Quantifying Qualitative Data for Understanding Controversial Issues

Michael Wojatzki[†], Saif M. Mohammad[◇], Torsten Zesch[†], and Svetlana Kiritchenko[◇]
[†]Language Technology Lab, University of Duisburg-Essen
{michael.wojatzki,torsten.zesch}@uni-due.de
[◇]National Research Council Canada
{saif.mohammad,svetlana.kiritchenko}@nrc-cnrc.gc.ca

Abstract

Understanding public opinion on complex controversial issues such as '*Legalization of Marijuana*' and '*Gun Rights*' is of considerable importance for a number of objectives such as identifying the most divisive facets of the issue, developing a consensus, and making informed policy decisions. However, an individual's position on a controversial issue is often not just a binary support-or-oppose stance on the issue, but rather a conglomerate of nuanced opinions and beliefs on various aspects of the issue. These opinions and beliefs are often expressed qualitatively in free text in issue-focused surveys or on social media. However, quantifying vast amounts of qualitative information remains a significant challenge. The goal of this work is to provide a new approach for quantifying qualitative data for the understanding of controversial issues. First, we show how we can engage people directly through crowdsourcing to create a comprehensive dataset of assertions (claims, opinions, arguments, etc.) relevant to an issue. Next, the assertions are judged for agreement and strength of support or opposition, again by crowdsourcing. The collected Dataset of Nuanced Assertions on Controversial Issues (NAoCI dataset) consists of over 2,000 assertions on sixteen different controversial issues. It has over 100,000 judgments of whether people agree or disagree with the assertions, and of about 70,000 judgments indicating how strongly people support or oppose the assertions. This dataset allows for several useful analyses that help summarize public opinion. Across the sixteen issues, we find that when people judge a large set of assertions they often do not disagree with the individual assertions that the opposite side makes, but that they differently judge the relative importance of these assertions. We show how assertions that cause dissent or consensus can be identified by ranking the whole set of assertions based on the collected judgments. We also show how free-text assertions in social media can be analyzed in conjunction with the crowdsourced information to quantify and summarize public opinion on controversial issues.

Keywords: controversial issues, judging assertions, crowdsourcing, stance detection, argument mining, sentiment analysis

1. Introduction

Controversy is a state of sustained public debate on a topic or issue that evokes conflicting opinions, beliefs, claims, arguments, and points of view. In this paper, we will refer to utterances of opinion, belief, claim, argument, or point of view relevant to an issue as *assertions*. People make assertions on a controversy (or controversial issue) both in the physical world and on social media. Others might agree or disagree with these assertions. An individual's *position* on an issue is not simply a binary support-or-oppose stance on the issue, but rather a cumulative sum of many nuanced beliefs and opinions. Thus assertions are a useful means of capturing one's position on a controversial issue. Examples of common controversial issues include *the legalization of marijuana, government policy on refugees,* and *gun rights*. Examples of assertions include '*Marijuana alleviates the suffering of chronically ill patients*' and '*Expanding legal use of marijuana makes illegal use easier*'.

Controversial issues are usually complex, not only because of the many diverse and inter-related assertions they evoke, but also because they often subsume many inter-related sub-issues (e.g., '*Should marijuana be legalized for medical purposes?*'). An issue has many *stake holders*—people that are directly or indirectly affected by it—e.g., patients undergoing chemotherapy and parents of teenage children are just two of the many stakeholders affected by a possible legalization of marijuana. Given these complexities, it is difficult to attain consensus, and people often tend to talk past each other without really listening to the merits of opposing arguments (amplifying existing echo chambers). Even decision making bodies, such as local and national governments, can make more informed choices if they have a comprehensive understanding of the controversial issue. Simply obtaining the percentage of people that support or oppose an issue is not sufficient. It is often more useful to obtain information about the different aspects of an issue, what aspects of an issue are considered more important, what people's beliefs and opinions on various aspects are, who the main stakeholders are, what opposing groups agree and disagree on, etc.

A common approach to understanding complex controversial issues is to hire experts and conduct surveys. However, such an approach has inherent limitations: the survey creators inadvertently bring in biases, often these surveys fail to cover all relevant aspects, and the process is time intensive and expensive. Further, such surveys often only make use of questions whose responses can be easily aggregated, e.g., multiple-choice questions and questions with a small set of possible responses. Quantifying qualitative responses often requires manual interpretation and does not scale up to large surveys. In contrast, people naturally express their nuanced positions on an issue through free-text utterances in the real world and in on-line social networks. Capturing and quantifying information in free-text assertions remains a significant challenge in understanding controversial issues.

This work has two broad goals. First, we propose a method to obtain and quantify qualitative information relevant to a controversial issue by **engaging people directly via crowdsourcing**. Specifically, we create a comprehensive dataset of nuanced assertions relevant to sixteen controversial issues in the United States. Next, we rank the assertions by both the degrees of agreement on them and by how strongly people support or oppose each of the assertions. We create this dataset by conducting crowdsourced surveys to: (1) collect about 150 unique assertions per issue, (2) determine agreement on these assertions by hundreds of respondents, and (3) rank the assertions based on how strongly people support or oppose them. For each of the sixteen issues we collect around 150 assertions (2,243 in total). For each of the assertions, we obtain 15 judgments on how strongly people support or oppose the assertions (67,290 in total) and about 45 judgments for agreement (101,133 in total). We will refer to this dataset as the dataset of *Nuanced Assertions on Controversial Issues (NAoCI)*.

We propose several metrics that can be calculated from the data and used for grouping, ranking, and clustering assertions and participants. We show how these metrics can be used to identify agreement and support on assertions, to rank issues based on controversial assertions, and to determine the similarity of assertions, users and groups. An analysis of the distribution of judgments shows that for all of the controversial issues there are more assertions with which the majority of the people agree than assertions with which the majority of the people disagree or assertions with which some agree and some do not (controversial assertions). However, the controversial assertions are often the ones that are supported or opposed to the greatest degree. The new approach to understanding argumentation proposed here goes well beyond simple positive–negative–neutral classification or overall stance detection from text. The NAoCI dataset will help foster new research that tackles difficult questions such as how people make arguments to support their stance on an issue, what is the distribution of assertions that people agree and disagree with across different groups, and how the position on a controversial assertion impacts overall stance.

Our second goal is to improve the **understanding of controversial issues using assertions made by people in social media**. As a first step towards achieving this goal, we propose several new natural language processing tasks in this paper. These tasks include identifying assertions implicit in free-text posts on social media, determining a speaker's position on various assertions, identifying the degree of agreement, support, and opposition for assertions by a large population of tweeters that post messages about a controversial issue, etc. These new natural language processing tasks are a way to summarize information about the controversial issues without necessarily having to do the crowdsourcing described above. However, the crowdsourced data will serve as a source of reference (gold) labels for the evaluation of these NLP algorithms.

