people's attitudes and in predicting their future actions or choices (Cambria, Schuller, Xia, & Havasi, 2013; Das & Chen, 2007; Li & Wu, 2010). Still, one should not assume that emotional content in the online discourse is a perfect reflection of the true affective responses people have for death. We believe that our methods offer a good approximation of these feelings, but there are many exogeneous reasons as to why a person chooses to use emotive language when describing a tragic and disastrous event.

As news and social media play a vital role in informing us about the important threats to humans and the environment, we need to better understand the significance of the numbing and lessened arousal that we find in reports involving greater mortality. On the one hand, it is good that such reports appear, in a sense, dispassionate, as one might want "just the facts" without emotion. But emotion is known to be a necessary motivator of action. Quantitative data, for example, as vital as they are, can be numbing and demotivating and may generate systemic biases that lead people to avoid taking action before it is too late (Meyer and Kunreuther 2017). We hope that the present study begins to point the way toward future research that will lead to an effective balance between facts and feelings that stimulate discourse about the major threats to humans and the planet.

Acknowledgement. We thank the Alfred P. Sloan Foundation for supporting this research through Grant Number G-2018-11100.

References

Bhatia, S. (2019). Predicting risk perception: New insights from data science. Management Science, 65(8), 3800-3823..

Bhatia, S., Goodwin, G. & Walasek, L. (2018). Trait associations for Hillary Clinton and Donald Trump in news media: A computational analysis. Social Psychological and Personality Science, 9(2), 123-130. https://doi.org/10.1177/1948550617751584

Bhatia, S., Mellers, B., & Walasek, L. (2019). Affective responses to uncertain real-world outcomes: Sentiment change on Twitter. PLOS ONE, 14(2), e0212489. https://doi.org/10.1371/journal.pone.0212489

Burns, W. J., Slovic, P., Kasperson, R. E., Kasperson, J. X., Renn, O., & Emani, S. (1993). Incorporating structural models into research on the social amplification of risk: Implications for theory construction and decision making. Risk Analysis, 13, 611–623.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2), 15–21. https://doi.org/10.1109/MIS.2013.30

Combs, B., & Slovic, P. (1979). Newspaper coverage of causes of death. Journalism Quarterly, 56(4), 837–843, 849

Cryder, C. E., Loewenstein, G., & Scheines, R. (2013). The donor is in the details. Organizational Behavior and Human Decision, 120, 15–23. https://doi.org/10.1016/j.obhdp.2012.08.002

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. Management Science, 53(9), 1375–1388. https://doi.org/10.1287/mnsc.1070.0704

Dickert, S., Kleber, J., Peters, E., & Slovic, P. (2011). Numeracy as a precursor to pro-social behavior: The impact of numeracy and presentation format on the cognitive mechanisms underlying donation decisions. Judgment and Decision Making 6(7), 638-650.

Ekman, P. (1992). An argument for basic emotions. Cognition & Emotion, 6(3-4), 169-200.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. Annual Review of Psychology, 59(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Fetherstonhaugh, D., Slovic, P., Johnson, S. M., & Friedrich, J. (1997). Insensitivity to the value of human life: A study of psychophysical numbing. Journal of Risk and Uncertainty, 14(3), 283–300. https://doi.org/10.1023/A:1007744326393

Gallus, J. & Bhatia, S. (2020). Gender, power and emotions in the collaborative production of knowledge: A large-scale analysis of Wikipedia editor conversations. Organizational Behavior and Human Decision Processes. 160. 115-130.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. Academy of Management Journal, 59(5), 1493-1507.

Genevsky, A., Västfjäll, D., Slovic, P., & Knutson, B. (2013). Neural underpinnings of the identifiable victim effect: Affect shifts preferences for giving. The Journal of Neuroscience, 33(43), 17188–17196. https://doi.org/10.1523/JNEUROSCI.2348-13.2013

Goldberg, E. (2015). After Syrian boy drowns, donations pour in to help refugees and migrants. Huffington Post. Available at https://www.huffingtonpost.co.uk/entry/syrian-boy-drowns-donations_n_55e8a0b0e4b0b7a9633c5915

Griffiths, T. L. (2015). Manifesto for a new (computational) cognitive revolution. Cognition, 135, 21-23.

Harlan, C. (2016). More than 700 migrants have drowned in Mediterranean in past few days.

Washington Post. Available at https://www.washingtonpost.com/world/europe/up-to-700-have-drowned-in-mediterranean-in-past-few-days/2016/05/29/f7473eaa-25b1-11e6-8329-6104954928d2_story.html.

Harlow, L. L., & Oswald, F. L. (2016). Big data in psychology: Introduction to the special issue. Psychological Methods, 21(4), 447.

Hawn, C. (2009). Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health affairs, 28(2), 361-368.

Hills, T. T., Proto, E., Sgroi, D., & Seresinhe, C. I. (2019). Historical analysis of national subjective wellbeing using millions of digitized books. Nature Human Behavior, 3(12), 1343-1343.

Humphreys, A., & Wang, R. J.-H. (2017). Automated text analysis for consumer research. Journal of Consumer Research, 44(6), 1274-1306.

Jenni, K., & Loewenstein, G. (1997). Explaining the identifiable victim effect. Journal of Risk and Uncertainty, 14(3), 235–257. https://doi.org/10.1023/A:1007740225484

Jones, M. N. (2016). Big data in cognitive science. Routledge, London.

Kogut, T. (2011). Someone to blame: When identifying a victim decreases helping. Journal of Experimental Social Psychology, 47(4), 748–755. https://doi.org/10.1016/j.jesp.2011.02.011

Kogut, T., & Ritov, I. (2005). The singularity effect of identified victims in separate and joint evaluations. Organizational Behavior and Human Decision Processes, 97(2), 106-116. http://dx.doi.org/10.1016/j.obhdp.2005.02.003

Kogut, T., & Ritov, I. (2005). The “identified victim” effect: an identified group, or just a single individual? Journal of Behavioral Decision Making, 18(3), 157–167.

https://doi.org/10.1002/bdm.492

Kosinski, M., & Behrend, T. (2017). Editorial overview: Big data in the behavioral sciences. Current Opinion in Behavioral Sciences, 18, iv-vi. https://doi.org/10.1016/j.cobeha.2017.11.007

Lazarus, R. S. (1982). Thoughts on the relations between emotion and cognition. American psychologist, 37(9), 1019.

Li, N., & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decision Support Systems, 48(2), 354–368. https://doi.org/10.1016/J.DSS.2009.09.003

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. Psychological Bulletin, 127(2), 267–286. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11316014

Mccaig, D., Bhatia, S., Elliott, M. T., Walasek, L., & Meyer, C. (2018). Text-mining as a methodology to assess eating disorder-relevant factors: Comparing mentions of fitness tracking technology across online communities. International Journal of Eating Disorders. https://doi.org/10.1002/eat.22882

Meyer, R. & H. Kunreuther (2017), The Ostrich Paradox: Why We Underprepare for Disasters, Wharton Digital Press.

Michel JB, Shen YK, Aiden AP, Veres A, Gray MK, et al. (2011) Quantitative analysis of culture using millions of digitized books. Science 331: 176–182

Mohammad, S. M., & Turney, P. D. (2010). Emotions evoked by common words and phrases:

Using mechanical turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text (pp. 26-34). Association for Computational Linguistics.

Olivola, C. Y., & Sagara, N. (2009). Distributions of observed death tolls govern sensitivity to human fatalities. Proceedings of the National Academy of Sciences, 106(52), 22151-22156.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1-135.

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161.

Schwarz, N. (1999). Self-reports: how the questions shape the answers. American psychologist, 54(2), 93.

