

# An Analysis of Political Views on Blogs

## October 27, 2008 Update

Todd Sullivan

todd.sullivan@cs.stanford.edu

## 1 Introduction

This week we are exploring political views expressed through a blogger's interests and communities joined. We have expanded our analysis from Obama, McCain, Biden, and Palin to also include the more generic classes of Republican, Democrat, Liberal, and Conservative. Due to expanding our analysis to include more generic mentions, we also examine FeelScore across network and gender (as in our previous update) for these new classes. In this update and all future updates, our use of "class" or "classes" refers to the entities Obama, McCain, Biden, Palin, Republicans, Democrats, Liberals, and Conservatives.

Some of the questions answered during this update include:

- Which classes have the most distinct author interests pertaining to them?
- Which classes have the most authors listing a positive or negative interest towards them?
- How likely is an author to list multiple interests pertaining to a single class?
- Which classes have the most distinct communities pertaining to them?
- Which classes have the most authors joining a positive or negative community about them?
- How likely is an author to join multiple communities about a single class?
- Are authors more likely to list multiple interests or join multiple communities pertaining to a single class?
- All of the questions from the October 20 update, except pertaining to the new

classes of Republican, Democrat, Liberal, and Conservative.

## 2 Author Interests

We searched for interests in the dataset that contained the last name of a candidate or the words "republican", "democrat", "liberal", or "conservative". This returned 348 distinct interests. Since this set is rather small, we hand-labeled each interest as being positive, negative, or neutral towards each class with each class only being labeled as non-neutral if the interest explicitly contains the class's name. Thus a positive interest for Republican does not count as being positive for McCain or negative for Obama. After filtering for irrelevant interests such as "palindromes", we are left with 302 distinct interests.

Most of these interests are from the LiveJournal network, with 273 interests from LiveJournal and 29 from Blogger. The dataset contains interests starting on August 5 2008. This fact will affect our results when doing a time-based analysis in later updates.

Table 2 summarizes the number of interests that are positive or negative for each class, as well as the number of distinct authors listing a positive or negative interest for each class. The counts clearly show that people list positive items in their interests much more frequently than negative items. Mentions for the left-leaning classes (Obama, Biden, Democrat, and Liberal) also far outweigh the right-leaning classes in terms of the number of distinct interests and the number of distinct authors prescribing to the interests. Of the candidates, Obama's mentions in the interests are far more numerous than all other candidates.

Biden remains the least mentioned candidate, but his trend of mostly positive mentions from our October 20 analysis continues with all of his mentions in interests being positive.

Table 2: Interest Counts by Class

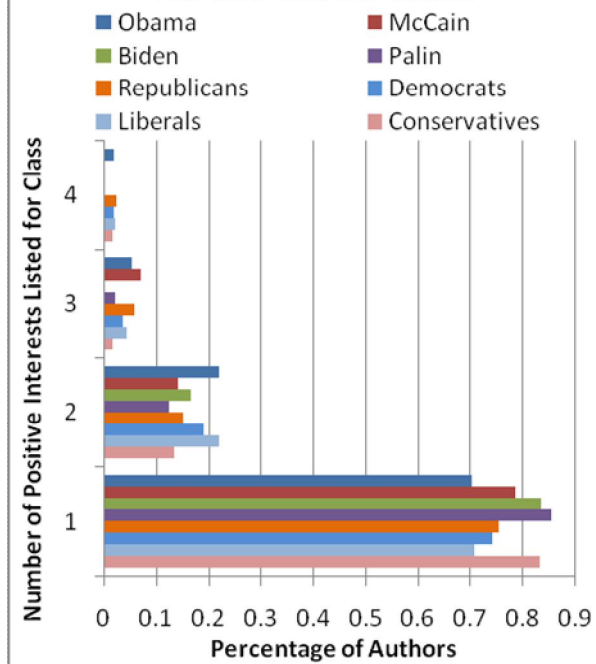
Positive Interests		
Class	Interest Count	Author Count
Obama	33	201
McCain	5	14
Biden	4	6
Palin	7	48
Republicans	25	118
Democrats	41	373
Liberals	100	605
Conservatives	38	112