All the annotation tasks described in this paper were approved by the National Research Council Canada's Institutional Review Board, which reviewed the proposed methods to ensure that they were ethical. All our data and crowdsourcing questionnaires are made available on the project webpage.[1]

[1]https://sites.google.com/view/you-on-issues

2. Related Work

In recent years, a number of web-based applications have been developed that help users share their opinions and explore public opinion on controversial issues. Specifically, Voting Advice Applications such as *votecompass.com*, *politicalcompass.org*, and *isidewith.com* ask visitors whether they agree or disagree with a set of pre-chosen assertions (Garzia and Marschall, 2012). The applications usually provide some form of visualization that depict which political party or candidate is closest to the user's own position. Voting Advice Applications have mainly been studied in political science, where researchers have examined questions such as: whether these applications have an effect on voting behavior (Ladner and Pianzola, 2010), what characteristics their users have (Wall et al., 2009), or how their design affects their outcome (Louwerse and Rosema, 2014). In contrast, here we focus on creating a language resource containing crowdsourced judgments on a large number of user generated assertions.

Our second goal of understanding and summarizing public opinion from posts on social media is related to work on detecting sentiment (Pang and Lee, 2008; Liu, 2012; Mohammad, 2016; Mohammad et al., 2018) and stance (Mohammad et al., 2016; Xu et al., 2016; Taulé et al., 2017), argumentation mining (Kwon et al., 2007; Walker et al., 2012; Rosenthal and McKeown, 2012; Stab and Gurevych, 2014; Peldszus and Stede, 2016; Habernal and Gurevych, 2016), and framing (Entman, 1993; Card et al., 2015; Tsur et al., 2015; Fulgoni et al., 2016; Johnson and Goldwasser, 2016). These approaches focus on identifying sentiment, stance, claims, premises, reasons, arguments, sentiment, etc. from individual utterances. In contrast, here we suggest quantifying information from a large number of social media utterances in order to gain the overall understanding of a complex issue. There exists work on opinion summarization (Hu and Liu, 2004; Zhuang et al., 2006; Titov and McDonald, 2008; Titov and McDonald, 2008; Ganesan et al., 2010; Gerani et al., 2014) and on the clustering of argumentative elements (Trabelsi and Zaıane, 2014; Misra et al., 2015; Boltužić and Šnajder, 2015; Barker and Gaizauskas, 2016); however, these studies mainly explore the grouping of utterances into positive and negative clusters, or extracting opinion utterances and claims to create a summary. Here, we first establish a comprehensive representation of assertions for a controversial issue, and then propose discovering elements of this representation from free-text utterances in social media.

3. Understanding Controversial Issues

Quantifiable and useful insights on a controversial issue can be obtained by having a large number of people vote on a large number of relevant assertions. As the manual creation of assertions is time-consuming, subject to personal bias, and potentially incomplete, we here rely on crowdsourcing to generate the assertions. In the subsections below, we describe how we: (1) engage people directly through crowdsourcing to obtain assertions and judgments on these assertions (Figure 1 gives the overview for an example issue), and (2) analyze and summarize the crowdsourced information through a number of ways

Issue	# of Assertions	# of Agreement Judgments	# of Support–Oppose Judgments
Black Lives Matter	135	6,154	4,050
Climate Change	142	6,473	4,260
Creationism in school	129	5,747	3,870
Foreign Aid	150	6,866	4,500
Gender Equality	130	5,969	3,900
Gun Rights	145	6,423	4,350
Marijuana	138	6,200	4,140
Same-sex Marriage	148	6,899	4,440
Mandatory Vaccination	134	5,962	4,020
Media Bias	133	5,877	3,990
Obama Care	154	6,940	4,620
US Electoral System	175	7,695	5,250
US in the Middle East	138	6,280	4,140
US Immigration	130	5,950	3,900
Vegetarianism & Veganism	128	5,806	3,840
War on Terrorism	134	5,892	4,020
Total	2,243	101,133	67,290

Table 1: Issues, number of generated assertions, and number of collected judgments.

including: ranking assertions based on how they were judged, clustering people and assertions by voting patterns, and ranking issues by degree of polarization. The sixteen controversial issues we explore are shown in Table 1. They were compiled from voting advice websites.

3.1. Collecting Public Opinion

Generating Assertions: Given an issue (name and a brief description), we asked each participant to come up with five assertions relevant to the issue. To guide the process of creating assertions, the participants were given the following directions. Participants had to formulate assertions in a way that a third person can agree or disagree with it. The assertions had to be self-contained and understandable. Hedged statements that included words such as '*maybe*', '*perhaps*', or '*possibly*' were prohibited.

As a quality control measure, participants were also required to answer test questions where they had to indicate whether the given assertions were applicable to the issue and whether they were in accordance with the rules. We discarded all responses from participants who incorrectly answered more than 10% of these questions.

We obtained 2,243 assertions from 69 participants for the sixteen issues. Duplicates and instances not in accordance with the guidelines were removed. Table 1 lists the number of remaining valid assertions for each issue.

Quantifying Agreement and Strength of Support and Opposition: Once the list of assertions was compiled, we obtained judgments on (a) whether people agree with the assertion, and (b) how strongly they support or oppose the assertion. A participant may not be inclined to judge all assertions. However, if a large enough number of judgments are obtained from many participants, then meaningful inferences can be drawn. Thus, individual participants were free to judge as many assertions as they wished. For both kinds of judgments, the exact questionnaires along with the directions and examples are shown in the Appendix.

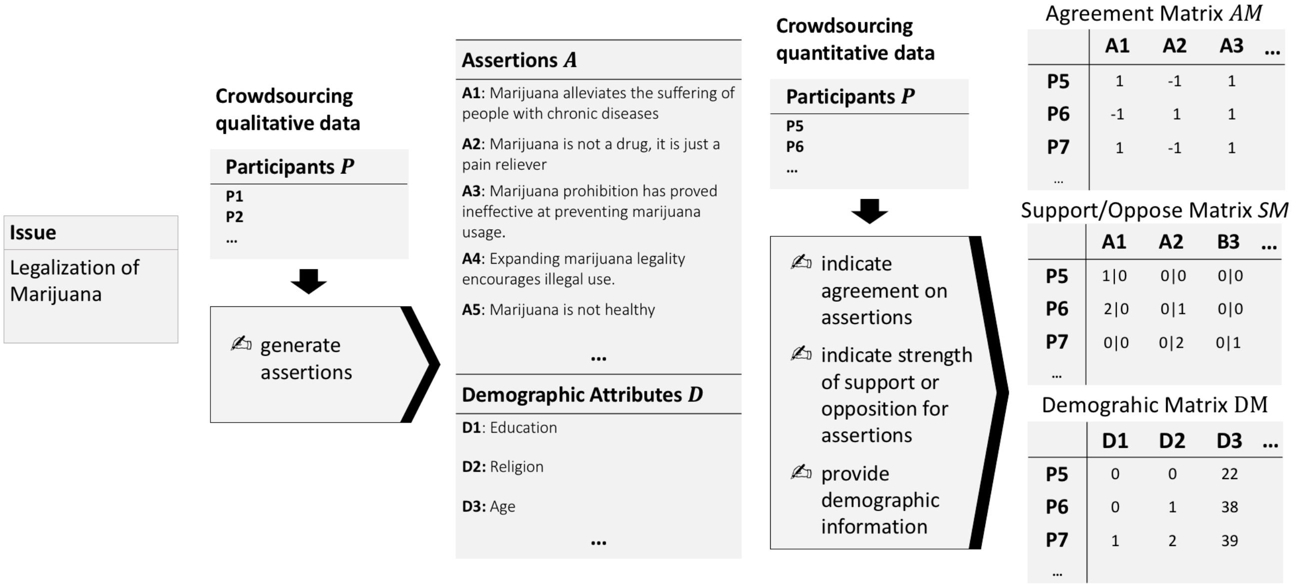
To obtain agreement judgments on the assertions, we simply asked the subjects to indicate to us whether they agree or disagree with the collected assertions. The average number of agreement judgments per assertion was 45. Table 1 shows how many judgments we collected for each issue. From the agreement judgments, we created the agreement matrix AM, which contains one column per assertion, and one row per participant. Each cell $ad_{p,a}$ in this matrix has the judgment provided by participant p for assertion a. Consequently, $\vec{ad_a}$ is the vector of all judgments provided for assertion a, and $\vec{ad_p}$ is the vector of all judgments provided by participant p.

