institute for SOFTWARE RESEARCH

Social Media ACTion:
SOLO Data Description

William Frankenstein
Kenneth Joseph
Kathleen M. Carley
February 2, 2016
CMU-ISR-16-103

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Center for the Computational Analysis of Social and Organizational Systems
CASOS technical report.

Carnegie Mellon

Abstract

This technical report summarizes the demographics and Socially Observed Linked Opinion (SOLO) dataset, which came out of the Social Media ACTion study that took place at Carnegie Mellon during the summer of 2015. 124 individuals rated 4,320 social media posts and 1,680 news clips along the three dimensions used in Affect Control Theory. The report includes a description of the data, the training materials provided, and the consent form used for this study.

Table of Contents

1   Introduction

This report describes the SOLO dataset, which is available to all researchers at Carnegie Mellon University. The report has the following sections: study background, study setup and interface, demographics, and data format description. This study was completed under IRB code HS14-670.

2   Study background

"State-of-the-art" research in sentiment analysis has three problems: the approaches were developed to analyze large bodies of text, it ignores the social context of social media, and it does not consider social media's international dimension. Social media text can be extremely short – making traditional machine learning approaches difficult, as the data to be classified has features not included in the training set. It is inherently social – frequently responding to individuals or events.

Most approaches focus exclusively on content[1]-[3]. For example, a user posting she is ill will receive positive, supportive posts on social media. The illness would be misclassified as a positive event due to the positive words in their responses. Finally, posts contain international content – cultures affect how individuals respond to events. Affect control theory formalizes the way that individuals respond to events by classifying evaluation, potency, and action, allowing for cross-cultural comparisons of events [4], [5].

To address these problems, the Social Media ACTion study had three primary goals:
1. To examine the role of context in evaluating valence of social media posts
2. To expand the lexicons available for Affect Control Theory
3. To develop a gold standard sentiment dataset of hand-labeled social media and news posts

To achieve the first goal, participants were asked to evaluate a set of Twitter posts twice: once, seeing a Twitter response post before seeing the original post, and the second time, seeing the response post directly beneath the original post. The second goal is product of analysis done on this dataset. The third goal is the primary focus of this technical report.

3   Study setup

3.1 Study Participants

Individuals were recruited for 45 minute sessions to evaluate 90 social media posts and received $8 compensation in the form of an Amazon Gift Card. Individuals were recruited from both the CMU Center for Behavioral and Decision Research (CBDR) as well as from flyers posted around the Oakland neighborhood of Pittsburgh. To qualify for the study participants had to be over 18 years of age and native English speakers to ensure that participants understood all social media posts. Individuals who did not finish the study were compensated at the rate of 6 cents per social media post. More

information about participant demographics is available in a later section of this technical report.

3.2 Study interface

Initially, participants were asked to attend in-person sessions and input answers using an internal CASOS server using a modified version of a survey designed for collecting medical informatics [6]. To facilitate collection over the course of the summer, however, we switched to using Qualtrics after 6 individuals had taken our study. We have updated the data collected from these participants so that all data is comparable, regardless of which platform the data was collected from. In particular, following best practice in Affect Control Theory coding there are three features of the interface that we manipulated to reduce framing and anchoring heuristics: double labeled axes, changing the lateral direction of intensity for "Activity" evaluations, and having an individual axis on each page seen by the participant [7].
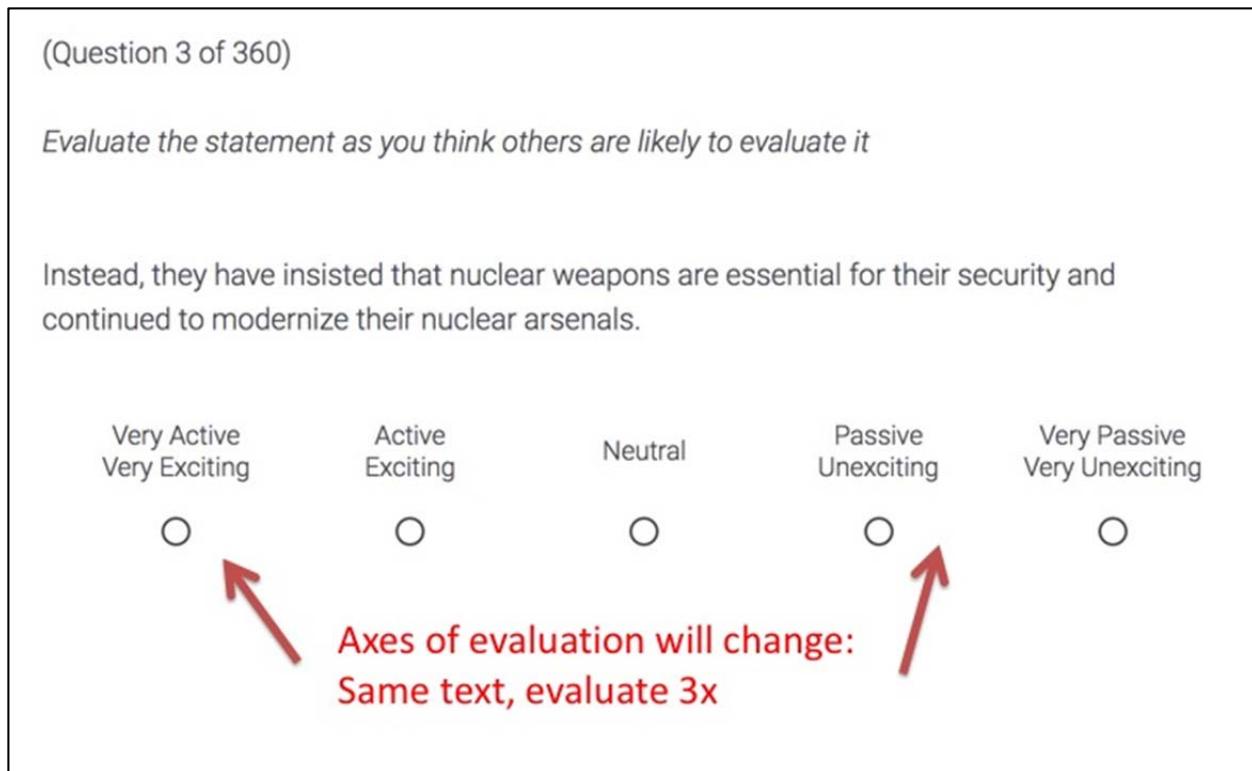
Figure 1. Sample evaluation screenshot taken from training slides.

