Given the data we have now, we plan to investigate three questions:

1. **Can response to crisis be used to predict political party?**

Political party is indicative of much more than just who a person will vote for. For example, we know that people’s neural reactions to “disgusting” (non-political) images can predict someone’s political party with 98% confidence [1], and that people from different political parties have different neural structures [2]. Specifically, conservative people have a more prominent right amygdala, which is associated with a heightened sense of fear and anger. Given these instances of political party becoming clear based on reaction to exposure to negative events, we plan to examine whether emotion manifests along party lines in response to a high-profile terrorist attacks. We hope to be able to identify party by reaction to the ISIS attacks, with the intuition that more conservative people will react more negatively to the terrorist attacks than more liberal people. To identify political party, we plan to look at user’s tweets about political candidates and use sentiment analysis to determine which candidate they support.

1. **Does crisis influence the way that people respond or does it just reinforce their preexisting views?**

For this question, we are trying to understand whether a person’s emotions are changed categorically or in magnitude by crises. We plan to make a sentiment “profile” for each user by using text analysis on their 20 most recent tweets. Once we have this profile, we plan to compare sentiment of any tweets made about ISIS around the time of the Paris attacks. The more that people’s tweets from that time period match their sentiment profile, the more it is evident that that the crisis has just confirmed or reinforced their views. However, if people’s sentiment changes, it is evidence that the crisis was able to change their minds.

1. **Does relative anonymity change the way that people respond?**

We plan to examine whether or not the forum in which users are posting influences the sentiment of their posts. Our suspicion is that the forum will influence what is posted since Reddit is relatively anonymous, whereas twitter is not. It could be the case that anonymity magnifies the sentiment profile that we find on twitter, or that people respond totally different in situations when their friends likely won’t be reading what they write. Any consistent difference would present an interesting window into how anonymity influences the way that people post.

[1] “Political Orientations Are Correlated with Brain Structure in Young Adults”; Kanai, Ryota et al.;

Current Biology , Volume 21 , Issue 8 , 677 – 680

[2] “Nonpolitical Images Evoke Neural Predictors of Political Ideology”; Ahn, Woo-Young et al.;

Current Biology , Volume 24 , Issue 22 , 2693 - 2699

**Methods**

The most interesting variable from our point of view is the text of the tweets and reddit posts. To make these into a useful metric, we plan to do text analysis using natural language processing techniques. We have not decided yet on a particular tool, but we plan to either classify the texts by emotion or on a scale of positivity to negativity.

We plan to use most our other variables to decide which tweets to use in our analysis. For example, if a user doesn’t have a profile picture, has no followers, or their tweets have few favorites or retweets, that would lend credence to the idea that they are not worth considering. If they are not, we will leave them out of further analyses.

Issues

Mismatched tags – our data has any tweets with the #is hashtag. That means we’re getting a lot of garbage from people who use the #is hashtag unrelated to the ISIS. For example, “#light #breakfast #on #diet #system #coffee #cup #is #must”

Reddit – Twitter User matching – At least so far, we are using the fairly naïve assumption that people share reddit and twitter usernames to match users. That is almost certainly wrong in some circumstances, and would be very difficult to figure out since we can’t know if the twitter users have reddit accounts under different names. It’s an even bigger problem when we match two users in this way who really aren’t the same person, because we wouldn’t know that we’ve matched the wrong person.

Overrepresentation of certain tweets – with a cursory look at the dataset, it’s clear that some tweets appear to be repeated very often while others are not. At this point, we’re not sure whether or not we want to be reusing the same tweets or if we only want to consider unique tweets.

Different languages – There are lots of tweets in the dataset that are in different languages. If we plan to use sentiment analysis on them, it might be the case that we can’t use the ones in foreign languages. In that case, we’d have to figure out some way of removing them.

Missing data – For the Twitter data, we encountered several hundred tweets with missing data entries, and have listed the number of missing entries for each attribute.