To obtain consistent and comparable fine-grained scores indicating the degree of support or opposition from multiple respondents, we used a technique known as Best–Worst Scaling (BWS) (Louviere et al., 2015b). BWS is an annotation scheme that addresses the limitations of traditional rating scale methods, such as inter- and intra-annotator inconsistency, by employing comparative annotations (Louviere, 1991; Louviere et al., 2015a; Kiritchenko and Mohammad, 2016; Kiritchenko and Mohammad, 2017). Annotators are given n items (an n-tuple, where $n > 1$ and commonly $n = 4$). They are asked which item is the *best* (highest in terms of the property of interest) and which is the *worst* (lowest in terms of the property of interest). When working on 4-tuples, best–worst annotations are particularly efficient because each best and worst annotation will reveal the order of five of the six item pairs. For example, for a 4-tuple with items A, B, C, and D, if A is the best, and D is the worst, then A $>$ B, A $>$ C, A $>$ D, B $>$ D, and C $>$ D. Real-valued scores of association between the items and the property of interest can be calculated from the BWS annotations (Orme, 2009; Flynn and Marley, 2014).

We generated 4,486 4-tuples of assertions from our list of 2,243 assertions using the code provided by Kiritchenko and Mohammad (2016). Participants were presented with four assertions at a time and asked two questions:

1. which of the assertions they support the most (or oppose the least),
2. which of the assertions they oppose the most (or support the least).

We obtained judgments from fifteen people for every 4-tuple. From the comparative judgments, we created the support–oppose matrix SM, that consists of one row per participant, and one column per assertion. In each cell $bd_{p,a}$ we store a tuple that indicates how often participant p has selected assertion a as the one they support the most (or oppose the least), and how often participant p has selected assertion a as the one they oppose the most (or support the least).



Agreement Matrix AM	A1	A2	A3	…
P5	1	-1	1	
P6	-1	1	1	
P7	1	-1	1	
…				

Support/Oppose Matrix SM	A1	A2	B3	…
P5	1|0	0|0	0|0	
P6	2|0	0|1	0|0	
P7	0|0	0|2	0|1	
…				

Demograhic Matrix DM	D1	D2	D3	…
P5	0	0	22	
P6	0	1	38	
P7	1	2	39	
…				

Figure 1: Quantifying qualitative assertions for the issue '*Legalization of Marijuana*'. The agreement matrix has values 1 for agreement and -1 for disagreement. The support/oppose matrix contains integer values indicating how many times the participants have selected an assertion as the one they most support (resp. least oppose) and most oppose (resp. least support). The demographic matrix contains the responses to the demographic questions such as education, religion, and age.

Demographic Data: A key determiner of one's beliefs and opinions is their personal experience, which in turn is often shaped by their demographic attributes. In order to determine the extent to which demographic attributes correlate with one's judgments on assertions, we asked the participants to provide us with their demographic information: age, gender, political affiliation, education, family status, profession, race, religion, whether the participants have ties to overseas and whether they are US citizens. Participants were free to not provide this demographic data if they wished.

Participants: Of the 230 subjects that participated in the quantitative phase, 85 (37%) submitted demographic information. This sample had a mean age of 34.9 years; 65% of the respondents were female. Just over 50% of the respondents had a bachelors or higher degree. With regard to political affiliation, a broad mix could be observed. However, the group of people who identified themselves as democrats was the largest. About 50% of respondents indicated that they were employed. Most of the participants had ties to overseas (68%) and were US citizens (89%). With respect to religion, the majority of the participants were Christians (Roman Catholic 28%, Protestant 25%, Russian Orthodox 1%). About 29% were atheists. 69% of our participants identified themselves as white, 10% as Hispanic, 8% as black or African American, 8% as Asian, 2% as American Indian or Alaskan native, and 3% as others. Further details of the demographic information can be found in the Appendix.

3.2. Summarizing Public Opinion

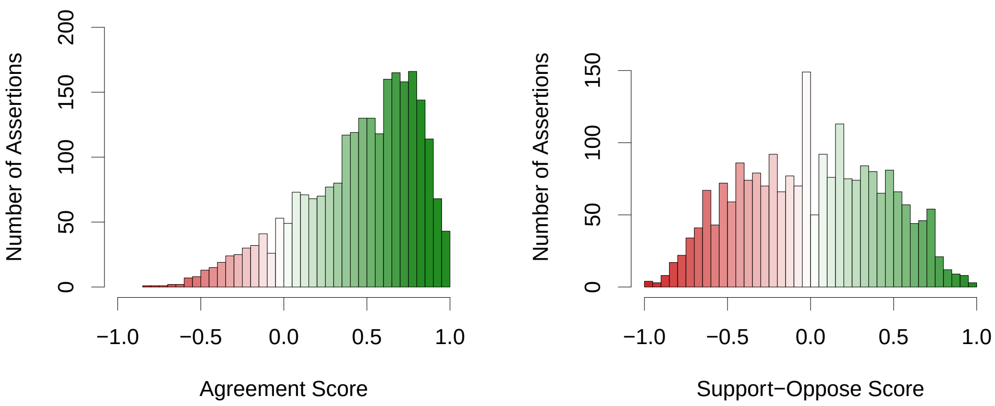
The data gathered from crowdsourcing efforts can be used to summarize, visualize, and gain insights into public opinion on controversial issues. In the subsections below, we describe how we calculate various metrics from the data that can be used to summarize various aspects of public opinion on an issue.