Sandhaus, E. (2008). The New York Times Annotated Corpus LDC2008T19. DVD. Philadelphia: Linguistic Data Consortium.

Slovic, P. (2007)."If I look at the mass I will never act": Psychic numbing and genocide. Judgment and Decision Making 2(2), 79-95.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. European Journal of Operational Research, 177(3), 1333–1352. https://doi.org/10.1016/j.ejor.2005.04.006

Slovic, S., & Slovic, P. (2015). The arithmetic of compassion. The New York Times. Available at https://www.nytimes.com/2015/12/06/opinion/the-arithmetic-of-compassion.html.

Slovic, P., & Västfjäll, D. (2010). Affect, moral intuition, and risk. Psychological Inquiry, 21(4), 387–398. https://doi.org/10.1080/1047840X.2010.521119

Slovic, P., Västfjäll, D., Erlandsson, A., & Gregory, R. (2017). Iconic photographs and the ebb and flow of empathic response to humanitarian disasters. Proceedings of the National Academy of Sciences of the United States of America, 114(4), 640–644. https://doi.org/10.1073/pnas.1613977114

Small, D. A., & Loewenstein, G. (2003). Helping a victim or helping the victim: Altruism and identifiability. Journal of Risk and Uncertainty, 26(1), 5–16. https://doi.org/10.1023/A:1022299422219

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. Journal of personality and social psychology, 48(4), 813.

Smith, R. W., Faro, D., & Burson, K. A. (2013). More for the many: The influence of entitativity on charitable giving. Journal of Consumer Research, 39(5), 961–976. https://doi.org/10.1086/666470

Summers, C., Slovic, P., Hine, D., & Zuliani, D. (1999). ’Psychophysical Numbing’: An Empirical Basis for Perceptions of Collective Violence. In C. Summers & E. Markusen (Eds.), Collective Violence: Harmful Behavior in Groups and Governments. Rowman & Littlefield.

Västfjäll, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion fade: Affect and charity are greatest for a single child in need. PLoS ONE, 9(6), e100115. https://doi.org/10.1371/journal.pone.0100115

Walasek, L., Bhatia, S. & Brown, G. (2018). Positional goods and the social rank hypothesis: Income inequality affects online chatter about high and low status brands on Twitter. Journal of Consumer Psychology, 28(1), 138-148. https://doi.org/10.1016/j.jcps.2017.08.002

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. https://doi.org/10.3758/s13428-012-0314-x

Table 1: Frequency distribution of deaths and death contexts across the three datasets, with percentage of total texts in dataset is shown in brackets.

	NYT	**NOW**	**RED**
Number of deaths			
1-9 deaths	7,720 (48%)	35,344 (44%)	11,803 (56%)
10-99 deaths	5,134 (32%)	30,063 (37%)	723 (3%)
100-999 deaths	2,659 (16%)	12,756 (16%)	499 (2%)
1000+ deaths	639 (4%)	2,749 (3%)	8,189 (39%)
Context of death			
Accident	1,520 (9%)	8,205 (10%)	292 (1%)
Crime	1,617 (10%)	6,612 (8%)	680 (3%)
Cyclone	21 (0%)	561 (1%)	9 (0%)
Disease	1,016 (6%)	3,218 (4%)	272 (1%)
Drought	192 (1%)	758 (1%)	13 (0%)
Earthquake	362 (2%)	2,395 (3%)	67 (0%)
Famine	144 (1%)	393 (0%)	322 (2%)
Flood	408 (3%)	3,046 (4%)	32 (0%)
Genocide	189 (1%)	935 (1%)	703 (3%)
Holocaust	186 (1%)	455 (1%)	736 (3%)
Hurricane	365 (2%)	1,015 (1%)	30 (0%)
Murder	1,757 (11%)	7,177 (9%)	1,575 (7%)
Suicide	1,631 (10%)	9,573 (12%)	916 (4%)
Terrorism	1,201 (7%)	6,615 (8%)	404 (2%)
Tornado	213 (1%)	732 (1%)	21 (0%)
War	4,638 (29%)	13,390 (17%)	3,482 (16%)
Total texts	16,152	80,912	21,214

*Table 2. Coefficients for log number of deaths (LogN) and death contexts in regressions predicting emotion. * indicates p < 0.05. Note that the coefficients of LogN have been multiplied by 1,000 and the coefficients for the various death contexts have been multiplied by 100. This has been done for display purposes as these coefficients are very small (this does not change their significance values or interpretation of results).*

NYT	**Valence**	**Arousal**	**Anger**	**Disgust**	**Fear**	**Joy**	**Sadness**	**Surprise**
LogN (x1000)	3.11*	-1.89*	-0.63*	0.11	-0.16	0.28*	0.43*	-0.19*
Accident (x100)	1.99*	-5.46*	-2.78*	-0.49*	-1.55*	-0.64*	0.94*	1.36*
Crime (x100)	-3.59*	3.32*	1.39*	0.48*	-0.03	0.08	0.08	-0.11
Cyclone (x100)	8.07	0.53	-0.87	0.57	-1.53	0.56	-0.62	1.73*
Disease (x100)	-6.91*	-0.75	-0.49*	2.62*	-0.18	0.01	1.34*	-0.51*
Drought (x100)	6.38*	-6.54*	-1.39*	-0.03	-1.92*	0.45	-0.41	-0.27
Earthquake (x100)	6.27*	-0.54	-0.11	0.17	-1.77*	-0.38*	-0.47*	-0.10
Famine (x100)	4.71*	-0.25	-0.93*	-0.30	-1.76*	0.91*	0.79*	0.07
Flood (x100)	6.62*	-5.17*	-0.69*	0.24	-1.51*	-0.08	-0.38	-0.21
Genocide (x100)	-4.39*	-0.12	0.83*	0.51*	0.92*	0.22	0.32	-0.45*
Holocaust (x100)	-0.97*	2.40	0.48	-0.07*	0.47*	0.39*	-0.78*	-0.14
Hurricane (x100)	1.53	1.17	0.57*	0.33*	0.03	0.14	2.26*	1.54*
Murder (x100)	-6.15*	4.63*	1.90*	0.84*	1.54*	0.33*	1.08*	-0.02
Suicide (x100)	-3.79*	4.14*	0.98*	-0.03	1.13*	0.13	0.49*	0.34*
Terrorism (x100)	-7.67*	4.38*	1.38*	1.05*	1.39*	-0.09	0.62*	0.37*
Tornado (x100)	6.27*	-1.67	0.07	0.66*	-0.61	0.11	0.88*	-0.16
War (x100)	-0.01*	0.02*	0 .00	0 .00	0 .00	0 .00	-0.01*	0 .00
NOW	**Valence**	**Arousal**	**Anger**	**Disgust**	**Fear**	**Joy**	**Sadness**	**Surprise**
LogN (x1000)	4.68*	-1.63*	-0.21*	0.71*	-2.15*	0.69*	-0.68*	-0.39*
Accident (x100)	-3.62*	-4.99*	-3.09*	-0.46*	-0.19*	-0.69*	2.99*	2.01*
Crime (x100)	-6.74*	4.44*	2.12*	0.56*	1.07*	0.02	0.13*	-0.27*
Cyclone (x100)	1.75	-1.68*	0.88*	0.42*	0.11	-0.86*	0.15	2.30*
Disease (x100)	-10.52*	-1.65*	-0.55*	3.25*	1.02*	-0.31*	2.12*	-0.87*
Drought (x100)	3.35*	-6.58*	-1.18*	-0.26*	-1.94*	-0.13	-0.62*	-0.48*
Earthquake (x100)	7.90*	-0.34	0.52*	0.07	-1.74*	-0.51*	-1.19*	-0.64*
Famine (x100)	3.08*	0.09	-1.26*	-0.24	-0.93*	1.09*	1.61*	0.12
Flood (x100)	6.59*	-7.24*	-1.68*	0.09	-1.56*	-0.11	-0.43*	-0.22*
Genocide (x100)	-4.37*	-0.21	0.35*	0.11	0.39	0.23*	0.49*	-0.32*
Holocaust (x100)	-1.38*	2.28	0.84	-0.08*	0.69*	0.4*	-0.52*	-0.11*
Hurricane (x100)	-3.30*	7.77*	1.22*	0.47*	1.69*	-0.10	3.19*	2.30*
Murder (x100)	-5.61*	4.24*	2.26*	0.84*	2.31*	0.21*	1.12*	-0.32*
Suicide (x100)	-5.99*	5.85*	1.13*	-0.38*	1.79*	-0.23*	0.21*	0.59*
Terrorism (x100)	-8.52*	7.89*	2.37*	1.64*	2.58*	-0.21*	0.77*	0.32*
Tornado (x100)	5.57*	3.89*	0.30	0.47*	-0.11	-0.13	0.03	-0.31*
War (x100)	-0.01*	0.02*	0.01*	0.00	0.01*	0 .00	-0.01*	0 .00
RED	**Valence**	**Arousal**	**Anger**	**Disgust**	**Fear**	**Joy**	**Sadness**	**Surprise**
LogN (x1000)	-0.62	-4.20*	-0.75*	0.51*	-2.71*	-0.41*	-0.68*	-0.87*
Accident (x100)	-12.39*	-0.16	-1.92*	-0.82	2.92*	-1.57*	4.37*	5.48*
Crime (x100)	-17.45*	7.91*	5.76*	0.25	-0.09	-0.33	-0.64	-0.60
Cyclone (x100)	-14.51	11.85	2.55	1.90	8.97	1.16	-3.13	10.31*
Disease (x100)	-17.63*	0.93	1.87*	3.93*	2.63*	-0.63	3.62*	-0.34
Drought (x100)	7.12	-9.28	-0.10	1.06	-2.07	1.13	-3.23	0.89
Earthquake (x100)	10.93	-6.74	-3.49	-1.67	-0.77	-0.42	-2.91	-1.43
Famine (x100)	-10.36*	5.67*	-2.81*	-1.75*	-2.59*	0.15	4.99*	-0.52
Flood (x100)	1.54	6.05	-2.96	-2.05	-1.23	0.89	-4.67	-1.08
Genocide (x100)	-25.25*	8.94*	-0.18	-0.42	-1.19	-0.03	-0.07	-0.05

