How do we feel when a robot dies? Emotions expressed on Twitter before and after hitchBOT's destruction

Kathleen C. Fraser
National Research Council Canada
Ottawa, Canada
kathleen.fraser@nrc-cnrc.gc.ca

Frauke Zeller
Ryerson University
Toronto, Canada
fzeller@ryerson.ca

David Harris Smith
McMaster University
Hamilton, Canada
dhsmith@mcmaster.ca

Saif M. Mohammad
National Research Council Canada
Ottawa, Canada
saif.mohammad@nrc-cnrc.gc.ca

Frank Rudzicz
University of Toronto and Vector Institute
Toronto, Canada
frank@spoclab.com

Abstract

In 2014, a chatty but immobile robot called hitchBOT set out to hitchhike across Canada. It similarly made its way across Germany and the Netherlands, and had begun a trip across the USA when it was destroyed by vandals. In this work, we analyze the emotions and sentiments associated with words in tweets posted before and after hitchBOT's destruction to answer two questions: Were there any differences in the emotions expressed across the different countries visited by hitchBOT? And how did the public react to the demise of hitchBOT? Our analyses indicate that while there were few cross-cultural differences in sentiment towards hitchBOT, there was a significant negative emotional reaction to its destruction, suggesting that people had formed an emotional connection with hitchBOT and perceived its destruction as morally wrong. We discuss potential implications of anthropomorphism and emotional attachment to robots from the perspective of robot ethics.

1 Introduction

A small group of Canadian researchers created the hitchBOT project in 2014, intersecting art, social robotics, and social science (Zeller and Smith, 2014; Smith and Zeller, 2017b). Its purpose was to kindle the public's engagement in arts and science, as well as spark discussions about our societies' attitudes towards robotics and technology. To this end, hitchBOT, shown in Figure 1, was designed to hitchhike alone across Canada (from Halifax to Victoria), relying on the kindness of strangers since it could not move on its own.

Figure 1: The hitchBOT robot.

The physical form of hitchBOT consisted of 'pool noodle' flotation devices for arms and legs, rubber boots, a plastic bin wrapped in solar panels for a body, and LED screens with facial animation on its head. GPS and 3G wireless allowed hitchBOT to communicate location and other diagnostics to the home server, and enabled speech recognition and automated dialogue via Cleverscript servers (Existor, 2016). Roughly the size of a five-year-old child, hitchBOT was designed to appear playful and non-threatening (Smith and Zeller, 2017a).

To a large extent, hitchBOT was successful. It traversed Canada, over 10,000 kilometres in 26 days, with no damage or adverse events, and gained broad international interest. With more than 35,000 followers on Twitter, 48,000 Likes on Facebook, and 12,000 followers on Instagram, hitchBOT incited a substantial level of engagement on social media. Moreover, hitchBOT at-

Proceedings of the 10th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 62–71
Minneapolis, June 6, 2019.

tracted significant international media interest, encompassing all traditional media forms (TV, radio, print media).

In 2015, a twin hitchBOT traveled Germany, the Netherlands, and the USA. The latter journey began on 17 July in Marblehead, MA, but abruptly ended in wilful destruction on 1 August, in Philadelphia, PA, only 500 km away.

In this paper, we examine the emotional connotations of the words used in the Twitter discourse around hitchBOT, using existing crowdsourced lexicons for emotion and sentiment. Others have started to investigate the emotional connections people build through personal interactions with robots (Young et al., 2009; Hirth et al., 2011; Hwang et al., 2013; Damiano et al., 2015). However, hitchBOT was exceptional in that the vast majority of its many Twitter followers would never meet it. In this sense, hitchBOT was similar to a public figure, or celebrity, and its destruction was a news-worthy event. As such, this represents a unique opportunity to measure widespread public opinion about robots and their treatment at the hands of humans, without the complicating factor of personal ownership. Darling (2016) argues that the degree of emotional connection we feel towards non-human entities, and specifically the emotional distress we feel when they are abused, is a major factor in whether we agree as a society to grant those entities legal protections beyond the simple property rights of the owner. Therefore it stands to reason that a better understanding of public sentiment could help to inform the debate over potential policies and regulations relating to robots and their use (e.g., Lin et al. (2011)).

We specifically explore two questions here:

(1) Were there differences in the type or scale of emotions expressed in each of the host countries? We compare the percentages of words associated with different emotions from the tweets produced during each trip, to examine any cultural factors in the public reaction to hitchBOT.

(2) What emotions were triggered when hitchBOT was destroyed? We compare the percentages of different emotion words and the distribution of positive and negative words produced before and after hitchBOT's destruction, to determine the dominant emotional responses to the event.

We begin with an overview of the related work studying human emotions towards robots, and then describe the corpus of tweets and the word–emotion association lexicons used in this work. We then present our findings, and conclude by discussing some examples from the data in relation to issues of anthropomorphism, emotion, and the question of how the ethical codes that govern our behaviour toward humans and animals may (or may not) apply to robots.