3.2.1. Ranking Agreement and Support

Assertions expressed by participants provide key insights into why an issue is controversial, what aspects of the issue people are particularly passionate about, etc. Thus organizing the assertions by amount of agreement (to quickly view the assertions with most and least agreement) and strength of support or opposition is particularly useful. We calculate the *agreement score* (ags) of an assertion a by simply subtracting the percentage of times the assertion was disagreed with from the percentage of times the assertion was agreed with:

$$ags(a) = \% \text{ agree}(a) - \% \text{ disagree}(a) \tag{1}$$

The agreement score can be used to rank assertions from **least agreement** (-1) to **most agreement** (1). A score of 0 indicates that an equal number of participants agree and disagree with the assertion and that the assertion is therefore highly **controversial**. We can identify the most controversial assertions by sorting assertions by the absolute value of the agreement scores and selecting those assertions which have the lowest absolute scores. These agreement scores can be used to better understand the debate on an issue. For instance, for the issue *legalization of same-sex marriage*, the assertion '*Love is a right for everyone.*' has the highest agreement score, '*Saying that gay people should get married is like saying that a brother can marry his sister both are at higher risk of disease.*' has the lowest agreement score, and '*Allowing same-sex marriage will create a slippery slope where people will begin to fight for other alter-*



(a) Distribution of Agreement Scores

(b) Distribution of Support–Oppose Scores

Figure 2: Distribution of agreement and support–oppose scores across all issues. We group the agreement and support–oppose scores into bins of size 0.05. For both the agreement and support–oppose scores, the colors encode how positive (green) or negative (red) the scores are.

native marriages such as polygamy.' is the most controversial. In the Appendix, we show the top three assertions according to these rankings for all issues.

We transform the comparative BWS judgments for support and opposition into scores ranging from -1 (maximum opposition) to 1 (maximum support) through a simple counting method proposed by Orme (2009). For an assertion a we calculate a *support–oppose score* (sos) by subtracting the percentage of times an assertion was chosen as the most opposed from the percentage of times the assertion was chosen as the most supported:

$$sos(a) = \%\,\text{most support}(a) - \%\,\text{most opposed}(a) \quad (2)$$

These scores can be used to rank assertions from most strongly supported (1) to most strongly opposed (-1). Selecting an assertion as 'most opposed' in the comparative annotations, may mean that one either most opposes or least supports the assertion. We can infer which of the two interpretations applies from the agreement judgments; i.e. 'most oppose' can be considered as 'least support' if a participant agrees to a statement and it remains 'most oppose' if the person disagrees with the assertion. Analogously, 'most support' can be interpreted as 'least oppose' if one disagrees to the assertion and it can be interpreted as 'most support' if one agrees to the assertion. Thus, we additionally calculate a *support score* (ranging from 'least supported' (0) to 'most supported' (1)) and an *oppose score* (ranging from 'least opposed'(0) to 'most opposed' (1)). To calculate these scores, we reuse the formula shown in equation 2. However, the percentages are now calculated only on the set of persons that have agreed to (*support score*) or disagreed with (*oppose score*) the assertion. These scores can be used to differentiate between assertions where a *support–oppose score* of about zero indicates that an assertion is both strongly supported and strongly opposed, and assertions which have a *support–oppose score* of about zero but that are rarely strongly supported or opposed. An example for the former is the assertion '*Freedom of the press prevents the government or other third parties from controlling the media.*' that has a *support–oppose score* of 0, but a fairly high *support score* of 0.71, and a high *oppose score* of 1. An example for the latter is the assertion '*Women's rights have well found legal basis.*' that has a *support–oppose score* of -0.07 and both a low *oppose score* (0) and *support score* (0.39).

Figure 2 shows histograms of the agreement and support–oppose scores. For the support–oppose scores, we observe that the scores have a normal distribution. This distribution can also be found when looking at the distributions per issue. For the agreement scores, we observe that the mass of the distribution is concentrated in the positive range of possible values, indicating that the participants tend to agree with the assertions more often than they disagree. (We observe a similar distribution in each of the individual issues as well.) This is consistent with the hypothesis that people often do not disagree with the individual assertions that the other side makes, however, they might disagree on the relative importance of that assertion among the various other assertions in terms of reaching an overall stance on the broader issue.

Examples of assertions with a particular high *agreement score* for the *Legalization of Marijuana* issue are '*Drug abuse can kill.*', '*Cannabis has legitimate medical effects.*', and '*Exceptions should be considered allow people with medical issues that will benefit from legalization of Marijuana.*'.

3.2.2. Ranking Controversial Issues

Governments and policy makers often have to work with not one but several controversial issues. Thus it is useful to know which issues are particularly polarizing so as to prioritize those issues or to allocate appropriate resources to

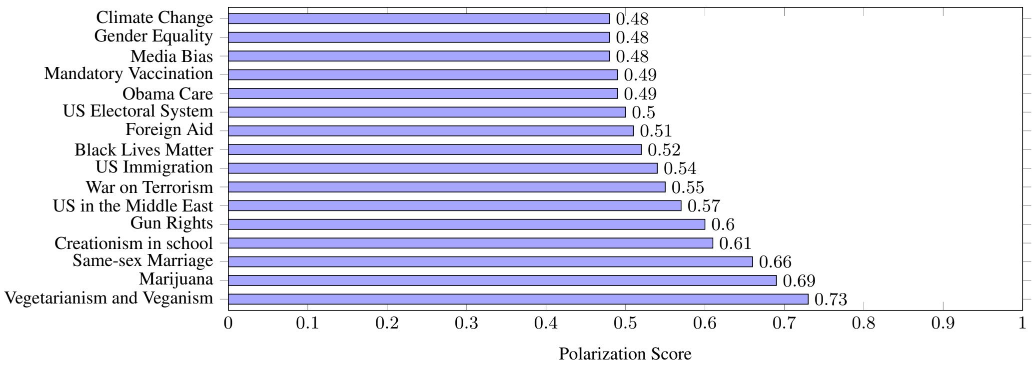


Figure 3: Issues ranked according to their polarization score.

them. One indicator of the degree of polarization is the extent to which the assertions associated with the issue evoke opposing responses. If all the assertions for an issue have an agreement score of zero (the number of respondents that agree is equal to the number of respondents that disagree), then the issue is maximally polarizing. Hence, we calculate the *polarization score* (ps) of an issue (a set of assertions) I by first calculating the average of the absolute value of the agreement score for each of the assertions, and then subtracting this value from one:

$$ps(I) = 1 - \frac{1}{|I|} \sum_{a \in I} |ags(a)| \tag{3}$$

A polarization score of 0 indicates that participants consistently agree or disagree with all assertions representing an issue. A polarization score of 1 indicates that – on average – an equal number of participants agree and disagree with the assertions on an issue. Consequently, a polarization score of 0.5 describes an issue in which more and less polarizing assertions keep a balance.

The polarization scores for the sixteen issues are shown in Figure 3. Interestingly, many of the polarization scores for the issues are around 0.5. For the issues *Climate Change*, *Gender Equality*, *Media Bias*, *Mandatory Vaccination*, and *Obama Care* the scores are even below 0.5, which means that on average there is more consensus than dissent in judging the assertions on these issues. However, as shown by the issues *Same-sex Marriage* (0.66), *Marijuana* (0.69) and *Vegetarianism & Veganism* (0.73), our data contains also more polarizing issues. In future work, we plan to examine whether linguistic properties of the assertions (e.g., whether they use superlatives) can be utilized to explain these differences. Also note that the score does not include external factors such as the social context in which a controversy takes place and should be used only as one (of the many possible ways) in which polarization can be measured.

