

SUPPLEMENTARY ONLINE MATERIALS FOR:

Sadness shifts to anxiety with time and distance from elementary school shooting in Newtown CT

Bruce Doré, Leonard Ort, Ofir Braverman, and Kevin Ochsner
Department of Psychology, Columbia University

STUDY 1: Using Twitter to track responses to the Newtown shooting over time and space

i. Tweet collection Procedure

Data collection scripts were written in Python (www.python.org), using the Tweepy API library (tweepy.org) to query the Twitter Streaming API (stream.twitter.com/1.1/statuses/filter.json) with the track parameter set to the keywords of interest (track = ['newtown', 'sandy hook']).

During data collection days, operation of the collection script was manually checked and it was restarted if it had errored out. After the initial collection period in December, active collection was resumed in January, and again in April, in order to investigate long-term changes in word use. (No data was excluded based on the time it was collected.)

ii. Tweet exclusions

Tweets meeting the following criteria were removed from the dataset, in an attempt to focus our analyses on original, English language content from unique users.

before geocoding (i.e., removed from raw .json database of all collected tweets)

- retweeted content (390840 of 1137216 tweets)

after geocoding & geocoding cleaning (i.e., removed from dataset of tweets with reported US location info)

- multiple tweets from individual users (31038 tweets)
 - median time between tweets = 8.56 hrs; *mode = 0.00 hrs]
 - we kept the first tweet posted by each user, to avoid giving disproportionate weight to any given user while maintaining precise measurement of time
 - many tweets removed at this step were duplicates, seemingly reflecting an error in processing within the Twitter API
- tweets with less than 60% English words (8939 tweets)
- tweets from different users with identical content (1177 tweets)

iii. Descriptive statistics for individual vs. newsmedia accounts

Four raters coded a random sample of 1000 tweets as having been posted by an individual, a news media organization, or other/unclear. Descriptive statistics for these tweets are reported below.

		total	affect	anx	sad	anger	cause	distance (miles)	time (months)
individual	mean	14.67	0.98	0.05	0.16	0.24	0.30	869.02	1.36
	(n=822) std dev	(5.82)	(1.02)	(0.21)	(0.42)	(0.49)	(0.57)	(818.35)	(1.32)
newsmedia	mean	14.8	1.07	0.04	0.33	0.29	0.15	855.74	1.11
	(n=52) std dev	(5.69)	(1.02)	(0.23)	(0.5)	(0.54)	(0.39)	(829.44)	(1.21)
unclear/other*	mean	15.31	0.90	0.00	0.18	0.10	0.31	1114.62	1.41
	(n=126) stdev	(5.33)	(0.91)	(0.00)	(0.45)	(0.31)	(0.47)	(994.74)	(1.43)

Table S1. Descriptive statistics for a random sample of 1000 tweets referencing “Newtown” or “Sandy Hook”, classified as having been posted by individuals, news media accounts, or unclear/other. *Note: in a forced choice rating, unclear/other tweets were classified as: individual person (31 tweets), person representing organization (95 tweets), non-human entity (0 tweets).

iv. Geocoding data cleaning procedure

Geocoded tweets were sorted by location field and manually checked to ensure that tweets indicating identical location were assigned coordinates that were identical (and correct according to Google Maps).

Next, tweets meeting any of these criteria were removed:

- geocoded coordinates outside of the continental US (121 tweets)
- more than one location specified in location field (e.g., "Newtown CT & Bington VT"; 481 tweets)
- location specified ambiguously in location field (e.g., "heart and mind in CT"; 68 tweets)

v. Base rates of word use relative to comparison corpora

	%AFFECT	%SADNESS	%ANXIETY	%ANGER
	words	words	words	words
<i>Pennebaker et al. 2007 corpora</i>				
emotional writing	6.02%	0.63%	0.68%	0.66%
control writing	2.57%	0.14%	0.21%	0.13%
blogs	5.84%	0.42%	0.30%	0.76%
<i>Newtown Twitter data</i>				
wave 1	7.27%	1.69%	0.27%	1.72%
wave 2	6.16%	0.58%	0.32%	1.86%
wave 3	5.61%	0.62%	0.62%	1.67%