2 Background and related work

As robots become more common in our everyday lives, there is a growing need to understand the factors influencing interactions between humans and robots, including the emotional component. One active area of research focuses on developing robots that can express emotion (Kühnlenz et al., 2013); here, in contrast, we consider the emotions expressed by humans towards robots. How do robots make us feel? Many robots are designed to promote anthropomorphism and zoomorphism (the attribution of human or animal characteristics to a non-human/animal entity), and it has been shown that the degree to which we anthropomorphize a robot affects our emotional connection with it (Riek et al., 2009). However, even robots with little physical resemblance to a human or animal can induce emotional attachments (Sung et al., 2007).

Our sentiments towards robots may depend partly on cultural differences. Bartneck et al. (2007a) administered a questionnaire on negative attitudes towards robots to 467 participants from seven different countries, including Germany, the Netherlands, and the USA. The questionnaire was divided into three subscales focusing on interaction, social influence, and emotion. In general, participants from the USA showed the most positive attitudes towards robots, particularly in their openness to interacting with robots, although they were more negative than the German or Dutch on the topic of robot emotion.

Social media has proven to be a rich source of data for sentiment and emotion analysis on a variety of topics, using lexicon-based and machine learning methods (e.g. Rosenthal et al. (2015); Giachanou and Crestani (2016); Mohammad et al. (2018)). However, very little work has focused on the emotions expressed towards robots. Friedman et al. (2003) analyzed 3,119 forum posts relating to the AIBO robot dog. They developed

a coding scheme to categorize posts as affirming or negating the following characteristics in AIBO robots: life-like essences, technological essences, mental states, social rapport, and moral standing. Interestingly, while most users affirmed aspects of life-like essences, mental states, and social rapport, only 12% expressed that the AIBO dogs have moral standing (e.g. a right not to be mistreated). Mubin et al. (2016) annotated 235 Twitter posts relating to the Nao robot, using a similar coding scheme, finding that over half the tweets expressed life-like essences and/or social rapport. Fink et al. (2012) compared forum posts about AIBO dogs, Roomba robot vacuum cleaners, and iPad tablets for topic and degree of anthropomorphism. They characterized anthropomorphic language as an attribution to the device of: life-likeness, emotional states or feelings, gender, personality, intention, names, or status as a family member. They found a generally higher frequency of anthropomorphic language in posts which also expressed a feeling or attitude towards the device, again supporting a link between anthropomorphism and emotion.

Other work on social media has focused specifically on users' interactions with chatbots, such as the infamous Tay chatbot. Tay was launched by Microsoft in 2016 and promptly shut down a day later, after her interactions with Twitter users resulted in her learning to generate toxic and offensive content. Neff and Nagy (2016) analyzed user responses to the incident and found that most reactions fell into two categories: those who saw Tay as a helpless victim of human behaviour, and those who viewed her as a threat or an example of technology spinning out of control. More generally, we expect that there will be individual differences in the degree to which artificial intelligence technologies are seen as useful and progressive versus threatening and dangerous, and this may be reflected in the emotional responses observed.

The questions of how we feel when a robot is harmed are open for debate. Friedman et al. (2003) describe the outrage and disgust expressed by some online forum users when an AIBO robot dog was thrown into the garbage on live TV; some Twitter users also expressed discomfort or sadness in response to a video of a Boston Dynamics employee kicking a robot dog (Parke, 2015).

The 'death' of a robot can be even more emotional. In Japan, when robot dogs break down permanently, they are sometimes honoured with Buddhist funerals (Burch, 2018). Other work has explored the attachments that soldiers develop with military robots, and the sense of loss that can follow their destruction in battle (Carpenter, 2016). Even the break-down of a Roomba can elicit "surprising" levels of emotional distress (Sung et al., 2007).

The prospect of 'killing' a robot can also be disturbing to many people. Bartneck et al. (2007b) report a study in which participants first interacted with a robot, and were then asked to destroy it with a hammer. Participants described feelings of guilt and uneasiness (although, notably, it appears that they all complied). Darling et al. (2015) reported that, when faced with a similar task, participants hesitated longer when the robot had been given a name and personified back-story.

To summarize the related work as it applies to our questions: we expect to see evidence for different attitudes towards hitchBOT across countries, with the USA expressing more positive sentiment and openness towards the robot (Bartneck et al., 2007a). After hitchBOT's destruction, we expect to see an increase in negative emotion, including sadness at the loss of hitchBOT and anger and disgust towards the perpetrator(s). However, people who feel distrustful of technology or artificial intelligence may express opinions supporting hitchBOT's destruction.

3 Methodology

In this section we first present the Twitter data collected for the analysis, then discuss our methodology for emotion analysis through the use of two large, publicly-available lexicons for sentiment and emotion.

3.1 Twitter data

The raw dataset comprises 188,082 tweets containing the token *@hitchBOT*, with the first tweet posted on 29 May, 2014, two months before hitchBOT's first trip, and the last tweet posted on 16 November, 2015, 3.5 months after its destruction.