Holocaust (x100)	-21.24*	12.07*	-0.55*	-1.03*	5.74*	-0.31*	-1.97*	-0.53*
Hurricane (x100)	-11.84	18.36*	4.14*	0.11	0.83	-1.45	3.63*	4.54*
Murder (x100)	-4.91*	-0.58	1.72*	0.94*	0.97	-0.39	0.78	-0.25
Suicide (x100)	-9.2*	8.57*	1.44*	-0.65	2.55*	0.02	1.50*	1.20*
Terrorism (x100)	-22.97*	14.79*	4.9*	5.29*	4.36*	-0.4	3.89*	-0.1
Tornado (x100)	7.65	2.88	0.73	0.80	1.59	-0.51	1.19	-0.58
War (x100)	-0.21*	0.12*	-0.01*	-0.01*	0.06*	0 .00	-0.02*	-0.01*

Table 3. Detailed outputs of our main regressions. The first four columns present coefficients (LogN), p-values (LogN-p), and 95% confidence intervals (LogN-L for lower and LogN-H for higher bounds) for log number of deaths. The remaining four columns provide the total number of observations (N), the R^2 values (R^2), the F-statistics (Fstat) and F-statistic p-values (Fstat-p) of the regressions.

NYT	**LogN** (x1000)	**LogN-p**	**LogN-L** (x1000)	**LogN-H** (x1000)	**N**	**R^2**	**Fstat**	**Fstat-p**
Valence	3.11	0.000	1.70	4.52	16,152	0.01	38.41	0.000
Arousal	-1.89	0.000	-2.74	-1.03	16,152	0.07	59.53	0.000
Anger	-0.63	0.000	-0.90	-0.36	16,152	0.08	80.36	0.000
Disgust	0.11	0.141	-0.04	0.26	16,152	0.10	93.47	0.000
Fear	-0.16	0.382	-0.53	0.20	16,152	0.03	28.33	0.000
Joy	0.28	0.003	0.10	0.46	16,152	0.01	15.25	0.000
Sadness	0.43	0.000	0.20	0.67	16,152	0.03	32.84	0.000
Surprise	0.43	0.000	0.20	0.67	16,152	0.04	55.29	0.000
NOW	**LogN** (x1000)	**LogN-p**	**LogN-L** (x1000)	**LogN-H** (x1000)	**N**	**R^2**	**Fstat**	**Fstat-p**
Valence	4.70	0.000	3.93	5.46	80,912	0.02	205.01	0.000
Arousal	-1.63	0.000	-2.12	-1.14	80,912	0.09	391.56	0.000
Anger	-0.21	0.005	-0.36	-0.07	80,912	0.13	604.66	0.000
Disgust	0.71	0.000	0.63	0.79	80,912	0.10	511.25	0.000
Fear	-2.15	0.000	-2.35	-1.95	80,912	0.04	230.86	0.000
Joy	0.69	0.000	0.59	0.79	80,912	0.02	64.05	0.000
Sadness	-0.68	0.000	-0.81	-0.55	80,912	0.07	343.44	0.000
Surprise	-0.39	0.000	-0.47	-0.30	80,912	0.07	347.57	0.000
RED	**LogN** (x1000)	**LogN-p**	**LogN-L** (x1000)	**LogN-H** (x1000)	**N**	**R^2**	**Fstat**	**Fstat-p**
Valence	-0.62	0.373	-1.99	0.75	20,999	0.04	55.21	0.000
Arousal	-4.20	0.000	-5.18	-3.21	20,999	0.03	31.60	0.000
Anger	-0.75	0.000	-1.08	-0.41	20,809	0.02	27.37	0.000
Disgust	0.51	0.000	0.26	0.76	20,809	0.04	44.04	0.000
Fear	-2.71	0.000	-3.13	-2.30	20,809	0.02	31.38	0.000
Joy	-0.41	0.001	-0.64	-0.17	20,809	0.00	3.79	0.000
Sadness	-0.68	0.000	-1.05	-0.32	20,809	0.02	22.08	0.000
Surprise	-0.87	0.000	-1.08	-0.67	20,809	0.02	20.55	0.000

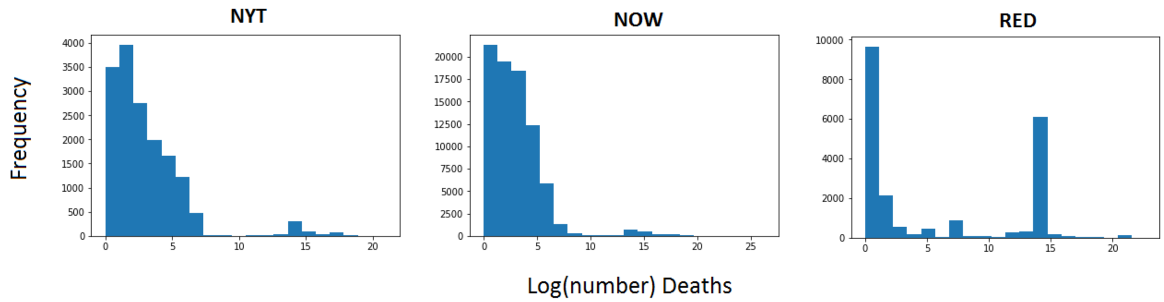


Figure 1. Distribution of log number of deaths across the three datasets analyzed in this paper.