3.2.3. Determining Similarity of Users and Assertions

The crowdsourced data can be used to determine which users show a similar response behavior and which assertions have been similarly voted on. Voting similarity between participants can be used to generate guesses about their judgments on assertions for which they have not voted. We determine the voting similarity between pairs of participants by computing the cosine of the vectors that represent the rows in the agreement matrix AM (see Figure 1):

$$cos(p1, p2) = \frac{\vec{ad}_{p1} \cdot \vec{ad}_{p2}}{|\vec{ad}_{p1}| \cdot |\vec{ad}_{p2}|} \tag{4}$$

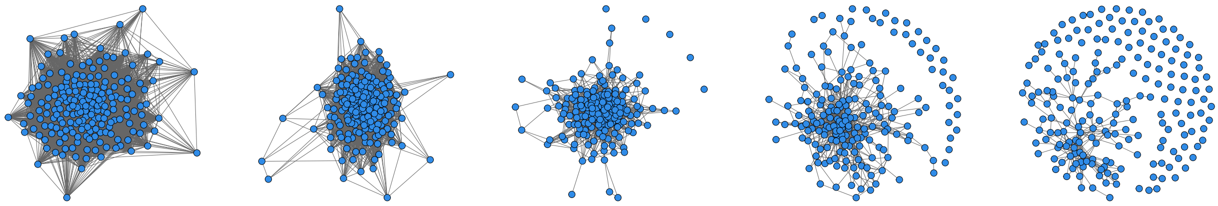
Being able to judge similarity between assertions helps identify inter-related assertions. We determine the degree by which two assertions are judged similarly by computing the cosine of the column vectors of the agreement matrix AM (see Figure 1):

$$cos(a1, a2) = \frac{\vec{ad}_{a1} \cdot \vec{ad}_{a2}}{|\vec{ad}_{a1}| \cdot |\vec{ad}_{a2}|} \tag{5}$$

The computed similarities between pairs of assertions are made public on the project's website. We manually inspected pairs of assertions with a particularly high or low judgment similarity. We found several reasons for high similarity between assertions (why people tend to agree with both assertions or they tend to disagree with both assertions): the two assertions are close paraphrases (e.g., '*Owning a gun can deter criminals.*' and '*Gun ownership deters crime.*'), one assertion entails the other (e.g., '*Oceans rise due to climate change.*' and '*Climate change has a big effect on the Earth.*'), underlying socio-cultural and political factors cause people to vote similarly on two (sometimes seemingly unrelated) assertions (e.g., '*US should hire skilled immigrants.*' and '*Inclusion should be facilitated for immigrants.*').

Reasons for low similarity between assertions (why people tend to agree with one assertion and disagree with the other) include: the two assertions are contradictory or contrasting (e.g., '*It is safe to use vaccines.*' and '*Vaccines cause autism.*'), and underlying socio-cultural and political factors cause people to vote dissimilarly on two (sometimes seemingly unrelated) assertions (e.g., '*Congress should immediately fund Trump's wall.*' and '*All immigrants should have the right to vote in the American elections.*').

To further explore the relation between judgment similarity and semantic textual similarity (Agirre et al., 2012), we compute the textual overlap between assertions using the Jaccard index (Lyon et al., 2001) and examine the agreement scores of textually similar assertions. We observe that assertions with high text similarity often have very similar



(a) cosine > 0.0 (b) cosine > 0.1 (c) cosine > 0.2 (d) cosine > 0.3 (e) cosine > 0.4

Figure 4: Similarity of participants visualized in an undirected graph for the issue *Black Lives Matter*. In the sub figures, we draw edges between two persons if their voting similarity is above a certain threshold.

agreement scores. An example for this case are the following assertions:

> '*Women should have the same rights than men.*' ($ags = 0.84$)
> '*Women should have the same rights as men.*' ($ags = 0.88$)
> '*Women should have the same right as men.*' ($ags = 0.86$)

Note that the first and third example are ungrammatical, which does not seem to significantly affect the rating.

However, there are also pairs with a high textual overlap and significantly different agreement scores. In these pairs, slight differences in wording have a marked influence on how the assertions are judged. Examples include:

> '*Foreign aid budget should be more effective.*' ($ags = 0.73$)
> '*The foreign aid budget should be made more effective.*' ($ags = 0.65$)

as well as

> '*Climate change is costing lives.*' ($ags = 0.76$)
> '*Climate change is already costing lives.*' ($ags = 0.52$).

3.2.4. Clustering Participants With Similar Positions

It is often useful to identify which groups of people have similar positions on a controversial issue. This allows for focused engagement with individual groups. The agreement judgments in our data can be used to cluster participants according to their judgments on assertions.

We use the voting similarity between participants (c.f. equation 4) to find groups of people with similar overall beliefs on an issue. The same methods described in section 3.2.1. to calculate the agreement and support–oppose scores for an assertion can be used to calculate agreement and support–oppose scores for individual groups (instead of for the whole population). Then separate ranked lists of assertions can be generated for each group. This ranking can then be used to summarize the judgments of the groups. We can also determine which assertions a group agrees the most with, which assertions receive similar judgments across two groups, and which assertions the two groups disagree on.

When we cluster participants by judgment similarity, several scenarios are possible: a binary split into persons that support and oppose the overall issue (e.g., one cluster includes people in favour of legalizing marijuana and one cluster includes people that are against), several clusters that correspond to persons with more specific positions (e.g., being against legalizing marijuana for medical purposes or favoring the idea that marijuana is a gateway drug), or single cluster which expresses a mainstream of positions.

To examine the distribution of similar participants, we create an undirected graph in which the nodes represent participants and edges are drawn if the person–person similarity exceeds a certain threshold. Next, we compare graphs with different thresholds. We find that it is not uncommon for an issue that at a low threshold almost all participants are connected; i.e. that all subjects have a certain similarity to each other. If we increase the threshold, we do not observe the formation of several clusters, but of a single cluster and an increasing amount of single disconnected outliers. This indicates that the majority of persons in our data belongs to a mainstream. We visualize this experiment for the issue *Black Lives Matter* in Figure 4. We manually inspect the judgments of disconnected persons and observe that these indeed tend to have rather radical positions (e.g., disagreeing with the assertion '*Everyone is equal.*'). Note that clustering using cosine similarity is just one of the several ways to identify groups of similar positions. In future work, we plan to examine other ways of determining similarity between vectors (e.g., by considering their overlap).

4. Understanding Controversial Issues from Assertions in Social Media

The second goal of this work is to summarize information about controversial issues without necessarily being dependent on the described crowdsourcing. We propose to make use of the abundance of opinions and beliefs that are expressed on social media, especially on controversial issues. We now briefly outline a number of NLP tasks that can be developed for the understanding of controversial issues by identifying and analyzing assertions made in social media. Notably, the dataset described in the previous section can be used as reference (gold) to evaluate these automatic methods.

- Identifying explicit and implicit assertions relevant to an issue in social media posts (e.g. tweets).
- Identifying social media posts that express the same assertions in different ways; Identifying posts that express contradictory or opposing assertions.
- Compiling a large list of assertions relevant to an issue from tweets; ranking them by an estimate of the degree of agreement; ranking them by an estimate of the support and oppose scores; automatically estimating polarization scores for issues.
- Determining semantic similarity between pairs of assertions; Clustering tweeters by the similarity of the assertions they make (agree with); Clustering assertions by their similarity.