In particular, note that:
1. We asked participants to evaluate the statement from a "general" perspective – which increases overall inter-rater agreement rates [8]
2. Participants rated statements on a 5-point Likert scale
3. Participants rated the same statement along Evaluation, Potency, and Activity scales immediately, on separate screens
4. Axes were given two reference points:

      a. Evaluation: Negative/Unpleasant to Positive/Pleasant
      b. Potency: Weak/Powerless to Strong/Powerful
      c. Activity: Active/Exciting to Passive/Unexciting
5. Activity was evaluated with "Active" on the left hand side and "Passive" on the right hand side to emphasize its distinction from potency

Participants each underwent a five minute training session to familiarize themselves with the ACT concepts of "Evaluation", "Potency", and "Activity". These slides, as well as the accompanying script, are available as an appendix to this technical report.

3.3 Questionnaire Structure

Participants who encoded Twitter posts rated posts in the following order (keep in mind ACT required participants to evaluate each post three times):

  I.   30 "standalone" Twitter posts (90 evaluations)
 II.   30 "conversation" Twitter posts (180 evaluations)
      a. Response seen first
      b. Original post seen second
III.   30 Response Twitter posts (90 evaluations)

The posts in section III were identical to the posts in section IIA; however, while in Section II they were presented in isolation, in section III they were presented together with the original post they responded to.

Participants who encoded news clips simply rated 120 sentences or sentence pairs.

3.4 Data Source

To ensure a diverse set of evaluations, we utilized four distinct topics across two platforms – Twitter and news articles pulled from Lexis Nexis. We ensured to the best of our ability that all tweets and news articles evaluated were in English; if non-English words were utilized, participants were instructed to mark the message as "Neutral".

For "general" topics – we utilized the "Gardenhose" Twitter dataset. It is part of a larger dataset available at CMU that is composed of 10% of the total Twitter firehose. We selected random tweets from this dataset to represent commonly used English on Twitter. For news articles, we utilized sentences from the 2014 New York Times set of editorials.

Table 1. Table of different topics used

	Nuclear	Arab Spring	General	Typhoon Haiyan
Dates Covered	Sep 2014 – Oct 2014	Oct 2009-Nov 2013	Sep 2013 – Aug 2014	November – December 2013
Sample Keywords	Nuclear proliferation, heavy water, uranium	Tahrir Square, Arab Spring	n/a for Gardenhose; New York Times editorials	Typhoon Haiyan, Typhoon Yolanda
Number of Twitter Posts	1,080	1,080	1,080	1,080
Number of News Clips	420	420	420	420

4   Demographics

The primary constraints on participants involved being able to attend an in-person coding session in Pittsburgh and being over 18 years old. We expected to have a body of participants that matched closely with the undergraduate population at Carnegie Mellon; while this was largely the case, we also had participants from the local Pittsburgh community. We had a total of 124 participants. All demographics questions were asked at the end of the study and were completely voluntary.

Table 2. Gender of participants

Female	77 (62%)
Male	47 (38%)
Other / Decline to state	0 (0%)

Table 3. Age distribution of participants

Under 25	73 (59%)
25-30	25 (20%)
31-40	10 (8%)
41-50	5 (4%)
Over 50	11 (9%)

While we screened for native English speakers, we asked participants to rate their own English ability. 4 individuals (3%) self-identified as speaking English "well" as opposed to 120 individuals (97%) identifying their English proficiency as native speakers.