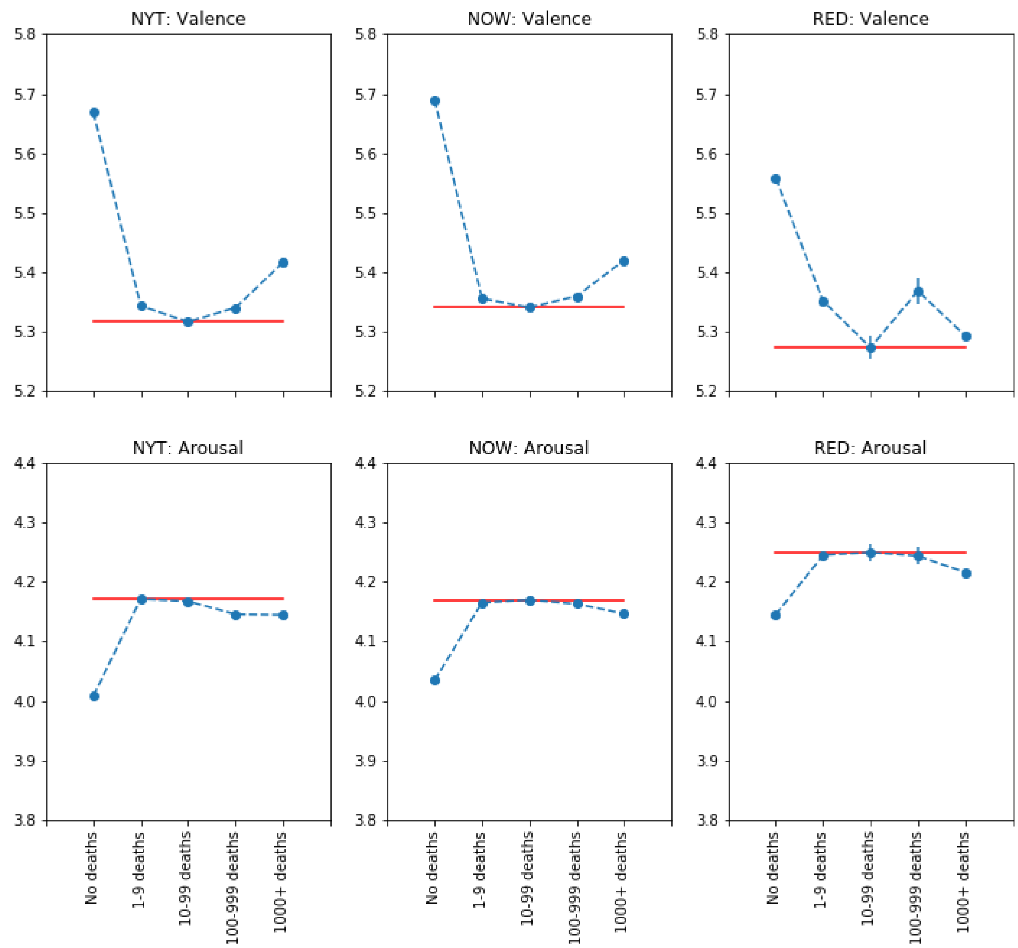


Figure 2. *Average valence and arousal in the texts in our three datasets, as a function of the number of deaths mentioned in the text. Red lines indicate minimum valence or maximum arousal in the dataset. Error bas indicate +/- 1 standard error. Note that due to the large sample sizes the error bars are not visible for most points.*

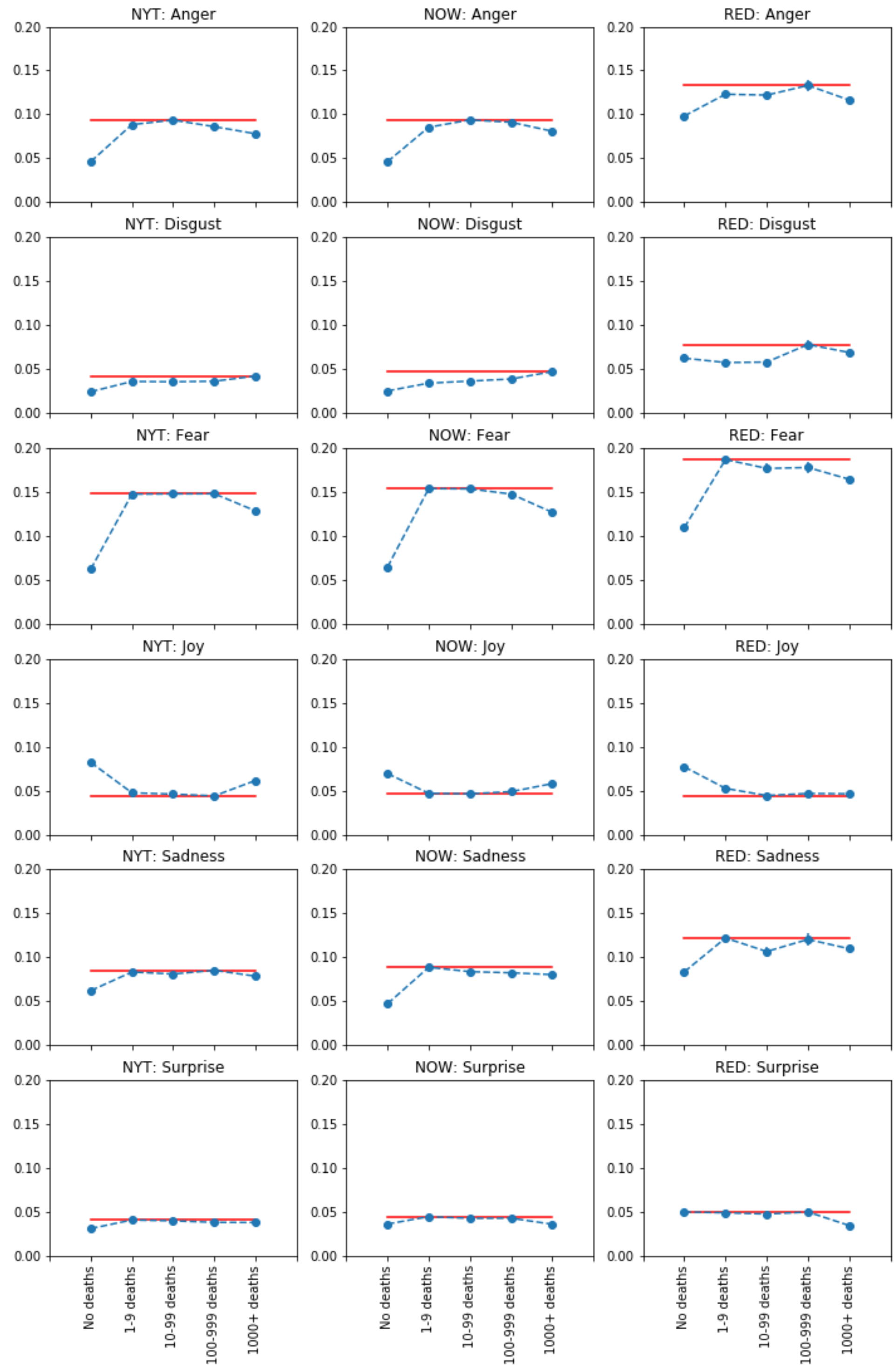


Figure 3. *Expressions of six basic emotions in texts, as a function of the number of deaths mentioned in the text. Error bas indicate +/- 1 standard error.*