As a first step, we have begun collecting tweets pertinent to each of the sixteen issues (that are part of this project) through a small manually identified list of query terms.[2] We have collected close to nine million tweets. The IDs of these tweets are available on the project webpage.

5. Conclusion

We proposed a method to quantify qualitative information relevant to a controversial issue by engaging people directly via crowdsourcing. This new approach to understanding argumentation goes well beyond simple positive–negative–neutral classification or overall stance detection from text. We created a dataset containing a comprehensive and nuanced list of assertions relevant to an issue, and then ranked them by both the agreement and degree of support for each assertion. We applied the proposed method for sixteen different controversial issues and collected a total of over 2,000 assertions. We obtained over 100,000 judgments of whether people agree or disagree with the assertions and about 70,000 judgments indicating how strongly people support or oppose the assertions. We also proposed several metrics such as an agreement score for an assertion and a polarization score for an issue that can be calculated from the data and used for grouping, ranking, and clustering issues, assertions and participants. Finally, we outlined a number of NLP tasks for the understanding of controversial issues through assertions made by people in social media. Our dataset, Nuanced Assertions on Controversial Issues (NAoCI), can be used as a source of reference labels in the evaluations of these tasks. We expect the NAoCI dataset and automatic algorithms produced as part of this project to be helpful—both to the lay person and expert—in making informed decisions about complex controversial issues.

6. Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant No. GRK 2167, Research Training Group "User-Centred Social Media". We would also like to thank Tara Small for the helpful discussions.

[2]The full list of query terms is available on the project webpage: https://sites.google.com/view/you-on-issues.

7. Bibliographical References

Agirre, E., Diab, M., Cer, D., and Gonzalez-Agirre, A. (2012). Semeval-2012 task 6: A pilot on semantic textual similarity. In *Proceedings of the Sixth International Workshop on Semantic Evaluation*, pages 385–393.

Barker, E. and Gaizauskas, R. (2016). Summarizing multi-party argumentative conversations in reader comment on news. In *Proceedings of the 3rd Workshop on Argument Mining*, pages 12–20, Berlin, Germany.

Boltužić, F. and Šnajder, J. (2015). Identifying Prominent Arguments in Online Debates Using Semantic Textual Similarity. In *Proceedings of the 2nd Workshop on Argumentation Mining*, pages 110–115, Denver, USA. Association for Computational Linguistics.

Card, D., Boydstun, A. E., Gross, J. H., Resnik, P., and Smith, N. A. (2015). The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, pages 438–444, Beijing, China.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4):51–58.

Flynn, T. N. and Marley, A. A. J. (2014). Best-worst scaling: theory and methods. *Handbook of Choice Modelling*, pages 178–201.

Fulgoni, D., Carpenter, J., Ungar, L. H., and Preotiuc-Pietro, D. (2016). An Empirical Exploration of Moral Foundations Theory in Partisan News Sources. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 3730–3736, Portorož, Slovenia.

Ganesan, K., Zhai, C., and Han, J. (2010). Opinosis: A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 340–348, Beijing, China.

Garzia, D. and Marschall, S. (2012). Voting advice applications under review: the state of research. *International Journal of Electronic Governance*, 5(3-4):203–222.

Gerani, S., Mehdad, Y., Carenini, G., Ng, R. T., and Nejat, B. (2014). Abstractive Summarization of Product Reviews Using Discourse Structure. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1602–1613, Doha, Qatar.

Habernal, I. and Gurevych, I. (2016). Which argument is more convincing? analyzing and predicting convincingness of web arguments using bidirectional lstm. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 1589–1599.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177, Seattle, USA.

Johnson, K. and Goldwasser, D. (2016). Identifying stance by analyzing political discourse on twitter. *Proceedings of the First Workshop on NLP and Computational Social Science 2016*, pages 66–75.

Kiritchenko, S. and Mohammad, S. M. (2016). Capturing reliable fine-grained sentiment associations by crowdsourcing and best–worst scaling. In *Proceedings of The 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, San Diego, California.

Kiritchenko, S. and Mohammad, S. M. (2017). Best-worst scaling more reliable than rating scales: A case study on sentiment intensity annotation. In *Proceedings of The Annual Meeting of the Association for Computational Linguistics (ACL)*, Vancouver, Canada.

Kwon, N., Zhou, L., Hovy, E., and Shulman, S. W. (2007). Identifying and classifying subjective claims. In *Proceedings of the 8th annual international conference on Digital government research: bridging disciplines & domains*, pages 76–81. Digital Government Society of North America.

Ladner, A. and Pianzola, J. (2010). Do voting advice applications have an effect on electoral participation and voter turnout? Evidence from the 2007 Swiss Federal Elections. *Electronic participation*, pages 211–224.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.

Louviere, J. J., Flynn, T. N., and Marley, A. A. J. (2015a). *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge University Press.

Louviere, J. J., Flynn, T. N., and Marley, A. A. (2015b). *Best-worst scaling: Theory, methods and applications*. Cambridge University Press.

Louviere, J. J. (1991). Best-worst scaling: A model for the largest difference judgments. Working Paper.

Louwerse, T. and Rosema, M. (2014). The design effects of voting advice applications: Comparing methods of calculating matches. *Acta politica*, 49(3):286–312.

Lyon, C., Malcolm, J., and Dickerson, B. (2001). Detecting short passages of similar text in large document collections. In *Proceedings of the 6th Conference on Empirical Methods in Natural Language Processing (EMNLP 2001)*, pages 118–125, Pittsburgh, USA.

Misra, A., Anand, P., Tree, J., and Walker, M. (2015). Using summarization to discover argument facets in online idealogical dialog. In *Proceedings of the NAACL HLT*, pages 430–440, Denver, USA.

Mohammad, S. M., Kiritchenko, S., Sobhani, P., Zhu, X., and Cherry, C. (2016). SemEval-2016 Task 6: Detecting Stance in Tweets. In *Proceedings of the International Workshop on Semantic Evaluation 2016*, San Diego, USA.

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

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. *Emotion Measurement*, pages 201–238.

Orme, B. (2009). Maxdiff analysis: Simple counting, individual-level logit, and HB. *Sawtooth Software*.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135.

Peldszus, A. and Stede, M. (2016). An annotated corpus of argumentative microtexts. In *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation, Lisbon 2015 / Vol. 2*, pages 801–815, London, UK. College Publications.

Rosenthal, S. and McKeown, K. (2012). Detecting opinionated claims in online discussions. In *Semantic Computing (ICSC), 2012 IEEE Sixth International Conference on*, pages 30–37. IEEE.

Stab, C. and Gurevych, I. (2014). Identifying argumentative discourse structures in persuasive essays. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 46–56. Association for Computational Linguistics.

Taulé, M., Martı, M. A., Rangel, F., Rosso, P., Bosco, C., and Patti, V. (2017). Overview of the task of stance and gender detection in tweets on catalan independence at ibereval 2017. In *Notebook Papers of 2nd SEPLN Workshop on Evaluation of Human Language Technologies for Iberian Languages (IBEREVAL), Murcia, Spain, September*, volume 19.