Table S2. Base-rate use of words from LIWC Affect, Sadness, Anxiety, and Anger categories (as percentage of total words) in corpora described in Pennebaker, Chung, Ireland, Gonzales, and Booth's (2007) LIWC Language manual, and in tweets referencing "Newtown" or "Sandy Hook".

vi. Model-predicted change in word use across time and distance

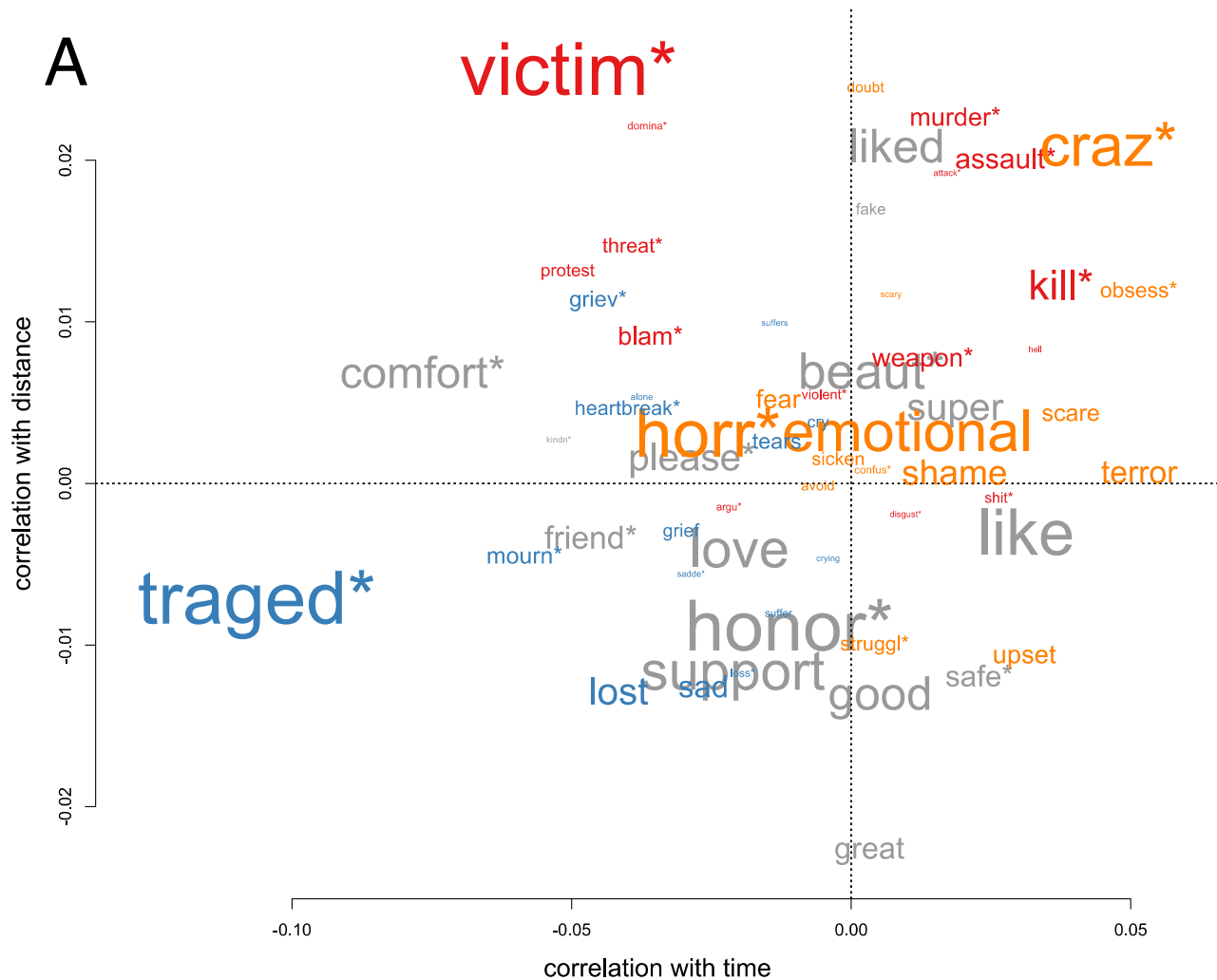
	beta (time)	model-predicted change in 1 month	beta (distance)	model-predicted change with 1000 miles
AFFECT	-0.073	-7%	-0.033	-3%
SADNESS	-0.377	-31%	-0.057	-5%
ANXIETY	0.116	+12%	0.105	+11%
ANGER	<i>ns</i>	<i>ns</i>	0.100	+11%