NYT			
1-9 deaths	**10-99 deaths**	**100-999 deaths**	**1000+ deaths**
anger	*afraid*	*disgusted*	*disgust*
joy	*angry*		*surprised*
sad	*fear*		
sadness	*happy*		
	surprise		

NOW			
1-9 deaths	**10-99 deaths**	**100-999 deaths**	**1000+ deaths**
angry	*disgust*	*anger*	*afraid*
joy	*sadness*	*fear*	*disgusted*
sad			*happy*
			surprise
			surprised

RED			
1-9 deaths	**10-99 deaths**	**100-999 deaths**	**1000+ deaths**
anger	*afraid*	*disgust*	
joy	*angry*	*disgusted*	
sad	*sadness*	*fear*	
		happy	
		surprise	
		surprised	

Figure 4. *Death categories in which different emotion words are most likely to occur in our three datasets.*

NYT

	Valence	Arousal	Anger	Disgust	Fear	Joy	Sadness	Surprise
Accident	5.36	4.11	0.06	0.03	0.13	0.04	0.09	0.05
Crime	5.29	4.21	0.11	0.04	0.15	0.05	0.09	0.04
Cyclone	5.44	4.15	0.08	0.04	0.13	0.05	0.08	0.06
Disease	5.30	4.14	0.08	0.06	0.14	0.05	0.10	0.03
Drought	5.43	4.08	0.07	0.04	0.12	0.05	0.08	0.04
Earthquake	5.36	4.17	0.09	0.04	0.14	0.05	0.10	0.05
Famine	5.42	4.13	0.07	0.04	0.12	0.06	0.09	0.04
Flood	5.42	4.10	0.08	0.04	0.13	0.05	0.08	0.04
Genocide	5.31	4.17	0.10	0.05	0.15	0.05	0.09	0.03
Holocaust	5.50	4.15	0.09	0.05	0.12	0.07	0.09	0.04
Hurricane	5.42	4.15	0.09	0.04	0.12	0.04	0.08	0.04
Murder	5.28	4.22	0.11	0.04	0.16	0.05	0.09	0.04
Suicide	5.30	4.21	0.10	0.04	0.16	0.05	0.08	0.04
Terrorism	5.25	4.22	0.11	0.05	0.16	0.05	0.09	0.04
Tornado	5.43	4.13	0.09	0.04	0.14	0.04	0.09	0.04
War	5.32	4.19	0.10	0.04	0.15	0.05	0.08	0.04

NOW

	Valence	Arousal	Anger	Disgust	Fear	Joy	Sadness	Surprise
Accident	5.33	4.11	0.06	0.03	0.15	0.04	0.11	0.06
Crime	5.28	4.23	0.12	0.04	0.17	0.05	0.09	0.04
Cyclone	5.42	4.12	0.09	0.04	0.14	0.04	0.08	0.06
Disease	5.29	4.13	0.08	0.07	0.15	0.05	0.10	0.03
Drought	5.42	4.08	0.07	0.03	0.12	0.05	0.08	0.04
Earthquake	5.34	4.22	0.10	0.04	0.16	0.05	0.11	0.06
Famine	5.42	4.15	0.08	0.04	0.13	0.07	0.10	0.04
Flood	5.44	4.08	0.07	0.04	0.13	0.04	0.08	0.04
Genocide	5.31	4.19	0.11	0.05	0.16	0.06	0.09	0.04
Holocaust	5.45	4.18	0.10	0.06	0.14	0.06	0.10	0.04
Hurricane	5.47	4.14	0.09	0.04	0.13	0.04	0.07	0.04
Murder	5.29	4.23	0.12	0.05	0.18	0.05	0.09	0.04
Suicide	5.28	4.24	0.11	0.04	0.17	0.05	0.09	0.05
Terrorism	5.25	4.27	0.12	0.05	0.18	0.05	0.09	0.05
Tornado	5.45	4.18	0.09	0.04	0.14	0.04	0.08	0.04
War	5.33	4.20	0.10	0.04	0.16	0.05	0.08	0.04

RED

	Valence	Arousal	Anger	Disgust	Fear	Joy	Sadness	Surprise
Accident	5.20	4.25	0.10	0.05	0.22	0.03	0.16	0.10
Crime	5.12	4.33	0.18	0.07	0.19	0.05	0.11	0.04
Cyclone	5.14	4.39	0.13	0.07	0.27	0.05	0.08	0.15
Disease	5.10	4.24	0.13	0.11	0.20	0.04	0.16	0.03
Drought	5.25	4.19	0.10	0.07	0.14	0.06	0.10	0.04
Earthquake	5.22	4.41	0.15	0.06	0.19	0.03	0.15	0.09
Famine	5.16	4.27	0.09	0.06	0.14	0.05	0.16	0.03
Flood	5.31	4.30	0.09	0.04	0.17	0.06	0.08	0.04
Genocide	5.04	4.30	0.12	0.07	0.17	0.04	0.12	0.03
Holocaust	5.36	4.11	0.18	0.15	0.21	0.04	0.18	0.02
Hurricane	5.37	4.22	0.09	0.04	0.18	0.04	0.09	0.05
Murder	5.27	4.24	0.14	0.07	0.19	0.05	0.12	0.04
Suicide	5.27	4.31	0.13	0.05	0.20	0.05	0.13	0.06
Terrorism	5.06	4.40	0.17	0.11	0.23	0.04	0.15	0.04
Tornado	5.44	4.25	0.12	0.07	0.18	0.04	0.12	0.04
War	5.13	4.32	0.12	0.06	0.21	0.04	0.10	0.03

Figure 5: Average expressions of valence and arousal, and six basic emotions in texts, as a function of death context. Here blue (red) shading of the cells indicates higher (lower) values for

the emotion in the dataset. The darkness of the shading is based on the minimum and maximum values for the column corresponding to the emotion for the dataset.