We first remove all retweets with no additional content (73,050 tweets), and all exact duplicates (30,334 tweets). We also remove all tweets from hitchBOT's own Twitter account[1] (494 tweets). We determine the language of a tweet using the Python `langdetect` library.[2] The vast majority

[1] Tweets from this account were written by a human.

[2] https://pypi.org/project/langdetect/

of tweets are written in English; to better capture the emotions in the countries through which hitchBOT travelled, we also include all tweets written in French (one of Canada's official languages), German, and Dutch. We exclude those written in any other languages (20,132 tweets). We then preprocess the tweets by replacing links, usernames, and RT tokens with ⟨*URL*⟩, ⟨*@USERNAME*⟩, and ⟨*RT*⟩, respectively. After this step, any tweets containing only links, usernames, and retweet tokens are also removed (435 tweets). As a result, we include 63,632 tweets in the final dataset.

3.2 Emotion analysis

There are different theories regarding the categorization and definition of emotions. In one view, there is a finite set of universal emotions. In pioneering work, Ekman et al. (1969) proposed a set of six culturally-universal emotions (joy, sadness, disgust, fear, anger, and surprise); Plutchik (1984) later developed a set of eight basic emotions (adding trust and anticipation).

An alternative theory seeks to describe emotions in terms of their underlying factors, or dimensions. Russell (2003) argues in favour of three largely independent dimensions, namely: valence (positive versus negative), arousal (active versus passive), and dominance (powerful versus weak).

In this work, we conduct our analysis from both the categorical and dimensional perspectives by using two lexicons: the NRC Emotion Lexicon (Mohammad and Turney, 2013), and the NRC Valence-Arousal-Dominance (VAD) Lexicon (Mohammad, 2018). Both lexicons were collected by crowd-sourcing annotations of emotional associations with words, and are publicly available.[3] The NRC lexicons offer wider coverage than most existing lexicons, and the use of best-worst scaling in the VAD Lexicon has been shown to lead to more reliable annotations than those obtained using rating scales (Mohammad, 2018). The NRC lexicons have been extensively validated for Twitter emotion and sentiment analysis (Tang et al., 2014; Yu and Wang, 2015; Chikersal et al., 2015).

Briefly, the Emotion Lexicon contains emotion labels for 14,182 unigrams. The labels are binary, indicating whether a word is associated with (a) any of Plutchik's eight basic emotions, and (b) positive or negative sentiment. The VAD Lexicon contains scores for 20,007 words along the dimensions of valence, arousal, and dominance. The scores are real-valued and range from 0 to 1 along each of the VAD dimensions. Note that the scores do not have intrinsic meaning; rather, they represent the relative rankings of words along each axis.

In both cases, the lexicons were originally created for English words; multi-lingual versions of the lexicons are also available, but were obtained by simply translating the English words to other languages. This can lead to some ambiguity, as one word may have multiple possible translations, and words may have different emotional connotations in different languages and cultures. However, Mohammad et al. (2016) showed that when words were automatically translated from English to Arabic, 90% of the Arabic words had the same sentiment associations as the original English word, and Afli et al. (2017) reported similar results for Irish.

In our analysis, the tweets are first tokenized using the NLTK tweet tokenizer. We ignore all words from the Cornell stoplist,[4] as well as the word token *robot*, which occurs in 30% of all tweets. From the remaining word tokens, we include only the subset of words which are listed in both the Emotion and VAD lexicons. For the basic emotions, we measure the percentage of words associated with that emotion (i.e. out of every 100 words, how many are associated with sadness, joy, etc.). For the VAD analysis, we focus primarily on valence, and report the average valence of all words (which are present in the lexicons), as well as the distributions of valence values.

The number of tweets and word tokens for each phase, as well as the number of word tokens which are represented in the lexicons, are given in Table 1. The 'Other' row includes tweets written before hitchBOT's destruction, but while it was not actively travelling (e.g. between trips). The 'Post-USA' row includes tweets posted after hitchBOT's destruction which ended the USA trip.

4 Analyses

4.1 A contrast of nations

In the first analysis, we aimed to compare the emotion words produced during each of the four trips (i.e., the first four rows in Table 1). However,

[3] http://saifmohammad.com/WebPages/lexicons.html

[4] http://www.lextek.com/manuals/onix/stopwords2.html

Phase	Tweets	Tokens	Lex.
1 Canada	8,490	131,846	21,843
2 Germany	1,625	23,171	3,457
3 Netherlands	211	2,970	478
4 USA	2,703	44,565	7,415
5 Other	5,316	82,090	13,430
6 Post-USA	45,287	714,441	116,752
Total	*63,632*	*999,083*	*163,375*

Table 1: Number of tweets and word tokens in the various phases of hitchBOT's existence, after preprocessing. The 'Lex.' column indicates the number of tokens appearing in both lexicons.

due to the relatively small number of tweets available for the Netherlands trip, we exclude these data and compare only Canada, Germany, and the USA (note that the USA data includes only those tweets produced *before* hitchBOT's destruction). This corresponds to lines 1, 2, and 4 in Table 1.