Table S3. Negative binomial model beta coefficients and predicted change across one-unit increases in time (1 month) and distance (1000 miles) for LIWC Affect, Sadness, Anxiety, and Anger categories. (% change = $1 - e^{\text{beta}}$)

vii. Reanalysis of effects of time with equal-length sampling waves

Because the waves of data collection in Study 1 were of irregular intervals, we went back and limited the dataset to tweets collected within the first 6 days of each of the three waves of data collection (n=27330), and ran the same models reported in the article. The results of these models correspond with those in the article. Time is associated with decreased affect words, $b = -.041$, 95%CI=[-.053, -.029], $p < .001$, decreased sadness words, $b = -.47$, 95%CI=[-.51, -.43], $p < .001$, increased anxiety words, $b = .15$, 95%CI=[.10, .19], $p < .001$, and increased causality words, $b = .043$, 95%CI=[.021, .065], $p < .001$.

viii. Visualizing effects of time and space at the level of individual words



B example tweets (paraphrased)

SADNESS: All of us are affected by the *tragedy* at Sandy Hook Elementary

ANXIETY: The shooting at Sandy Hook is senseless and *horrifying*

ANGER: A memorial will be built for the Sandy Hook *victims*

OTHER AFFECT: We *honor* and remember those who died at Newtown.

Figure S1. A) Words from LIWC Sadness (in blue), Anxiety (in orange), Anger (in red), and Other Affect (in grey) categories most frequently used in tweets referencing “Newtown” or “Sandy Hook”. Text size is proportional to (within-category) frequency of use, spatial arrangement reflects correlations with time and distance (* indicates wildcard) B) Paraphrased example tweets containing the most common word from each category.

Study 2: Focusing on the abstract causes of this tragedy evokes a pattern of sadness and anxiety that mirrors effects of time and distance observed on Twitter

i. Exploring the relationship between causal uncertainty and anxiety

In order to investigate the possibility that anxiety is related to perceptions of uncertainty regarding the causes of the Sandy Hook tragedy, we conducted an exploratory analysis of the data collected in Study 2.

Two raters rated the texts generated in Study 2 on a 5-pt scale (1-very certain of cause to 5-very uncertain of cause), and we conducted a regression analysis with this variable as a predictor and self-reported anxiety as an outcome. The inter-rater reliability for this measure was acceptable (Cronbach's $\alpha = .79$). The mean rating of uncertainty was 2.57 (SD=1.2). The bivariate relationship between uncertainty of cause and anxiety was negative and non-significant, $b = -2.7$, 95% CI = [-9.6, 4.2], $p = .41$. Entering uncertainty and perceived risk/unresolvedness as simultaneous predictors did not change the significance of perceived risk/unresolvedness as a predictor of anxiety. Additionally, one rater categorized the causes identified in these texts. The most common cause identified was shooter mental health (66%), followed by societal mental health policy (28%), societal gun policy (24%), parenting/home environment (12%), school security (8%), and terrorism motive (4%).