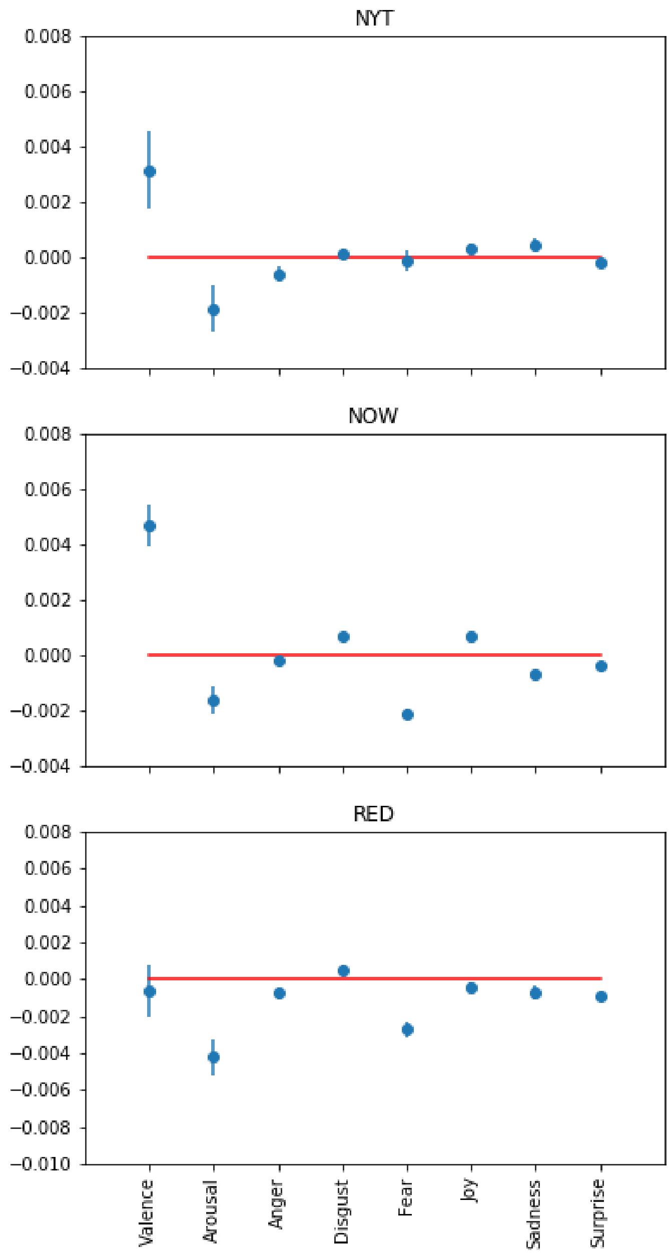
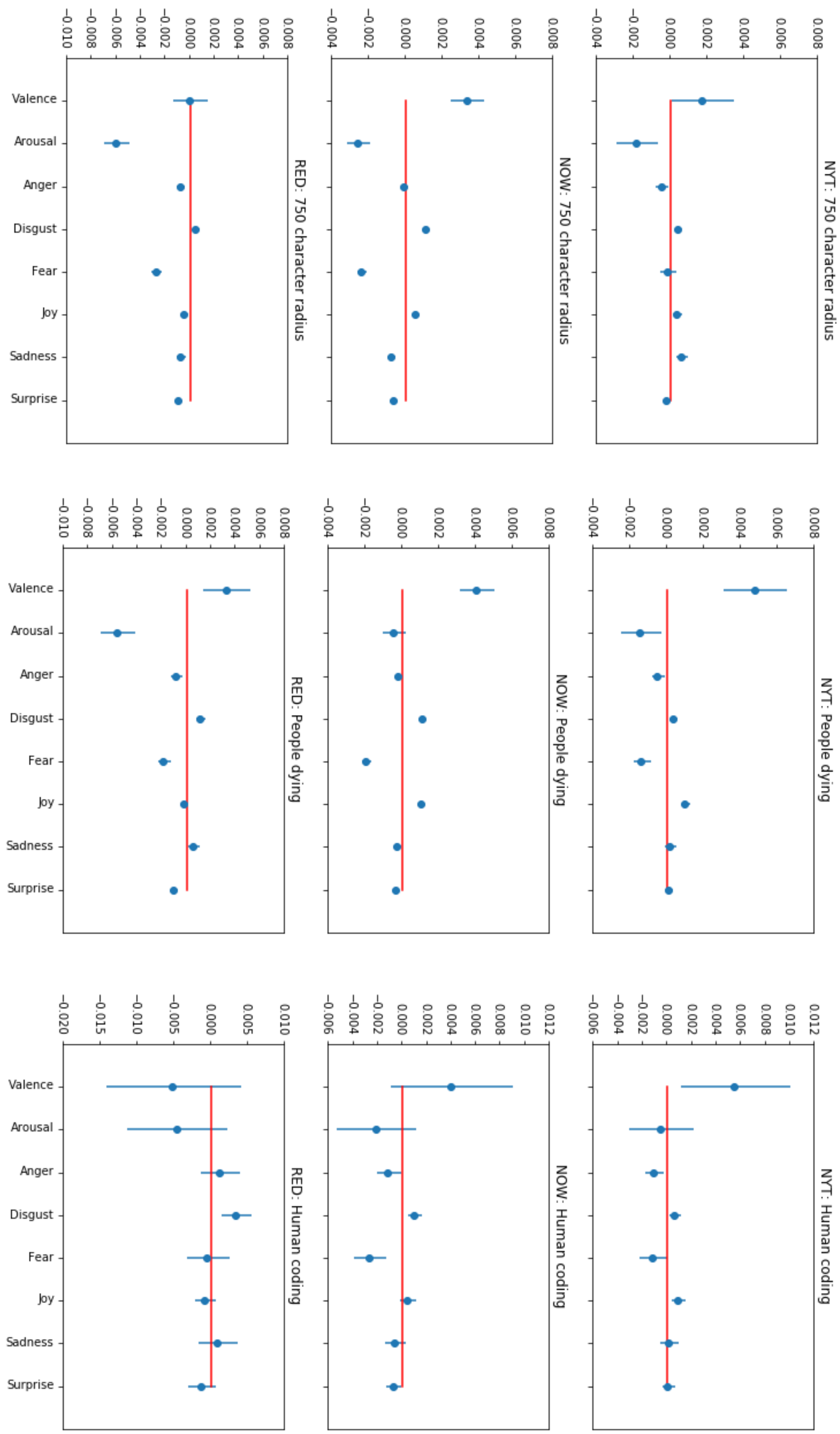


Figure 6: Coefficients for the effect of log-number of deaths on emotional expressions in texts. Error bars indicate 95% confidence intervals.



NYT: 750 character radius
NOW: 750 character radius
RED: 750 character radius
NYT: People dying
NOW: People dying
RED: People dying
NYT: Human coding
NOW: Human coding
RED: Human coding
Valence
Arousal
Anger
Disgust
Fear
Joy
Sadness
Surprise

Figure 7: Coefficients for the effect of log-number of deaths on emotional expressions in texts, for robustness tests involving analyses of only words in a 750-character window around the death phrase, analyses of only death phrases explicitly referring to people dying, and analyses of only texts verified as being correctly classified by human coders. Error bars indicate 95% confidence intervals.