Only a small fraction of tweets are labelled with location information, and so for each country we include all tweets posted within the duration of hitchBOT's visit to that country, with the assumption that much of the Twitter content will be generated from inside the country of interest. There is some evidence to support this: during the Germany trip, 75% of tweets were written in German (compared to 7% during the Canada trip 1% during the USA trip).

Figure 2 shows the percentage of words associated with each emotion during each phase. Qualitatively, the distributions are similar across the trips, with Twitter users in all countries producing more positive than negative emotion words, and more words associated with anticipation and joy than anger, disgust, fear, and sadness. However, there are some differences as well. To determine whether the differences between countries are significant, we first perform a χ^2 test for each emotion, comparing the observed word counts for each emotion for each of the three countries to the expected counts under the null hypothesis of no difference between the countries. The χ^2 test is appropriate in the case of unequal sample sizes, as we have here. Since we repeat this test 10 times, we use a Bonferroni-adjusted α of 0.005 as the significance threshold. In cases where a significant difference is observed, we conduct a post-hoc pairwise proportion test to determine between which countries the relevant differences occur. Since the pairwise procedure involves three comparisons, we use $\alpha = 0.016$ as the threshold for significance.

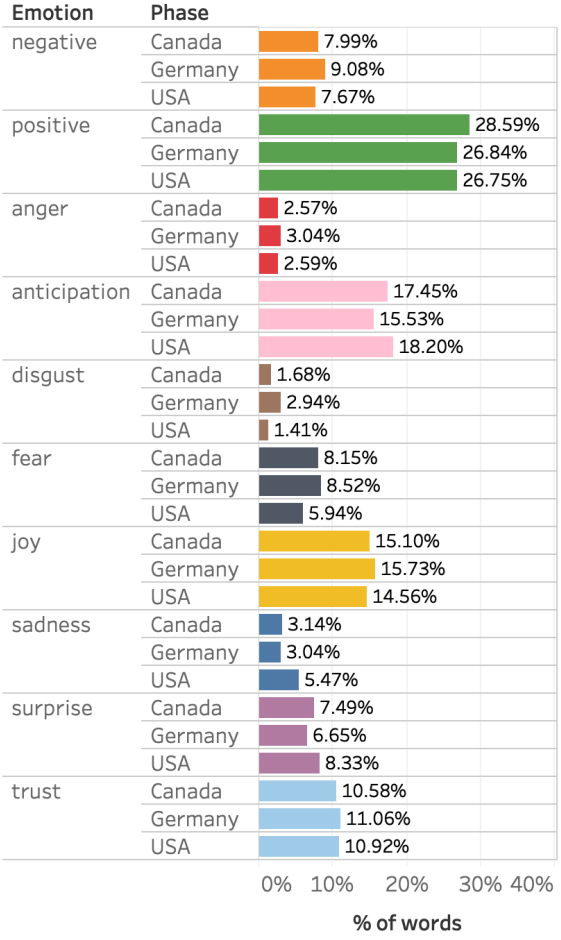

Figure 2: A comparison of the emotions expressed in tweets while hitchBOT travelled through different countries. For all words in the tweets which are contained in the emotion lexicon, we show the percentage of those words that are associated with the various emotions, by country.

Considering first the overall sentiment, there is no significant different in the percentage of negative words produced in the three countries. Canada produces the highest percentage of positive words, although the difference is only significant compared to the USA ($p = 9.8 \times 10^{-5}$).

For the basic emotions, there are no significant differences between the countries on anger, anticipation, joy, surprise, or trust. For disgust, Germany has a significantly higher percentage than both Canada ($p = 6.6 \times 10^{-5}$) and the USA ($p = 7.0 \times 10^{-6}$). The USA has the lowest percentage of fear words, significantly lower than both Canada ($p = 2.2 \times 10^{-8}$) and Germany ($p = 6.1 \times 10^{-5}$). Finally, the USA has the highest percentage of words associated with sadness compared to both Canada ($p = 4.2 \times 10^{-18}$) and Germany

($p = 6.8 \times 10^{-6}$).

While it is not possible here to analyze each of these trends in detail, we do consider two illustrative examples of what kinds of words are driving these differences:

Why were people sadder during the USA trip, even before hitchBOT's death? This pattern turns out to be driven by multiple discouraged tweets around the start of hitchBOT's American journey, when the robot did not manage to leave its starting point for a week, e.g. *hitchhiking robot's cross-country trek off to a* ***sluggish*** *start* and *a cross-country hitchhike is* ***tough*** *if no one will help you* ***leave*** *massachusetts*. Since hitchBOT's destruction cut the trip short after only two weeks, these early tweets have a larger impact than if the trip had been completed as expected.