We also asked individuals to identify other languages spoken at home. 88 (71%) of participants reported that they only spoke English at home.

Table 4. Count of other languages spoken at home by participants

No other languages spoken at home	88
American Sign Language	1
Chinese (Mandarin)	4
Czech	1
Spanish	6
French	3
Guajarati	2
Hindi	6
Tamil	5
Marathi	4
Telugu	1
Kannada	1
Taiwanese	1
Punjabi	1
Russian	2
Urdu	2
Vietnamese	1

We asked participants to identify their race and ethnicity. Participants were able to select more than one category of race. 4 individuals (3%) self-identified as being of Hispanic, Latino, or Spanish origin.

Table 5. Participant ethnic and racial distribution. Individuals could select multiple categories.

American Indian or Alaska Native	0 (0%)
Asian or Pacific Islander	43 (35%)
Black or African American	12 (10%)
White	72 (58%)
Other	4 (3%)

5   Data Format

The data is available on the CASOS Megadon server, which can be accessed at megadon.casos.cs.cmu.edu. The data is located on the D:// drive under "Public SOLO Data". An additional folder containing the questionnaires uploaded to Qualtrics, for researchers interested in replicating the study, is available upon request.

Eval_Tweets and Eval_News contain Evaluative ratings of tweets and news clips respectively; Power_Tweets and Power_News contain Power ratings, and Active_Tweets and Active_News contain Activity tweets.

Tweet row names have the format:
X[[NUMBER1]]_[R/S]_[[NUMBER2]].[[NUMBER3]].

News row names have the format: [[LETTER1]]_X[[NUMBER1]]_S_[[NUMBER2]].

[NUMBER1] refers to the Tweet ID - this is located in either ArabSpring, Garden, Haiyan, or NukeTweets.tsv. [LETTER1] identifies the news topic – "A" for Arab Spring, "G" for General, "N" for Nuclear, and "T" for Typhoon.

S indicates that the tweet or message was evaluated in isolation.

R indicates that the tweet or message was evaluated in context. You can view the mapping of what this tweet responded to in the XX_Pairs.tsv text file.

[NUMBER2] refers to the EPA rating, where 0 = evaluative, 1 = power, 2 = activity.

[NUMBER3] occasionally some tweets were selected twice. If this is the case, there will be a .1 or .2 after the main string identifying the tweet.

For Evaluation, the 5 point Likert goes from Most negative = 1 to most positive = 5.
For Power, the 5 point Likert goes from Weakest = 1 to Strongest = 5.
For Activity, the 5 point Likert goes from Most Active = 1 to Most Passive = 5.

6   Lessons Learned

Over the course of conducting this study, several lessons were learned. I hope this document can serve to improve future studies.

For Carnegie Mellon researchers – I highly recommend advertising the study on the Center for Behavioral and Decision Making Research (CBDR) website (http://cbdr.cmu.edu). Studies performed online (such as those which would normally be done through Amazon Mechanical Turk) can also be posted to that website. I had significantly more success recruiting participants through CBDR than through flyers.

Carnegie Mellon also maintains a subscription to Qualtrics. This is particularly useful as Qualtrics allows for individuals to create relatively customized questionnaires very easily, as outlined in their technical support [9]. What isn't mentioned in their reference notes is that you can utilize basic HTML formatting – which significantly improves the presentation of the questionnaire.

7 References

[1] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining," presented at the Proceedings of LREC, 2006.

[2] J. W. Pennebaker, C. K. Chung, and M. Ireland, "The development and psychometric properties of LIWC2007," LIWC.net, Austin, TX, USA, LIWC2007, 2007.

[3] P. J. Stone, *User's Manual for The General Inquirer*. MIT Press (MA), 1968.

[4] L. S. Lovin, "Affect control theory: An assessment*," *Journal of Mathematical Sociology*, vol. 13, no. 1, pp. 171–192, Jan. 1987.

[5] D. R. Heise, "Affect control theory: Concepts and model," *Journal of Mathematical Sociology*, vol. 13, no. 1, pp. 1–33, Jan. 1987.

[6] M. M. Benham-Hutchins, B. B. Brewer, K. M. Carley, M. Kowalchuk, and J. A. Effken, "Development of a Social Network Analysis Data Collection Application," *Under Review*, Jan. 2016.

[7] G. P. Morgan and J. H. Morgan, "Surveyor 3.0: A Note on an Open Source Application for Sentiment Analysis," *Sociological Methods & Research*, pp. 1–12, Oct. 2014.

[8] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," presented at the … AAAI Conference on Weblogs and Social Media, 2014.