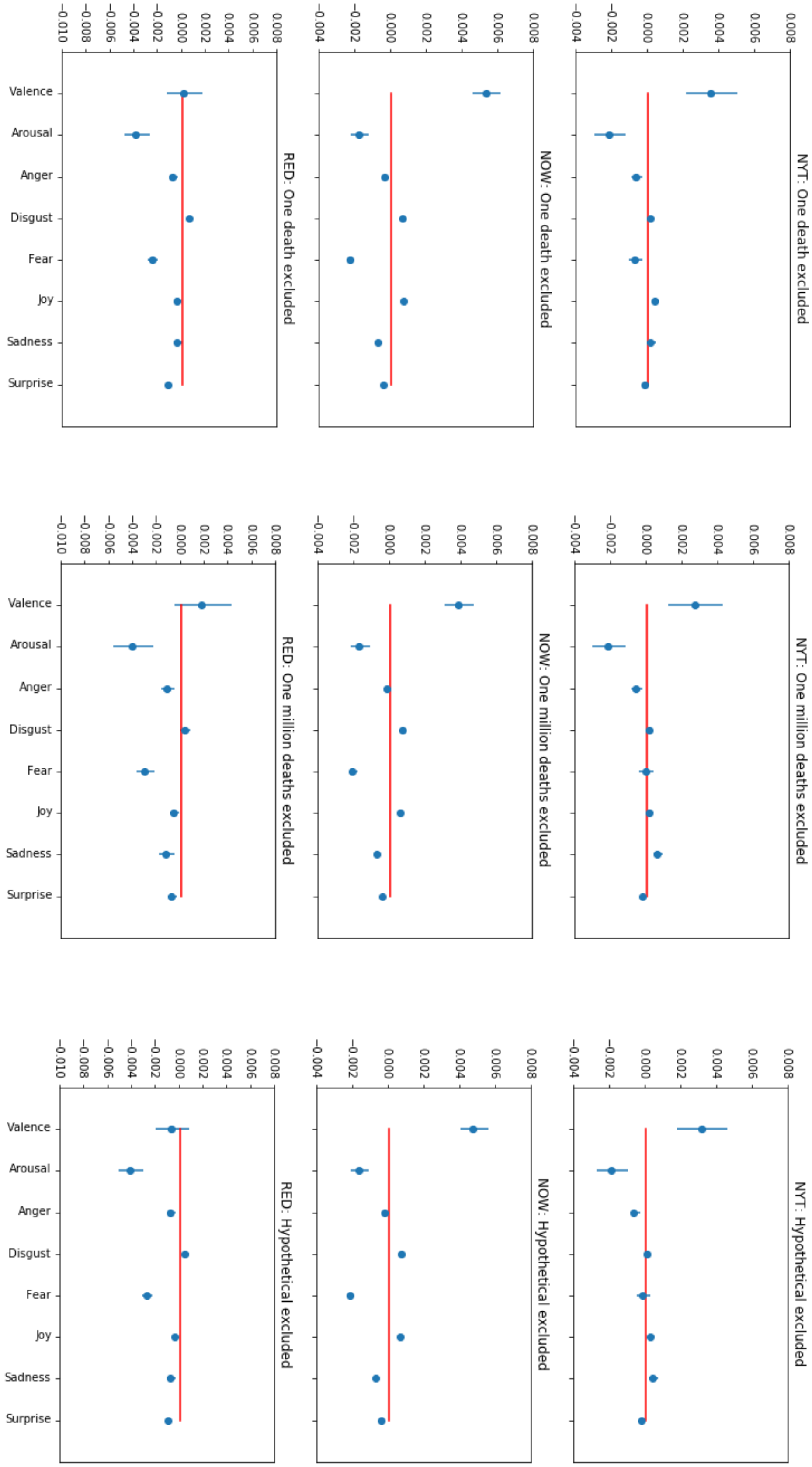

NYT: One death excluded
NOW: One death excluded
RED: One death excluded
NYT: One million deaths excluded
NOW: One million deaths excluded
RED: One million deaths excluded
NYT: Hypothetical excluded
NOW: Hypothetical excluded
RED: Hypothetical excluded
Valence
Arousal
Anger
Disgust
Fear
Joy
Sadness
Surprise

Figure 8: Coefficients for the effect of log-number of deaths on emotional expressions in texts, for robustness tests involving analyses of texts mentioning more than one death (left panels), analyses of texts not mentioning one million deaths (middle panels), and analyses of texts not mentioning the word "if" within a fifty character radius of the death phrase (right panels). Error bars indicate 95% confidence intervals.

Appendix

Death Phrases

We used the presence of 130 distinct death phrases to determine whether or not a text referred to people dying. These death phrases were developed based on experimenter intuition, and were intended to form a comprehensive linguistic set of death references. They were also preregistered.

The set of death phases used was {*dead, men dead, man dead, women dead, woman dead, children dead, child dead, people dead, person dead, died, men died, man died, women died, woman died, children died, child died, people died, person died, have died, has died, men have died, man has died, women have died, woman has died, children have died, child has died, people have died, person has died, perished, men perished, man perished, women perished, woman perished, children perished, child perished, people perished, person perished, have perished, has perished, men have perished, man has perished, women have perished, woman has perished, children have perished, child has perished, people have perished, person has perished, killed, men killed, man killed, women killed, woman killed, children killed, child killed, people killed, person killed, were killed, was killed, men were killed, man was killed, women were killed, woman was killed, children were killed, child was killed, people were killed, person was killed, have been killed, has been killed, men have been killed, man has been killed, women have been killed, woman has been killed, children have been killed, child has been killed, people have been killed, person has been killed, murdered, men murdered, man murdered, women murdered, woman murdered, children murdered, child murdered, people murdered, person murdered, were murdered, was murdered, men were murdered, man was murdered, women were murdered, woman was murdered, children were murdered, child was murdered, people were murdered,* *person was murdered, have been murdered, has been murdered, men have been murdered, man has been murdered, women have been murdered, woman has been murdered, children have been murdered, child has been murdered, people have been murdered, person has been murdered, committed suicide, men committed suicide, man committed suicide, women committed suicide, woman committed suicide, children committed suicide, child committed suicide, people committed suicide, person committed suicide, has committed suicide, men have committed suicide, man has committed suicide, women have committed suicide, woman has committed suicide, children have committed suicide, child has committed suicide, people have committed suicide, person has committed suicide, death, murder, suicide, fatalities, fatality, lives were lost, life was lost*}.

Deviations from Preregistration Plan

Our analysis included a number of deviations from our preregistration plan, motivated by the difficulties of accurately extracting death phrases from the corpora, which we could not anticipate prior to collecting our data. Specifically, the flagging algorithm specified in our preregistration excluded year numbers from the analysis only if the text used the sentence structure in YEAR PHRASE (e.g. in 1987 women were killed), the YEAR PHRASE (e.g. the 1987 murders), or a YEAR PHRASE (e.g. a 1987 suicide). However, human coding revealed that there were many additional YEAR PHRASE statements that were not preceded by in, the, or a, and that all instances in which there was a death phrase preceded by a year between 1800 and 2019 involved references to years rather than instances in which thousands of people died. Our decision to exclude all such statements ensures that there are no such errors in our classification.

The flagging algorithm in our preregistration also did not exclude texts in which the word preceding the number was a month, month abbreviation, or an age word. Again, it was human coding that revealed that the numbers mentioned in all such instances refer to dates or ages rather than the number of deaths in the event.

The flagging algorithm only applied to *million* and not to *billion*, *thousand*, or *hundred.* Additionally, it did not explicitly consider verbal representations such as *half a*, *quarter of a*, *tens of*, or *hundreds of*. Human coding revealed that such instances were present and needed to be explicitly corrected.

A fourth deviation from our preregistration involves our use of five (rather than seven) categories for combining texts. Particularly, our preregistration specified that we would divide our texts into the following seven categories, in order to avoid data sparsity issues: 1. Texts that do not mention any deaths; 2. Texts that mention between 1-9 deaths; 3. Texts that mention between 10-99 deaths; 4. Texts that mention between 100-999 deaths; 5. Texts that mention between 1,000-9,999 deaths; 6. Texts that mention between 10,000-99,999 deaths; and 7. Texts that mention 100,000+ deaths. Our primary analysis would then calculate the average emotional content of texts in each category and test for statistically significant differences across these categories. After performing our categorization, we realized that there were very few texts in category 5 and 6. In fact category 6 (10,000 to 99,999 deaths) only had a total of 234 total texts across our three datasets (compared to, for example, 10,000+ total texts in category 7 and 15,000+ texts in category 4). Thus we pooled categories 5, 6 and 7 into a single category of texts involving 1,000+ deaths. Of course, we also perform additional analysis, not part of our preregistration plan that does not involve discretizing our results into these categories. This analysis shows that our results are robust to how we discretize the number of deaths our data.

Our preregistration plan specified that we would only examine the word occurrences for *murder*, *accident*, and *suicide*, however, as specified above, we expand our analysis to consider an additional 16 possible death contexts. In addition to analyzing valence and arousal using Warriner et al.'s dataset, (2013) we also analyzed six basic emotions using the NRC word-emotion lexicon (Mohammed & Turney, 2010). Our use of the NRC word-emotion lexicon was not preregistered, but is useful for a more nuanced understanding of the emotional content of the texts.

There are a number of additional extensions (rather than modifications) to our preregistration plan, such as our extensive robustness checks, which were not preregistered. We do not discuss these extensions here (though we do identify them as non-preregistered extensions in the main text).