Negative Interests		
Class	Interest Count	Author Count
Obama	2	3
McCain	3	5
Biden	0	0
Palin	5	6
Republican	15	28
Democrat	2	3
Liberal	12	13
Conservative	10	17

Figure 2 shows an overhead view of how many authors prescribe to a given number of positive interests per class. We see that for each class 70% to 80% of authors only list one positive interest while 10% to 20% list two interests. The data contains a few outliers that are now shown, with one author prescribing to 12 positive interests for the Democrat class and another prescribing to 9 positive interests for the Liberal class. The result seems to correlate well with the number of interests pertaining to a particular class. Classes with a large number of interests, such as Democrats or Liberals, tend to have more authors prescribing to multiple interests for the class.

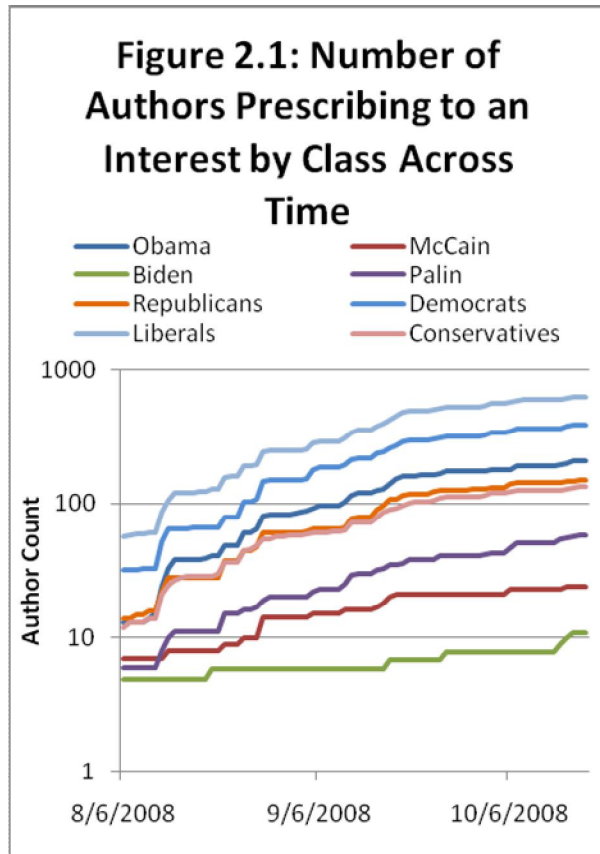
Figure 2: Authors Listing Multiple Interests for the Same Class



## 2.1 Interests Across Time

Figure 2.1 plots the number of authors prescribing to an interest by class across time. This data uses the date that each author-interest connection was inserted into the database. We had hoped that this information would be a good rough estimate of when the author actually added the interest to his or her profile.

Each class' curve exhibits the same stair-step behavior regardless of the class' left or right-leaning association. Unfortunately, this behavior is most likely due to the scheduling of the crawler that collected the data, with batches of interests collected and inserted into the system every three to four days. This time-based view of authors' interests will still be useful when examining each class' FeelScore across time, but it is not useful for examining interests alone.



## 2.2 Interests by Gender, Age, Etc.

The vast majority of authors with relevant interests do not have a listed gender, birthday, or location. For example, of all authors with an interest pertaining to Obama only 2 have a gender that is not “unspecified”. We believe that this is mostly due to the format of LiveJournal and Blogger rather than any overwhelming connection between listing a political view in your interests and staying anonymous in terms of gender.

In fact, LiveJournal does not contain any gender information for authors while Blogger contains gender information for 45.1% of authors. On the age front, the LiveJournal data contains birthday information for only 1.9% of authors and Blogger for 27.2%.