Why were people more disgusted during the Germany trip? The most frequent word associated with disgust during the Germany trip is the German *schade*, which in the NRC lexicons is translated as the English *bummer*, which is associated with disgust. However, *schade* could also be translated as *shame* or *pity*; in the Emotion Lexicon, *shame* is also associated with disgust, but *pity* is not. This illustrates how different translations of the same word may have slightly different emotional connotations. (A manual review of the German tweets reveals that most occurrences of this word correspond to the sense of "what a pity," rather than explicit disgust towards hitchBOT.)

While these differences certainly merit further investigation, the overall impression is of remarkably similar emotional profiles in each of the three countries visited.

4.2 The death of hitchBOT

In the second analysis, we partition the dataset into those tweets written before and after hitchBOT's destruction (lines 1–5 versus line 6 from Table 1). For convenience, we refer to these time periods as *Life* and *Death*, respectively. Note that these tweets could have been posted from anywhere in the world, as long as they were written in English, French, German, or Dutch. Figure 3 shows the percentages of words associated with the eight basic emotions as well as positive and negative sentiment. The difference in emotion word percentages between life and death is significant for every emotion and sentiment (according to a χ^2 test and corrected for multiple comparisons).

Most trends are as expected, with increases in anger, disgust, fear, sadness, surprise, and negative sentiment after hitchBOT's death. Similarly, we observe a decrease in anticipation, joy, and positive sentiment. Counter-intuitively, the percentage of trust words shows a small but significant increase after death. An examination of the data suggests multiple reasons for this, including: the negation of trust words (e.g. *hitchhiking not* ***safe*** *for robots either in us*), irony (e.g. *welcome to the city of* ***brotherly*** *love*), and word-sense ambiguity (e.g. *adorable hitchhiking hitchbot* ***found*** *mutilated*).

In terms of the magnitude of the changes, the greatest relative difference is seen in the emotions of anger (4.7 times greater after death) and disgust (3.8 times greater), followed by sadness (3.6 times greater). This pattern seems reasonable, given the deliberate nature of the destruction.

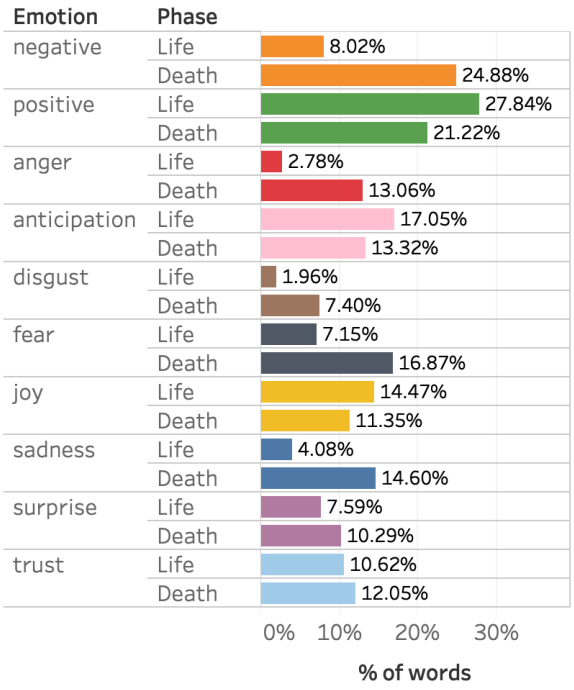

Figure 3: A comparison of the emotions expressed in tweets before and after hitchBOT's destruction. For all words in the tweets which are contained in the Emotion Lexicon, we show the percentage of those words that are associated with the various emotions.

We then consider the valence distribution of the words produced before and after hitchBOT's destruction. Valence is similar in some ways to the positive-negative sentiments discussed above, but contains much richer information about the inten-