[9] "Import and Export Surveys - Qualtrics" [Online]. Available: http://www.qualtrics.com/university/researchsuite/advanced-building/advanced-options-drop-down/import-and-export-surveys/

8   Appendix A: Training Slides

Slide 1
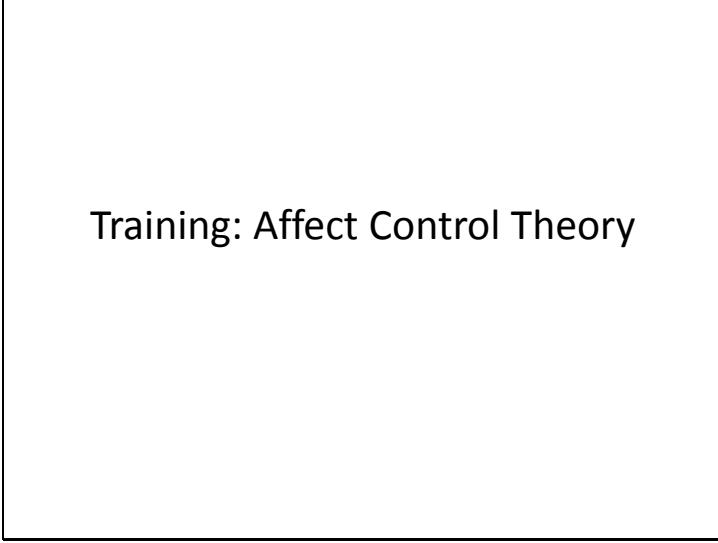
Training: Affect Control Theory

Slide 2

Study Goal: Expand Concepts of Sentiment

- Goal of study is to develop understanding of sentiment based on different feelings through a framework called Affect Control Theory

Slide 3

Why are we trying to do this?

- Most tools currently developed only look at positive and negative sentiment:

- We know, however, there are lots of other aspects of sentiment, as well as a social theory behind sentiment

Slide 4

Affect Control Theory (ACT)

- We perceive social identities through dimensions of sentiment
- Social events change sentiment and evoke emotion within us
- Structuring sentiment along three specific dimensions allows for cross-cultural comparisons of emotion

Heise, "Social action as the control of affect". *Systems Research and Behavioral Science* 1977 (22).
Heise, *Expressive Order: Confirming Sentiments in Social Actions.* (2007).

Slide 5

ACT Captures Multiple Dimensions of Sentiment
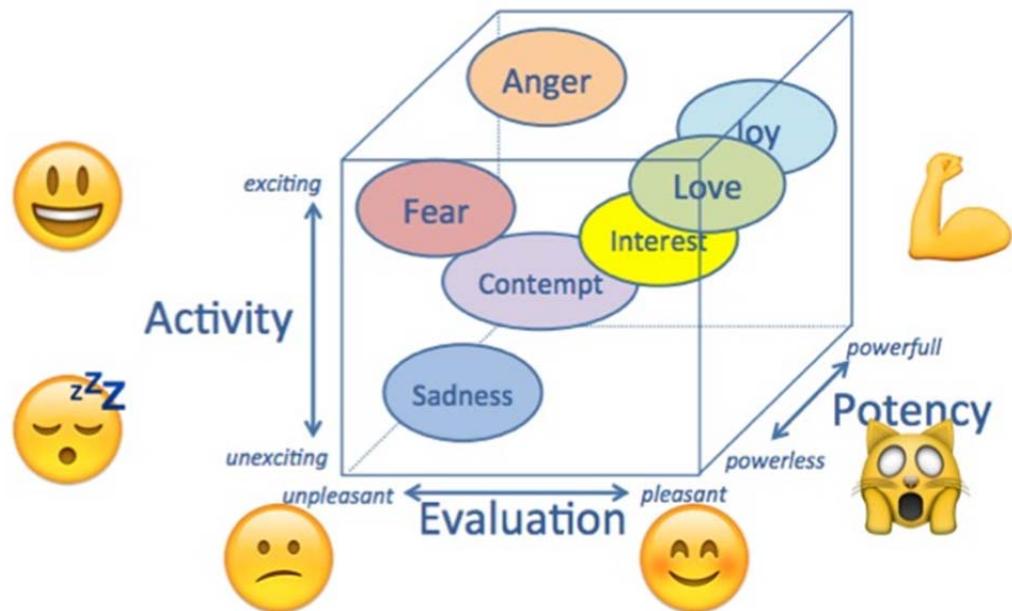
Illustration from Tobias Schröder's ACT lectures:
https://cs.uwaterloo.ca/~jhoey/research/bayesact/lectures/lecture1/lecture1Slides.pdf

17

Dimensions of Sentiment (cont.)

- ACT's three dimensions of sentiment are:
  - **Evaluation** – how "good" or "bad", "positive" or "negative", "pleasant" or "unpleasant" something is
  - **Potency** – how "powerful" or "powerless" something is, the degree of status something or someone exhibits
  - **Activity** – how "active" or "passive" something is, the level to which it provokes excitement

Slide 7 – Two versions provided. Initially used "Breaking Bad" example to illustrate how Twitter users had conversations on the platform, switched to President Obama's "peas in guacamole" comment after it was made during the summer of 2015.

