Thesis Workshop Proposal

Julian McClellan

Due 5/7/17

Persuasion

Persuasion, a goal in a number of settings, from political and marketing campaigns to friendly or professional conversations, has been the subject of significant research efforts (Tan et al. 2016; Dillard and Pfau 2002). Before the advent of social media websites like Facebook or Reddit, these research efforts were mostly confined to laboratory settings, but thanks to the increasing number of social interactions online, interpersonal persuasion has become observable on a massive scale (Fogg 2008).

Tan et al. explored persuasion in /r/changemyview (CMV), an area of the popular social media website Reddit (2016). CMV is particularly conducive to the study of mass interpersonal persuasion, as posters must state the reasoning behind their views, and successful arguments must be awarded with explicit confirmation. Thus, the outcome of the persuasion efforts, reasoning behind people's views, and the full interactions are accessible.

With access to this information, Tan et al. focused primarily on how interaction dynamics and choice of language within arguments were associated with a successful change in someone's opinion. A third focus of the study was an attempt to determine the malleability of an opinion, i.e. the likelihood that the holder of that opinion would award successful arguments to change it. Assuming that at least 10 unique challengers to the opinion were present, and that the holder of the opinion responded at least once, Tan et al. analyzed the way in which the opinion was presented and attempted to predict whether or not it could be changed.

This last task, attempting to determine the malleability of an opinion without respect to any of the arguments attempting to change it, was difficult indeed, and Tan et al. only achieved an ROC AUC of .54. Still, using weighted logistic regression, they found some significant features consistent with self-affirmation theory (Cohen, Aronson, and Steele 2000; Correll, Spencer, and Zanna 2004).

Self-Affirmation Theory

> "[S]elf-affirmation theory . . . suggests that every person strives for positive self-regard and, to achieve it, draws on successes in important domains in her or his life. These domains constitute aspects of individual identity, including important social roles, abilities, and beliefs. Because the individual depends on a constellation of domains for feelings of adequacy, a threat to one of the domains can prompt a defensive reaction. A crucial tenet of self-affirmation theory, though, is that the ultimate goal of a defensive reaction is the security of the global sense of self-worth, no the security of the domain, per se. The individual should defend a given domain only to the degree that the more general sense of self-worth is compromised by its loss."

(Correll, Spencer, and Zanna 2004)

In psychology, self-affirmation, which can be thought to reinforce one's global sense of self-worth, has been found to indicate open-mindedness and make beliefs more likely to yield (Correll, Spencer, and Zanna 2004; Cohen, Aronson, and Steele 2000).

Tan et al. found that within the text of an opinion, the use of first person pronouns were strong indicators of malleability, but first person plural pronouns correlated with resistance.

> "[I]ndividualizing one's relationship with a belief using the first person pronouns affirms the self, while first person plurals can indicate a dilluted sense of group responsibility for the view."

(Tan et al. 2016)

While Tan et al. attempted to derive the level of self-affirmation present within the stating of an opinion, the user stating that opinion can have other sources of self-affirmation. Returning to Correll:

> "[I]f global self-worth is temporarily bolstered by success in a second, unrelated domain, the individual should be more willing to tolerate a threat to the domain of interest."

Looking at the wording of the opinion itself is a related domain, but it is reasonable to assume that if a Redditor has previous submission and comment history, that some of that history is unrelated to the opinion they are presenting for change in CMV. Additionally, self-affirmation theory does not restrict the source of bolstering one's global self-worth; it can be self-affirmation or affirmation from third parties.

Thus, within past submissions and comments one can look at the same features as Tan et al., first person singular and plural pronouns for self-affirmation, but also for features that are indicative of third party affirmation, like the score, given by other users, of the comment or submission in question.

All of a Reddit user's past submission history is available for perusal, as well as 1000 of the most recent, top-scoring, "hottest", and most "controversial" comments. Exploring the affirming nature of this history allows a deeper testing of self-affirmation theory, as a lab experiment can only really test the history created within the lab settings itself. Cohen et. al (2000), for example. . .