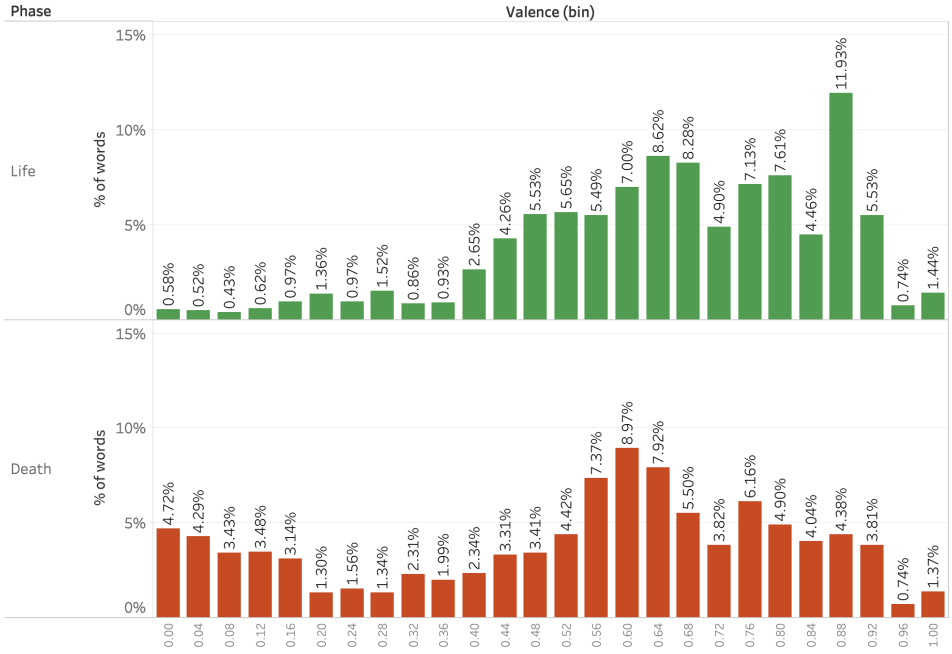

Figure 4: The valence distribution before/after hitchBOT's destruction, for words contained in the VAD lexicon.

sity of the emotion. If we consider only the mean valence, we do see a reduction from 0.67 in life, to 0.55 in death. However, Figure 4 offers a more detailed picture of how the valence distribution changes. A Kolmogorov–Smirnov test indicates that the two distributions are significantly different ($p < 0.001$). Specifically, we observe a substantial increase in the lowest-valence words (i.e. those expressing strong negative emotion) after death.[5]

To qualitatively examine the words which are found in these lowest-valence bins, the most highly-frequent words in the three lowest bins are given in Table 2. An interesting feature of these words is how many of them reflect some level of anthropomorphism and/or moral judgment. For example, words like *die*, *death*, *demise*, *killing*, and *kill*, imply the end of a life, at least metaphorically. The word *murder* is even stronger, since it denotes specifically the unlawful ending of a *human* life. In terms of moral judgments, the high frequency of words such as *blame*, *shame*, *terrible*, and *wrong* suggest the belief that there was something ethically wrong with destroying the robot. The word *crime* is also significant in this context, implying that this action was not just ethically but also legally unacceptable.

[5]The distributions for arousal and dominance, as well as interactive visualizations for all the figures in this paper, can be accessed at http://saifmohammad.com/WebPages/hitchbot.html

However, these views are far from universal. We also observe many words which are not usually associated with actions against animate beings, such as *destroyed* and *destruction*, as well as *broken* and *wrecked* (not visible in Table 2, with frequencies of 55 and 53, respectively). Furthermore, we note an apparent dissociation between the degree of anthropomorphism expressed in the tweet, and the polarity of the sentiment regarding hitchBOT's destruction. For example, among tweets expressing dismay at the incident, some mourn the loss of hitchBOT merely as a piece of technical equipment in a science experiment:

> *so a canadian robotics students' long, successful experiment in trust ended a few weeks after entering the states :(*

while others attribute personality and mental state to hitchBOT, and even refer to it with a nickname:

> *oh 'merica, what did you do to our sweet, sweet @hitchbot poor little hitchy.*

Tweets which celebrate hitchBOT's destruction are in a minority, but there are several, and they similarly range from describing hitchBOT simply as an object (albeit, an object that could be 'killed'):

i'm glad we killed hitchbot before it became trendy to transport roadside trash around the country

to attributing gender and personality:

philadelphia saves the world again, kills #hitchbot. he was a smug bastard and deserved to die

Attributing human-like characteristics to a robot, but then not ascribing it moral standing (e.g. the right not to be harmed), has interesting parallels to the findings of Friedman et al. (2003) with respect to AIBO dogs, and may also relate to the observations of Neff and Nagy (2016) that Tay the chatbot was sometimes viewed as a threat to, rather than a victim of, humanity. However, the examples here are merely anecdotal and additional work will be required to annotate the data for these various attitudes before we can draw further conclusions.

5 Discussion

There is a potential gap between what people write on Twitter and how they truly feel about robots and their destruction. On such a platform, there may be a tendency to use emotionally provocative language to attract attention and retweets. Even ignoring this effect, clearly we can say, for example, that a battery is 'dead' without thinking that it was ever really alive. Friedman et al. (2003) also discuss this disconnect between language and belief, observing that in many cases, anthropomorphic language is used playfully and as an informal shorthand (even in this paper, we find it simpler to refer to hitchBOT's *life* rather than *the period of time prior to hitchBOT's destruction*). In their work, Friedman et al. (2003) conclude that, "we are not saying AIBO owners believe literally that AIBO is alive, but rather that AIBO evokes feelings *as if* AIBO were alive." Similarly, we certainly do not propose that the use of anthropomorphic language indicates that Twitter users actually believed hitchBOT was a living thing, but rather that their lexical choices reflect an emotional connection with the robot, and subsequent empathetic reaction to its destruction, akin in some ways to that which might be evoked by a living being.