Conversations example 1:

Conversations on Twitter

THE most comprehensive and important analysis on US N waste. Breaking Bad: A Nuclear Waste Disaster [URL]

← Original post

@originalposter you call this a disaster? I've passed gas that was more toxic than what leaked out of WIPP.

@originalposter Good info.

Response posts
Have @user at front of message

Conversations on Twitter

Add green peas to your guacamole. Trust us. [[URL]]

← Original post

respect the nyt, but not buying peas in guac. onions, garlic, hot peppers. classic. [[URL]]

.@nytimes Don't stop there. Ever get them on your ballpark frank? Mmm, mm. [[URL]]

Response posts normally have @user at front of message

Slide 8

# Twitter Peculiars

- Retweets: how information is spread across twitter
  – Sometimes prefaced with RT, or quotes around the tweet
  – E.g. "@potus: respect the nyt, but not buying peas in guac. onions, garlic, hot peppers. classic. [[URL]]"
  – MT: 'Modified Tweet', effectively the same
- Hashtags: #whatsupwiththat #topic #trending
  – Use # to indicate topics
  – Some tweets have multiple hashtags
- Responding to others: @user1
  – Sometimes posts have @user at beginning of tweet
  – Messages with @user later in message used to alert others

Slide 9
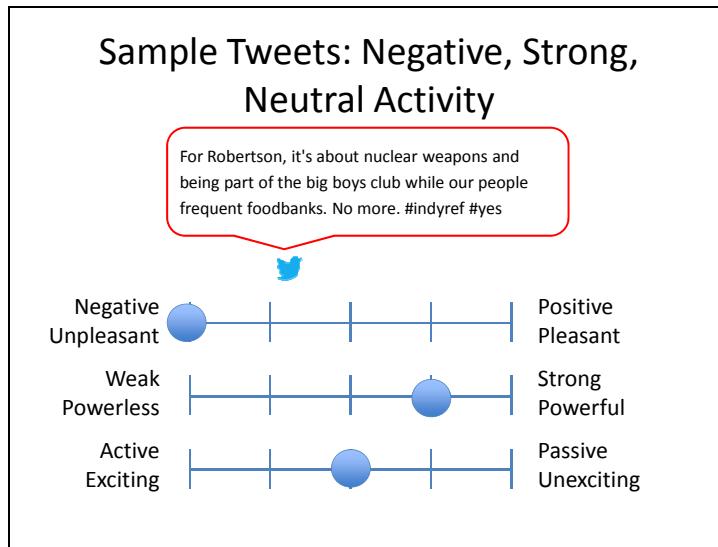
Message clearly negative on Robertson. Robertson a powerful individual, "part of the big boys club"; neither active nor passive as it's unclear what action he is taking.

Slide 10
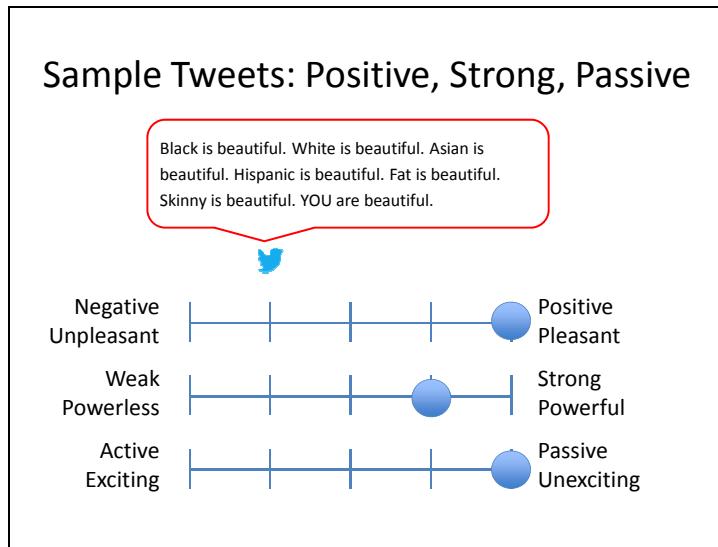
All positive statements. All empowering statements – however, also all passive. Unclear what action is being taken here.