> [A]sked half of their participants to write a paragraph about an important value (to affirm their sense of self-worth) before exposing them to arguments that challenged their views on capital punishment or abortion. Compared with control participants who wrote about less important values, those who wrote about a central value were more willing to recognize the strengths of the challenging argument.

Utilizing a Reddit user's past submission and comment history, on the other hand, not only allows a more extended look into instances in which self and third party affirmation may have occurred, but it also provides instances of participation with the community that the view is being exposed to. If a Reddit user posited their opinion to CMV, but also had previous submissions and comments within the CMV community, then that participation would be relevant, but previous participation in other subreddits can also be more or less relevant as well. CMV does not establish any barriers (besides following the subreddit rules), against first-time participation from other Reddit users new to CMV. If a user posts an opinion on CMV, but has also attempted to change an opinion on CMV, then it's possible that these attempts may or may not have changed the view (receiving or not receiving explicit recognition). Looking at past participation within CMV, a notion of reciprocity might be present; users who have had their comments awarded for changing other opinions might be more yielding with their own comments.

Tan et al. extended self-affirmation theory as it was onto an online debate space. Utilizing an opinion author's past interactions in CMV can possibly extend self-affirmation theory itself, by looking further back in time to potential affirmation, but also shed more light on the unique characterstics of Reddit and CMV in general

Research Design: Data and Methodology

From 2013 - 2016 there are over 27,000 submissions to /r/changemyview. As mentioned previously, for each unique author, one can access all of their previous submission history as well as their 1000 most recent, top-scoring, most controversial, and "hottest" comments. Assuming no overlap between these comment categories, up to 4000 comments can be retrieved, though this is not easily attainable. However, because the most recent 1000 comments can be retrieved, the period of commenting history with 100% accuracy can be identified.

With these data available, a number of features to track (self-)affirmation can be created. Affirmation could come in the form of upvotes (or downvotes), from the number of comments to a submission or another comment, following Tan et al., the use of first-person pronouns, and the sentiment of replies.

Features that might have less to do with affirmation can also be tracked, such as simple counts of comments and submissions over a given period, the number of words in each comment (as measures of activity), and the subreddits participated in.

With this data, binary classification can be used to predict the likelihood that a given submission in /r/changemyview contained a delta from the original poster, i.e., to predict whether the original poster change their view.

Given that there might be different affects for a user's first post versus their second post, and third post, and so on, if the data has enough repeat authors, the `nth` submissions on /r/changemyview from authors can be compared to see if the affects change as a Redditor gains more submission experience in posting to /r/changemyview. It would be interesting to see whether Redditors became more or less likely to change their minds. Perhaps, one might think, repeated posting is indicative of a more open mind and willingness to engage, thus more malleable, or perhaps the most malleable opinions appear in the first posts, and as users posts more, they run out of easily changed opinions and are more intractable.

Cohen, Geoffrey L, Joshua Aronson, and Claude M Steele. 2000. "When Beliefs Yield to Evidence: Reducing Biased Evaluation by Affirming the Self." *Personality and Social Psychology Bulletin* 26 (9). Sage Publications Sage CA: Thousand Oaks, CA: 1151–64.

Correll, Joshua, Steven J Spencer, and Mark P Zanna. 2004. "An Affirmed Self and an Open Mind: Self-Affirmation and Sensitivity to Argument Strength." *Journal of Experimental*

Social Psychology 40 (3). Elsevier: 350–56.

Dillard, James Price, and Michael Pfau. 2002. *The Persuasion Handbook: Developments in Theory and Practice.* Sage Publications.

Fogg, BJ. 2008. "Mass Interpersonal Persuasion: An Early View of a New Phenomenon." In *International Conference on Persuasive Technology*, 23–34. Springer.

Tan, Chenhao, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. "Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-Faith Online Discussions." In *Proceedings of the 25th International Conference on World Wide Web*, 613–24. International World Wide Web Conferences Steering Committee.