Titov, I. and McDonald, R. (2008). A Joint Model of Text and Aspect Ratings for Sentiment Summarization. In *Proceedings of ACL-08: HLT*, pages 308–316, Columbus, USA.

Trabelsi, A. and Zaıane, O. R. (2014). Finding Arguing Expressions of Divergent Viewpoints in Online Debates. In *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)@ EACL*, pages 35–43, Gothenburg, Sweden.

Tsur, O., Calacci, D., and Lazer, D. (2015). A Frame of Mind: Using Statistical Models for Detection of Framing and Agenda Setting Campaigns. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, pages 1629–1638, Beijing, China.

Walker, M. A., Tree, J. E. F., Anand, P., Abbott, R., and King, J. (2012). A Corpus for Research on Deliberation and Debate. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2012)*, pages 812–817, Istanbul, Turkey.

Wall, M., Sudulich, M. L., Costello, R., and Leon, E. (2009). Picking your party online–an investigation of ireland's first online voting advice application. *Information Polity*, 14(3):203–218.

Xu, R., Zhou, Y., Wu, D., Gui, L., Du, J., and Xue, Y. (2016). Overview of NLPCC Shared Task 4: Stance Detection in Chinese Microblogs. In *International Conference on Computer Processing of Oriental Languages*, pages 907–916, Kunming, China.

Zhuang, L., Jing, F., and Zhu, X.-Y. (2006). Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 43–50. ACM.

Appendix

Table 2 summarizes the information provided by the participants in the demographic survey. Note that 85 (37%) of the total 230 participants submitted responses to the demographic survey. For each of the sixteen issues explored in this study, Tables 3, 4 and 5 list assertions that most people agree with, assertions that most people disagree with, and assertions that are the most controversial. Below, we show an example questionnaire used for collecting assertions and an example questionnaire used for collecting judgments on the collected assertions:

QUESTIONNAIRE I: CREATING ASSERTIONS ON CONTROVERSIAL ISSUES

Provide at least five relevant assertions on the given controversial issue. The assertions must be expressions that one can agree or disagree with. They can be claims, beliefs, opinions, reasons, arguments, or any statement that can be used to inform or support one's position on the issue. The assertions do not have to be reflective of your own opinions. The assertions can be about a sub-issue or an aspect of the issue.

The assertions should:

- support a position that is relevant to the issue.
- cover a diverse set of positions on the issue. (Avoid claims that rephrase the same argument in slightly different ways.)
- be formulated in a way that a third person can agree or contradict the assertion.
- be self contained and understandable without additional context. (Do not use 'it', 'she/her' or 'he/him/his' etc. to refer to an issue, a person or something else that is not directly mentioned in your assertion.)
- be precise. (Avoid vague formulations such as maybe, perhaps, presumably or possibly.)

The assertions should NOT:

- be a simple expression of agreeing/supporting or disagreeing/rejecting the overall issue.
- contain multiple positions (e.g. migrants are friendly and hardworking).
- contain expressions of personal perspective (e.g. I don't like immigrants).
- be the same as any of the provided examples; or simple negations or other variants of a provided example.

Issue:	Marijuana
Description:	This issue is about legalization of cannabis. This includes the legalization for recreational or medical usage and other positive or negative consequences of legalizing cannabis.

Q1: True or False: This issue is about risks of consuming Cocaine.

( ) true

(✓) false

Q2: Choose the assertion which meets the format requirements:

(✓) The government should discourage any drug usage.

( ) Maybe, the government should discourage any drug usage.

Q3: Enter assertion 1 about 'Marijuana': ______
Q4: Enter assertion 2 about 'Marijuana': ______
Q5: Enter assertion 3 about 'Marijuana': ______
Q6: Enter assertion 4 about 'Marijuana': ______
Q7: Enter assertion 5 about 'Marijuana': ______

QUESTIONNAIRE II: JUDGING ASSERTIONS ON CONTROVERSIAL ISSUES

We want to better understand common controversial issues such as immigration and same-sex marriage. Therefore, we have collected a large amount of assertions relevant to these issues. Your task is to:

- Indicate whether you agree or disagree with these assertions.
- Indicate how strongly you support or oppose these assertions. Since it is difficult to give a numerical score indicating the degree of support or degree of opposition, we will give you four assertions at a time, and ask you to indicate to us:
  - Which of the assertions do you support the most (or oppose the least)?
  - Which of the assertion do you oppose the most (or support the least)?
  - *If you support two assertions equally strongly, then select any one of them as the answer. The same applies for oppose.*

Q1: Indicate whether you agree or disagree with the given assertions on the issue 'Black Lives Matter'.

- Every race has experienced racism.
  ( ) agree ( ) disagree
- There is racial discrimination in the U.S..
  ( ) agree ( ) disagree
- The Black lives matter movement is important.
  ( ) agree ( ) disagree
- Black Lives Matter encourages racial hate.
  ( ) agree ( ) disagree

Q2: Which of these assertions on the issue 'Black Lives Matter' do you support the most (or oppose the least)?

( ) Every race has experienced racism.

( ) There is racial discrimination in the U.S..

( ) The Black lives matter movement is important.

( ) Black Lives Matter encourages racial hate.

Q3: Which of these assertions on the issue 'Black Lives Matter' do you oppose the most (or support the least)?

( ) Every race has experienced racism.

( ) There is racial discrimination in the U.S..

( ) The Black lives matter movement is important.

( ) Black Lives Matter encourages racial hate.

Variable	Answers	%
Education	Bachelor	39%
	some college	15%
	high school	12%
	Master	12%
	associate degree	11%
	vocational certification	9%
	PhD	2%
Race	White	69%
	Hispanic	10%
	Black / African American	8%
	Asian	8%
	other	3%
	American Indian	2%
Religion	Roman Catholic	28%
	Atheist / Agnostic	25%
	Protestant	19%
	other	11%
	Muslim	2%
	Buddhist	2%
	Jewish	1%
	Russian Orthodox	1%
Affiliation	Democrat	45%
	independent	22%
	none	16%
	Republican	15%
	other	1%
Family Status	married	29%
	single (living alone)	28%
	single (living with partner)	20%
	single (living with parents)	19%
	other	4%
Profession	employee	49%
	unemployed	14%
	self-employed	13%
	student	9%
	other	6%
	retired	6%
	civil servant	2%
U.S. Citizen	yes	89%
	no	11%
Ties to overseas	yes	73%
	no	27%

Table 2: Demographic information of the 85 subjects that participated in the voluntary demographic survey.