Slide 11
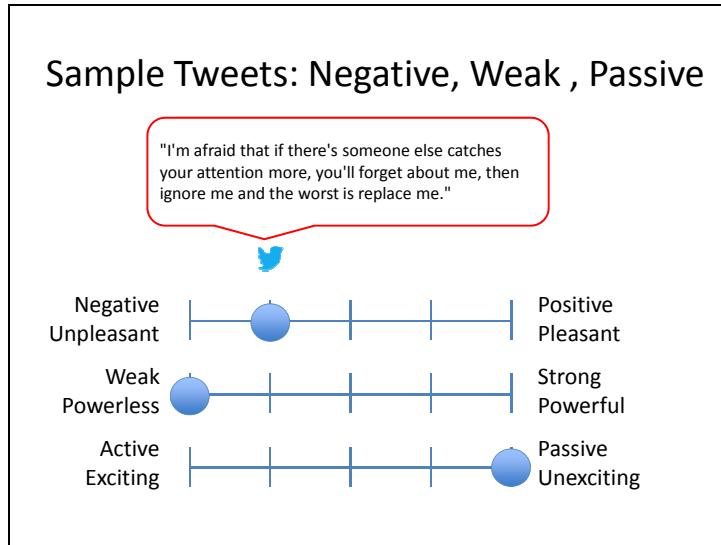
Message of a 14-year-old angsty teenager that you want to reach out and hug – clearly, a negative message, they feel weak, and they feel inactive.
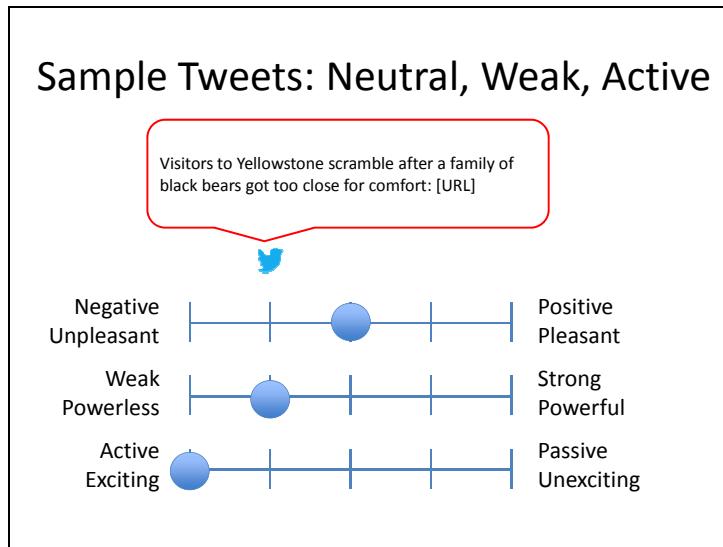
Slide 12

Here's an example of a neutral tweet – neither positive or negative; it's just reporting the news. However, it's weak as the people in the sentence had to run away – which is itself an active process.

Slide 13

Sample Tweets: Negative, Strong and Passive
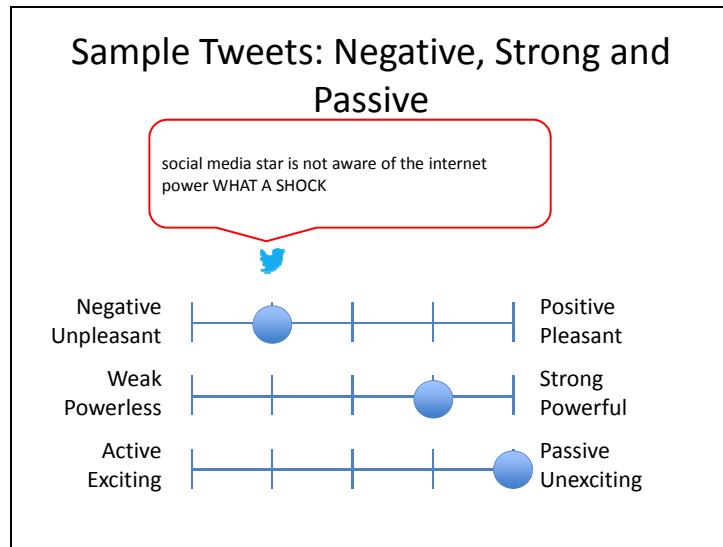
Finally, a sarcastic tweet. A bit negative, mocking a powerful "social media star" – someone with status. But also passive; the message mocks the "star" for their inaction.

Twitter evaluators shown the following three slides, which highlight "User 1" and "User 2"

Slide 14

Slide 15

Interface of tool: Context

User 1:

All those deaths, all those Martyrs, all that blood , for what? #Syria

How does seeing the original context of the tweet impact evaluation?

Evaluate User 2's statement

User 2:

"@user1: All those deaths, all those Martyrs, all that blood , for what? #Syria Never underestimate Religious Cults & Big Money :-)

Very Negative Unpleasant	Negative	Neutral	Positive	Very Positive Pleasant
○	○	○	○	○

Individuals evaluating news clips first saw the news clip examples, followed by screenshots of the interface.

Slide 17

<div style="border:1px solid black; padding:1em;">

360 questions

- Breakdown of questions:
- 120 news clips

</div>

Slide 18

Slide 19

Interface of tool

(Question 3 of 360)

Evaluate the statement as you think others are likely to evaluate it

Instead, they have insisted that nuclear weapons are essential for their security and continued to modernize their nuclear arsenals.