Examples of Mentions of Death

In Table A1, we present mentions of deaths from our three datasets. We present one mention for each of four death categories (1-9 deaths, 10-99 deaths, 100-999 deaths, and 1,000+ deaths) per dataset. These mentions of death were randomly selected from the set of texts given to our human coders, and we also present results of the coding analysis (indicating whether or not our initial death classification was correct). In the examples shown here, one of our twelve texts were classified incorrectly by the initial algorithm. This was the basis of modifications to our algorithm, such as the exclusion of death phrases in which the number of deaths was equivalent to year (e.g. 1974).

Plots of Aggregate Emotionality

In Figure A1 we provide an additional plot showing the aggregate valence, arousal, anger, disgust, fear, joy, sadness and surprise for texts with at least one death mention. In these plots we also shade the color of the cell to capture the relative expression of that emotion. Cells reflecting more positive emotions (e.g. higher valence, or lower fear) are shaded green, whereas cells reflecting more negative emotions (e.g. lower valence, or higher fear) are shaded red.

Nonlinear Regressions

As many of the trends identified in Figures 2 and 3 are non-linear, we reran our primary linear regression with a non-linear component. Specifically, this regression predicted the emotionality of the text using both the log-number of deaths and the log-number of deaths squared. As with our regressions in the main text, this regression also had fixed effects for each of the sixteen death contexts, as well as random effects for the 130 different death phrases. The coefficients on the linear and quadratic terms revealed in this regression are shown in Table A2. To allow readers to easily interpret the quadratic component of the model (which cannot be easily done by just visually inspecting the coefficients) this table also provides model predictions for N = {10,100,1000…10000000} deaths. Here we can see that the numbing patterns observed in our prior analysis replicate. Thus, predicted valence is higher for higher deaths in both the NYT and NOW data (these data seem to display a convex relationship between valence and number of deaths). Interestingly, we also find a U shaped relationship for the RED data, which was not apparent in the linear regression. Thus, valence initially drops, but then increases, as number of deaths increases. This is consistent with the numbing hypothesis.

Table A1. Examples of mentions of deaths, randomly selected from the list of texts offered to our human coders.

	# Deaths	**Correct**
NYT		
was also a student. One instructor got out but two were killed, Air Force officials said	2	Y
and 15 wounded. Moscow has confirmed 11 dead on its side, while a spokesman for the insur	11	Y
nd 252 people killed, last year 11 crashes and 221 deaths. But perhaps none of them were a	221	Y
B. Henderson. (Norton, $22.95.) An enigmatic 1974 murder case	1,974	N
NOW		
ut so far authorities have confirmed at least four deaths. In the province of Surat Thani, one bom	4	Y
ther militant attacks since June, 2014. Twenty people were killed this month in an attack on a s	20	Y
ussein said in a statement. His office said 331 deaths were recorded between September 6	331	Y
ings and vicious assaults have left more than 6300 people dead and at least 11,500 injured in violen	6,300	Y
RED		
uring implementation of this programs at least one person was murdered that we know of and	1	Y
ases, including a lady with a rescue who had ten dead on her property and a dozen more in critical	10	Y
nfected fleas from airplanes, where over a hundred people died before quarantined the area	100	Y
brought small pox to the Americas. Some million dead natives later and we have our winner	1,000,000	Y

Table A2. Outputs of a linear regression with a quadratic component on the log-number of deaths. The first two columns indicate coefficients of this regression for the linear (LogN) and quadratic (LogN^2) components, with * corresponding to $p < 0.05$. The remaining columns indicate predictions of the model for various numbers of deaths.

			Emotion Prediction for N Deaths						
NYT	**LogN**	**LogN^2**	**10**	**100**	**1000**	**10000**	**100000**	**1000000**	**10000000**
Valence	-2.32	0.40*	5.39	5.39	5.39	5.40	5.42	5.43	5.46
Arousal	-5.69*	0.28*	4.15	4.14	4.13	4.13	4.13	4.13	4.14
Anger	-0.2	-0.03	0.08	0.08	0.08	0.08	0.08	0.07	0.07
Disgust	-0.13	0.02	0.03	0.03	0.03	0.03	0.03	0.04	0.04
Fear	1.59*	-0.13*	0.14	0.14	0.14	0.14	0.14	0.14	0.13
Joy	-1.16*	0.11*	0.05	0.05	0.05	0.05	0.05	0.06	0.06
Sadness	1.63*	-0.09*	0.08	0.09	0.09	0.09	0.09	0.09	0.08
Surprise	-0.19*	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NOW	**LogN**	**LogN^2**	**10**	**100**	**1000**	**10000**	**100000**	**1000000**	**10000000**
Valence	4.13*	0.04	5.42	5.43	5.44	5.45	5.47	5.48	5.49
Arousal	-5.48*	0.30*	4.14	4.13	4.13	4.13	4.13	4.13	4.14
Anger	1.07*	-0.10*	0.08	0.08	0.08	0.08	0.08	0.08	0.07
Disgust	1.26*	-0.04*	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Fear	-3.04*	0.07*	0.14	0.13	0.13	0.12	0.12	0.12	0.12
Joy	0.35*	0.03*	0.05	0.05	0.06	0.06	0.06	0.06	0.06
Sadness	-1.59*	0.07*	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Surprise	-0.88*	0.04*	0.04	0.04	0.04	0.04	0.04	0.04	0.04
RED	**LogN**	**LogN^2**	**10**	**100**	**1000**	**10000**	**100000**	**1000000**	**10000000**
Valence	-7.97*	0.49*	5.40	5.39	5.38	5.38	5.38	5.39	5.41
Arousal	-7.77*	0.24	4.21	4.19	4.18	4.17	4.16	4.16	4.16
Anger	0.97	-0.11*	0.12	0.12	0.12	0.11	0.11	0.11	0.10
Disgust	1.17*	-0.04	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Fear	-5.08*	0.16*	0.16	0.15	0.15	0.14	0.13	0.13	0.13
Joy	-0.77	0.02	0.06	0.05	0.05	0.05	0.05	0.05	0.05
Sadness	-1.66*	0.06	0.11	0.10	0.10	0.10	0.10	0.10	0.10
Surprise	-0.31	-0.04	0.05	0.04	0.04	0.04	0.04	0.04	0.03

NYT

# Deaths:	1-9	10-99	100-999	1000+
Valence	5.34	5.32	5.34	5.42
Arousal	4.17	4.17	4.15	4.14
Anger	0.09	0.09	0.09	0.08
Disgust	0.04	0.04	0.04	0.04
Fear	0.15	0.15	0.15	0.13
Joy	0.05	0.05	0.04	0.06
Sadness	0.08	0.08	0.09	0.08
Surprise	0.04	0.04	0.04	0.04

NOW

# Deaths:	1-9	10-99	100-999	1000+
Valence	5.36	5.34	5.36	5.42
Arousal	4.17	4.17	4.16	4.15
Anger	0.09	0.09	0.09	0.08
Disgust	0.03	0.04	0.04	0.05
Fear	0.15	0.15	0.15	0.13
Joy	0.05	0.05	0.05	0.06
Sadness	0.09	0.08	0.08	0.08
Surprise	0.04	0.04	0.04	0.04

RED

# Deaths:	1-9	10-99	100-999	1000+
Valence	5.35	5.27	5.37	5.29
Arousal	4.24	4.25	4.24	4.22
Anger	0.12	0.12	0.13	0.12
Disgust	0.06	0.06	0.08	0.07
Fear	0.19	0.18	0.18	0.16
Joy	0.05	0.04	0.05	0.05
Sadness	0.12	0.11	0.12	0.11
Surprise	0.05	0.05	0.05	0.03

Figure A1. Aggregate emotionality of texts in various death categories. Cells are shaded based on the emotionality of the death category for the emotion relative to other death categories for the same emotion. Cells reflecting more positive (negative) emotions are shaded green (red).