Issue	Metric	Top 3 Assertions
Black Lives Matter	highest agreement assertions	Every human is equal. Black people are as human as white people. We can all make a better world if we work together.
	most controversial assertions	People of color are more likely to be born into poverty. The Black Lives Matter movement encourages racial hate. Police racial profiling reduces minorities.
	lowest agreement assertions	Blacks are the scum of society. The world would be safer without black people. Not all people are equal.
Climate Change	highest agreement assertions	Global warming can change our climate. Climate change has a big affect on the Earth. Climate change will cause problems for future generations.
	most controversial assertions	Different changes in weather does not mean global warming. The climate change is caused by most developed country. Rising levels of atmospheric CO2 do not necessarily cause global warming.
	lowest agreement assertions	Global warming is not real. Global warming has nothing to do with the change in weather. Trump won the election due to his positions on the environment.
Creationism in school	highest agreement assertions	Freedom of expression is vital to our liberties. Each person should be free to choose what to believe in. People in schools should not be made fun of for their religion.
	most controversial assertions	Religion has no place in a school. Teaching creationism's truth is necessary in upholding the truth of the Bible. Creationism should not be tought in a classroom because it goes against some beliefs.
	lowest agreement assertions	Atheists should be banned from schools. Creationism should include just Christian beliefs versus evolution. Christian religion should be promoted in schools.
Foreign Aid	highest agreement assertions	People have the right to know where there taxes are going and why. We should clearly know where our funding is going and why. Foreign aid should be corruption free
	most controversial assertions	Helping other countries will only increase those nations' dependence on the U.S. Stop borrowing money and raise tax payers dollars. Foreign aid lead to further difficulties for both countries.
	lowest agreement assertions	US spending should focus on defence rather than aid. Foreign aid should be focused on African countries. Foreign aids uplift corruption.
Gender Equality	highest agreement assertions	All people should be treated equally. Wages should not be based on gender. Women can be as successful as men.
	most controversial assertions	Instead of fighting for gender equality, we must fight for gender equity. A woman's physical condition makes her unsuitable for certain jobs. Places don't hire people based on gender.
	lowest agreement assertions	Women should not be in the army (direct combat forces). Gender equality is stupid. The wage gap is a made up thing.
Gun Rights	highest agreement assertions	Gun owners need to be required to have a background check. Gun owners should register their arms. Gun owners should be required to take a gun safety course.
	most controversial assertions	In a certain part to eliminate the arms would be to end the delinquency. Guns should only be issued for hunting. People who own guns are not more likely to mass kill.
	lowest agreement assertions	Everyone should own a gun. The gun industry is too heavily regulated. Guns should be legal for everyone.
Marijuana	highest agreement assertions	Drug abuse can kill. Exceptions should be considered allow people with medical issues that will benefit from legalization of Marijuana. Marijuana is proven to have medical benefits.
	most controversial assertions	Legalization of marijuana will result to people not pushing hard drugs. Marijuana use will not increase just because it is legalized. People who use marijuana are mentally stuck at the age they were when they began using the drug.
	lowest agreement assertions	Marijuana should be as readily available as cigarettes are. Allowing the legal use of marijuana will prevent drug dealers from selling it. Cannabis is nontoxic

Table 3: The top three ranked assertions according to different metrics of the agreement score (issues 1-7).

Issue	Metric	Top 3 Assertions
Same-sex Marriage	highest agreement assertions	God loves all the people for equal. One person's believes should not affect another persons rights. Love is a right for everyone.
	most controversial assertions	Same sex marriage is unlike any other marriage. Allowing same-sex marriage will create a 'slippery slope' where people will begin to fight for other alternative marriages such as polygamy. Marriage was God's idea. He defined it as between a man and a woman only. End of conversation.
	lowest agreement assertions	Saying that gay people should get married is like saying that a brother can marry his sister both are at higher risk of disease. There is more domestic abuse in homosexual couples. Children of same-sex couples will turn out gay.
Mandatory Vaccination	highest agreement assertions	The vaccines must be given by people who are very prepared to put them. Vaccines should be free. Medical treatments like vaccines should be for all the people.
	most controversial assertions	Vaccines can cause serious and fatal side effects. Vaccinations are expensive and yet not 100% safe. The theory of herd immunity has never been proven.
	lowest agreement assertions	Vaccinations cause autism. Vaccines increase the risk of autism. Vaccines are not safe.
Media Bias	highest agreement assertions	The job of a reporter is to tell the truth. Every body has its owns preferences. The society deserves the truth.
	most controversial assertions	If you don't watch the media , bias isn't a problem. Journalists only cover stories that support their own opinion. Media is liberal biased because the liberals are right.
	lowest agreement assertions	Fox News is the only unbiased news source. Journalism focusses on Democrats views. Fake news is not a real problem.
Obama Care	highest agreement assertions	Every deserves a chance at life. Everyone needs insurance. Health care should not punish people for pre-existing conditions.
	most controversial assertions	People may rely on free health care to substitute there own role in maintaining unhealthy lifestyle choices. Obama care should be removed. An unregulated market ensures cheap medicine.
	lowest agreement assertions	Medical treatments should be only for people that can afford treatments. By giving people free health care it can lead to them being lazy. Free health care leads to people not taking care of their health.
US Electoral System	highest agreement assertions	The people should elect there leader. The electoral system must be fair. States should do what they can to avoid voter fraud.
	most controversial assertions	The electoral college protects minority interests. Popular vote should never be used as a determining factor because it give total control of the country to large coastal states. Our good electoral system puts US as a world example.
	lowest agreement assertions	Voting for an independent candidate only ensures that the Republican will win. Gerrymandering is a valid system. The voting age should be changed to 16.
US in the Middle East	highest agreement assertions	US Engagement in the middle east needs to be carefully addressed. US should protect America against terrorism. Middle east countries should be allowed to determine their own future.
	most controversial assertions	Withholding funds to Palestinian National Authority prevents anti-Semitism. Usa's engagement in middle east is not a public interest topic. The US needs to show its support for Israel by moving its embassy to Jerusalem. It would make a big statement to the world.
	lowest agreement assertions	All Muslims should be regarded as terrorists. It is our duty to spread Western ideals to the Middle East. The war in Iraq was worth the costs.
US Immigration	highest agreement assertions	Migrants have a positive impact on the economy. A person in need should not be denied help and life regardless of where they are from. Legal immigration has brought some great scientists to the United States.
	most controversial assertions	Immigration takes opportunities away from those born in the USA. Immigration should be elevated because it brings the best brains to the economy. Illegal immigrants take jobs away from Americans.
	lowest agreement assertions	Immigrants are more often criminals. Immigrants are bad for the US. Immigrants are better workers than US workers.

Table 4: The top three ranked assertions according to different metrics of the agreement score (issues 8-14).

Issue	Metric	Top 3 Assertions
Vegetarianism & Veganism	highest agreement assertions	People have the right to choose the kind of meal they want to eat. A persons diet should be there choice and the individual freedoms of choosing one food should not be shamed. Creating a nutrition plan is left for the person.
	most controversial assertions	Vegetarianism is healthier than veganism. Killing animals is harmful. Veganism is like a religion.
	lowest agreement assertions	Killing animals is against Gods law. I like cows, so I don't eat them. Humans are meant to be vegan.
War in Terror	highest agreement assertions	Terrorism is an international threat. The government must take care that the taken measures on counter terror does not affect innocents. Terrorism has destroyed too much and needs to end.
	most controversial assertions	As a bonus, war on terror almost guarantees safety worldwide. The war on terrorism is an invasion of privacy. It is ok that our taxes support the war on terrorism.
	lowest agreement assertions	All Muslims are terrorists. Terrorism is only a problem in the middle east. War on terriorism kept us safe so far.

Table 5: The top three ranked assertions according to different metrics of the agreement score (issues 15 & 16).