Very Active Very Exciting	Active Exciting	Neutral	Passive Unexciting	Very Passive Very Unexciting
○	○	○	○	○
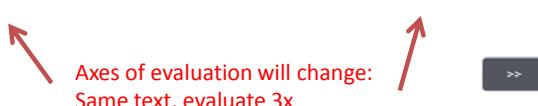
Axes of evaluation will change: Same text, evaluate 3x

The following news clips examples were provided to individuals rating news statements.

Slide 20
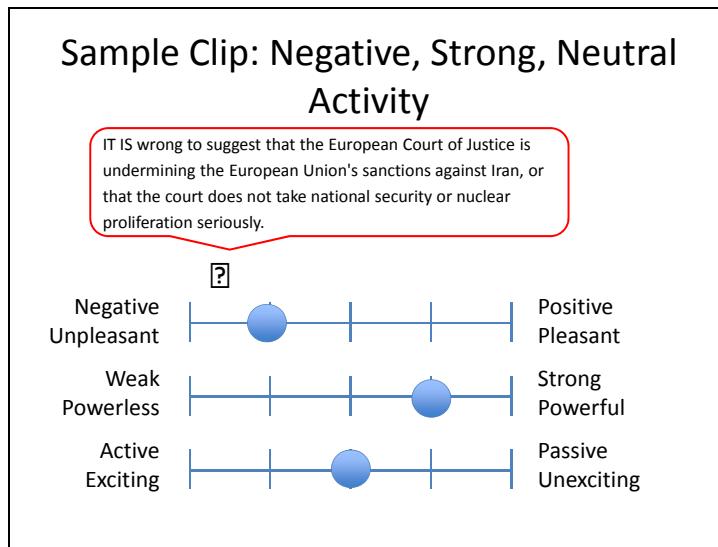
Slightly negative – the Court of Justice is clearly missing something here. These are powerful institutions. However, since we don't know what action is taking place, neither active nor passive.

Slide 21
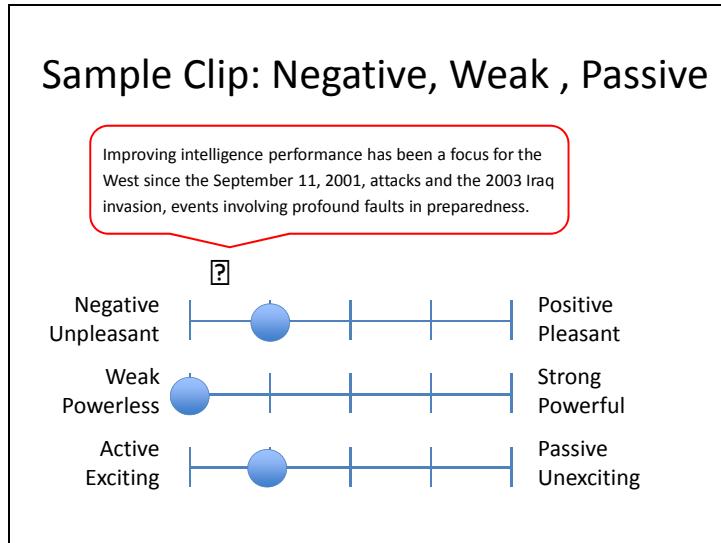
A negative statement – "faults in preparedness". The institution is being brought up in the context of weakness – 9/11 – so weak. But very active – the statement is focusing on how to improve intelligence.

Slide 22
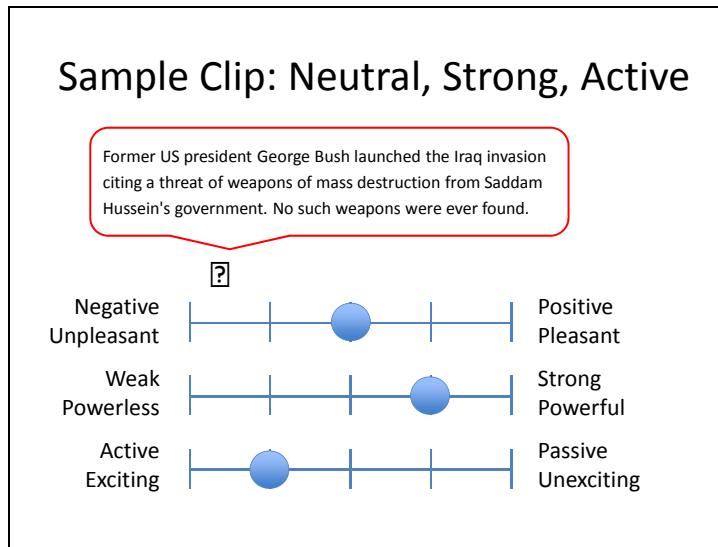
Neutral statement – while written in a slightly negative tone, alone these sentences are neutral. Strong institutions referenced. Active positions taken.

Slide 23

Sample Clip: Negative, Strong and Passive

The Senate is scheduled to hold hearings today on a dangerous new treaty negotiated by the Clinton Administration that would lift longstanding controls on nuclear trade.