The human tendency towards anthropomorphism can have far-reaching consequences in terms of what we view as ethical behaviour. In a thought-provoking discussion of whether social robots should be extended any type of legal protection, Darling (2016) argues that many of our existing laws protecting animals from abuse are based on our anthropomorphism and emotional connection with animals, rather than, e.g., biological factors (the fact that it is legal to slaughter a cow for food but not a horse seems based primarily in our cultural emotional connection to horses). Darling (2016) also writes that one interpretation of the purpose of law is to codify a social contract: "We construct behavioural rules that most of us agree on, and we hold everyone to the agreement." From that perspective, it is important to start gathering data on the nature and extent of society's emotional attachments to robots of various kinds.

Freq.	Token	Freq.	Token	Freq.	Token
3061	destroyed	214	shame	127	blame
1477	demise	205	shit	124	terrible
739	death	185	hate	119	hell
417	destruction	178	wrong	105	die
413	murder	175	tragic	104	dangerous
360	kill	156	destroying	101	crime
277	mutilated	155	upset	93	violence
234	doomed	147	damn	91	war
228	killing	143	violent	87	sadly
216	fake	132	hurt	85	incident

Table 2: The highest frequency words in the lowest valence bins after hitchBOT's destruction.

The current analysis is limited in a number of ways. Emotion is analyzed on the word level, rather than the sentence level, and as such we do not take into account negations or any other context, nor do we attempt to detect sarcasm. In particular, we cannot ensure that hitchBOT is the actual entity to which the emotion is attached (e.g., *when he comes to this **great** nation's **beautiful** capital, i want to be able to drive him through it.*). Furthermore, the amount of data in each phase is not balanced, with the majority of tweets occurring after hitchBOT's destruction, and the particularly small number during the trip to the Netherlands limited our cross-cultural analysis.

Nonetheless, while somewhat exploratory in nature, these preliminary analyses suggest several avenues for future research. By analyzing tweets on the sentence-level and conducting a topic analysis, we can get a better sense of what attitudes and beliefs are underlying people's emotional word choices. Additionally, by manually annotating tweets for attributions to hitchBOT of life-likeness, emotional states, intention, and so on

(following the work of Fink et al. (2012)), we can start to draw a clearer link between anthropomorphic language and emotional attachment. In future work we also plan to look more specifically into different cultures and their perceptions, using various lexicons. We can also consider machine learning approaches to emotion analysis, as well as recent advances in lexicon-based approaches (Buechel and Hahn, 2016). Finally, although the corpus is not currently publicly available, we do plan to release the data to other researchers in the future.

6 Conclusion

We have presented an analysis of the emotion words produced by Twitter users about hitchBOT. When comparing tweets written during each of hitchBOT's trips across Canada, Germany, and the United States, the emotion word percentages were generally similar, although some significant differences were observed, with Canadians expressing the most positive sentiment, and Americans expressing the least fear and the most sadness. While Germans expressed significantly more disgust than the others, this effect may be due to a near-synonym translation with a different emotional connotation than the original German word.

When examining the tweets written before and after hitchBOT's 'death', significant differences were observed in all of the basic emotions, with marked increases in the percentage of words associated with anger, disgust, and sadness. The proportion of words with very low valence scores (i.e. those expressing negative sentiment) also increased dramatically. A qualitative analysis of these low-valence words suggests that Twitter users perceived the actions of the vandals as morally corrupt, with an intensity of emotion that seems incommensurate with an interpretation of the event as simple property damage. These findings will hopefully provoke future questions probing how humans should behave towards robots and towards discussions around robot ethics.

References

Haithem Afli, Sorcha McGuire, and Andy Way. 2017. Sentiment translation for low resourced languages: Experiments on Irish general election tweets. In *18th International Conference on Computational Linguistics and Intelligent Text Processing*.

Christoph Bartneck, Tomohiro Suzuki, Takayuki Kanda, and Tatsuya Nomura. 2007a. The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & Society*, 21(1-2):217–230.

Christoph Bartneck, Marcel Verbunt, Omar Mubin, and Abdullah Al Mahmud. 2007b. To kill a mockingbird robot. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 81–87. ACM.

Sven Buechel and Udo Hahn. 2016. Emotion analysis as a regression problemdimensional models and their implications on emotion representation and metrical evaluation. In *Proceedings of the Twenty-second European Conference on Artificial Intelligence*, pages 1114–1122. IOS Press.

James Burch. 2018. In Japan, a Buddhist funeral service for robot dogs. *National Geographic*. [Online; accessed 20-November-2018].

Julie Carpenter. 2016. *Culture and human-robot interaction in militarized spaces: A war story*. Routledge.

Prerna Chikersal, Soujanya Poria, and Erik Cambria. 2015. SeNTU: Sentiment analysis of tweets by combining a rule-based classifier with supervised learning. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 647–651.

Luisa Damiano, Paul Dumouchel, and Hagen Lehmann. 2015. Towards human–robot affective co-evolution overcoming oppositions in constructing emotions and empathy. *International Journal of Social Robotics*, 7(1):7–18.

Kate Darling. 2016. Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In Ryan Calo, Michael Froomkin, and Ian Kerr, editors, *Robot Law*. Edward Elgar Publishing.

Kate Darling, Palash Nandy, and Cynthia Breazeal. 2015. Empathic concern and the effect of stories in human-robot interaction. In *Robot and Human Interactive Communication (RO-MAN), 2015 24th IEEE International Symposium on*, pages 770–775. IEEE.

Paul Ekman, E Richard Sorenson, and Wallace V Friesen. 1969. Pan-cultural elements in facial displays of emotion. *Science*, 164(3875):86–88.

Existor. 2016. Cleverscript - turn scripts into bots. `http://www.cleverscript.com`. [Online; accessed 7-February-2016].

Julia Fink, Omar Mubin, Frédéric Kaplan, and Pierre Dillenbourg. 2012. Anthropomorphic language in online forums about Roomba, AIBO and the iPad. In *Proceedings of the IEEE International Workshop on Advanced Robotics and its Social Impacts (ARSO 2012)*, pages 54–59.

Batya Friedman, Peter H Kahn Jr, and Jennifer Hagman. 2003. Hardware companions?: What online AIBO discussion forums reveal about the human-robotic relationship. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 273–280. ACM.

Anastasia Giachanou and Fabio Crestani. 2016. Like it or not: A survey of Twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*, 49(2):1–41.

Jochen Hirth, Norbert Schmitz, and Karsten Berns. 2011. Towards social robots: Designing an emotion-based architecture. *International Journal of Social Robotics*, 3(3):273–290.

Jihong Hwang, Taezoon Park, and Wonil Hwang. 2013. The effects of overall robot shape on the emotions invoked in users and the perceived personalities of robot. *Applied Ergonomics*, 44(3):459–471.

Barbara Kühnlenz, Stefan Sosnowski, Malte Buß, Dirk Wollherr, Kolja Kühnlenz, and Martin Buss. 2013. Increasing helpfulness towards a robot by emotional adaption to the user. *International Journal of Social Robotics*, 5(4):457–476.

Patrick Lin, Keith Abney, and George Bekey. 2011. Robot ethics: Mapping the issues for a mechanized world. *Artificial Intelligence*, 175(5-6):942–949.

Saif M. Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, pages 174–184, Melbourne, Australia.

Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. Semeval-2018 Task 1: Affect in tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.

Saif M Mohammad, Mohammad Salameh, and Svetlana Kiritchenko. 2016. How translation alters sentiment. *Journal of Artificial Intelligence Research*, 55:95–130.

Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Omar Mubin, Aila Khan, and Mohammad Obaid. 2016. #naorobot: Exploring Nao discourse on Twitter. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, pages 155–159. ACM.

Gina Neff and Peter Nagy. 2016. Talking to bots: Symbiotic agency and the case of Tay. *International Journal of Communication*, 10:4915–4931.

Phoebe Parke. 2015. Is it cruel to kick a robot dog? *CNN*. [Online; accessed 20-November-2018].

Robert Plutchik. 1984. Emotions: A general psychoevolutionary theory. *Approaches to emotion*, pages 197–219.

Laurel D Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. 2009. How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE international Conference on Human Robot Interaction*, pages 245–246. ACM.

Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. Semeval-2015 task 10: Sentiment Analysis in Twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 451–463.

James A Russell. 2003. Core affect and the psychological construction of emotion. *Psychological Review*, 110(1):145–172.

David Harris Smith and Frauke Zeller. 2017a. The death and lives of hitchBOT: The design and implementation of a hitchhiking robot. *Leonardo*, 50(1):77–78.

David Harris Smith and Frauke Zeller. 2017b. hitchBOT: The risks and rewards of a hitchhiking robot. *Suomen Antropologi: Journal of the Finnish Anthropological Society*, 42(3):63–65.

Ja-Young Sung, Lan Guo, Rebecca E Grinter, and Henrik I Christensen. 2007. "My Roomba is Rambo": Intimate home appliances. In *International Conference on Ubiquitous Computing*, pages 145–162. Springer.

Duyu Tang, Furu Wei, Bing Qin, Ting Liu, and Ming Zhou. 2014. Coooolll: A deep learning system for twitter sentiment classification. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 208–212.

James E Young, Richard Hawkins, Ehud Sharlin, and Takeo Igarashi. 2009. Toward acceptable domestic robots: Applying insights from social psychology. *International Journal of Social Robotics*, 1(1):95.

Yang Yu and Xiao Wang. 2015. World Cup 2014 in the Twitter world: A big data analysis of sentiments in US sports fans tweets. *Computers in Human Behavior*, 48:392–400.

Frauke Zeller and David H. Smith. 2014. The Hitchbot's guide to travelling across a continent. http://theconversation.com/the-hitchbots-guide-to-travelling-across-a-continent-31920. [Online; accessed 27-February-2016].