Negative Unpleasant		Positive Pleasant
Weak Powerless		Strong Powerful
Active Exciting		Passive Unexciting
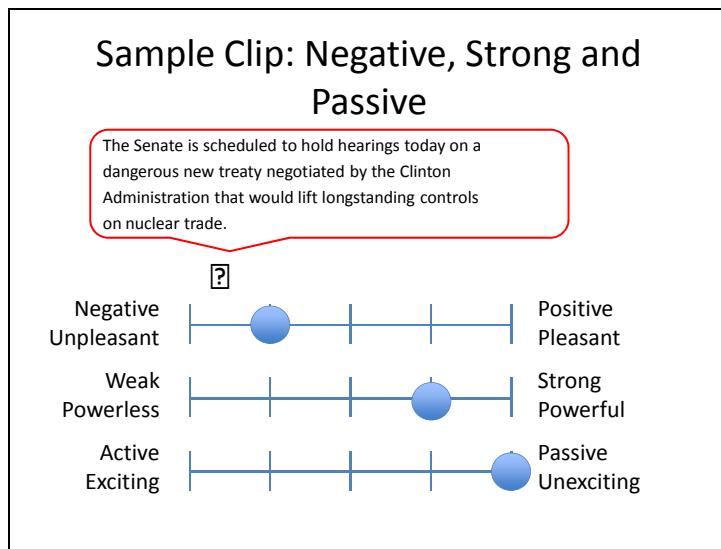
Slightly negative here – a "dangerous" new treaty. "Senate", "Clinton Administration" powerful institutions. However, passive – it's not that the Senate is holding hearings – they're "scheduled to hold" hearings.

9   Appendix B: Consent Form

This social media coding is part of a research study conducted by Will Frankenstein and Kenneth Joseph at Carnegie Mellon University and is funded by Crosswalk- Graduate Student Small Project.

The purpose of the research is to develop a 'gold standard' of social media posts encoded by affect control theory. The three dimensions measured in affect control theory are: emotion (positive to negative), powerful (weak to strong), and action (lively to quiet).

Procedures

Participants will view a variety of short, anonymous social media posts, and rate them along a 5 point Likert scale for each of the three dimensions of affect control theory. In some cases, participants will see a social media post twice; this will be done to provide more context for the original post.

The study is expected to take 45 minutes to complete.

Participant Requirements

Participation in this study is limited to individuals age 18 and older. Participants must be native English speakers.

Risks

The risks and discomfort associated with participation in this study are no greater than those ordinarily encountered in daily life or during other online activities. The primary risk to participants is boredom or fatigue from reading several social media posts in one sitting.

Benefits

There may be no personal benefit from your participation in the study but the knowledge received may be of value to humanity.

Compensation & Costs

Participants will be paid $8 in Amazon gift cards for completion of the study. Individuals who do not complete the study will be compensated at a rate of 6 cents per social media post viewed in Amazon gift cards.

There will be no cost to you if you participate in this study.

Confidentiality

The data captured for the research does not include any personally identifiable information about you. We will capture some summary demographic information about you, but it will not be linked to yourself or the data provided.

Your data and consent form will be kept separate. Your consent form will be stored in a locked location on Carnegie Mellon property and will not be disclosed to third parties. By participating, you understand and agree that the data and information gathered during this study may be used by Carnegie Mellon and published and/or disclosed by Carnegie Mellon to others outside of Carnegie Mellon. However, your name, address, contact information and other direct personal identifiers in your consent form will not be mentioned in any such publication or dissemination of the research data and/or results by Carnegie Mellon.

Right to Ask Questions & Contact Information

If you have any questions about this study, you should feel free to ask them by contacting the Principal Investigator, Will Frankenstein, PhD Candidate in Department of Engineering & Public Policy, Baker Hall 129, 5000 Forbes Avenue, Pittsburgh, PA 15213 / frankenstein@cmu.edu / 412-589-9788. If you have questions later, desire additional information, or wish to withdraw your participation please contact the Principal Investigator by mail, phone or e-mail in accordance with the contact information listed above.

If you have questions pertaining to your rights as a research participant; or to report objections to this study, you should contact the Office of Research integrity and Compliance at Carnegie Mellon University. Email: irb-review@andrew.cmu.edu . Phone: 412-268-1901 or 412-268-5460.

Voluntary Participation

Your participation in this research is voluntary. You may discontinue participation at any time during the research activity. However, not completing the study will mean that you will not be compensated for your time.

[Design the web page so that the following questions must be answered appropriately before the individual can proceed to the study task.]

I am age 18 or older. ☐ Yes   ☐ No

I have read and understand the information above. ☐ Yes   ☐ No

I want to participate in this research and continue with the coding ☐ Yes   ☐ No

[if the answer is no to any of the above questions, the individual cannot participate and should not be allowed to proceed to the next question.]