## 2.3 Example Interests

Many of the positive interests simply state a class, such as “obama”, “mccain 08”, or “republican”. Most of the negative interests fol-

low a well-defined format of “doing something to \_class\_” where “\_class\_” is replaced by the respective class. The “doing something” part ranges for civil comments to rather harsh comments. Some examples are: “anti”, “teasing”, “baiting”, “bashing”, “hating”, “pissing off”, “cockpunching”, “forced lobotomization of”, and “death to”.

## 3 Author Communities

We generated a list of relevant communities in the same way as the interests. The community data comes entirely from LiveJournal with 128 communities pertaining to at least one of the classes. The data contains 602 distinct authors from LiveJournal.

Table 3 summarizes the amount of positive and negative communities pertaining to each candidate. The characteristics are similar to the interests examined earlier. Obama has the most communities out of all candidates. Biden has the least positive communities but does not have any negative communities. Palin is the candidate with the most negatives communities. The left-leaning communities vastly outnumber the amount of right-leaning communities.

Figure 3 shows the percentage of authors that join a given number of positive communities about each class. Since there are far less communities in comparison with interests, one might expect Figure 3 to have higher percentages at a frequency of 1 in comparison to Figure 2. Somewhat surprisingly, this only happens with Biden and Conservatives. All other classes have a higher percentage of authors joining at least two communities.

This behavior of joining multiple communities more often than listing multiple interests is intuitive when examined from another angle. Most people probably find making a single statement in their interests section about a class is sufficient. There is little gained from listing both “obama” and “obama for prez” in your interests section. On the other hand, by joining multiple communities about a

single class you are maximizing the number of people that you encounter with similar political feelings. Thus one would expect bloggers to be more likely to join multiple communities for a single class than list multiple interests for the class.

Similar to interests, the community data contains the date that each author-community pair was added to the database. This time information exhibits the same stair-stepping behavior as interests and is most likely due to the same reasons discussed in Section 2.1. Also like Section 2.2, the community data contains insufficient information for an analysis by gender, age, or location.

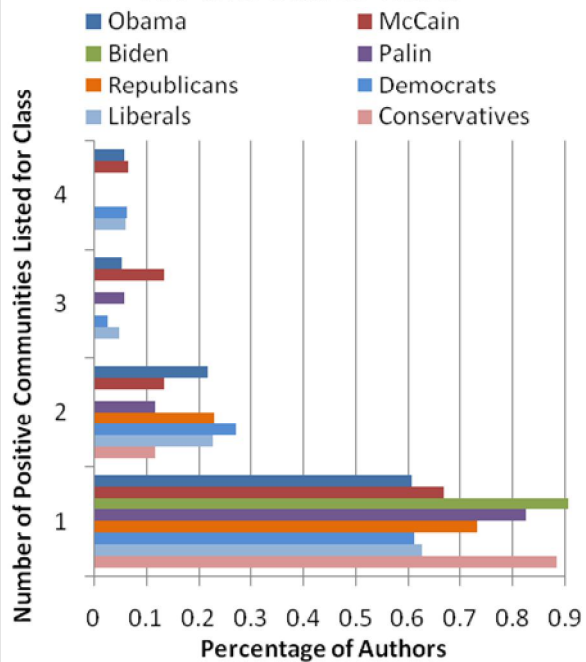
Table 3: Community Counts by Class

Positive Communities		
Class	Community Count	Author Count
Obama	26	206
McCain	6	15
Biden	1	3
Palin	5	17
Republicans	8	26
Democrats	19	111
Liberals	45	281
Conservatives	9	17

Negative Communities		
Class	Community Count	Author Count
Obama	1	9
McCain	2	13
Biden	0	0
Palin	3	11
Republicans	0	0
Democrats	0	0
Liberals	3	10
Conservatives	0	0

Figure 3: Authors Joining Multiple Communities for the Same Class



### 3.1 Example Communities

The majority of positive communities list a class along with some kind of demographic information such as race or location. Examples include “Obama PA”, “GothsForObama”, and “TeenRepublicans”. The few negative communities that exist generally contain “anti”, “no”, or something else such as “wtf”.

## 4 Republicans, Democrats, Liberals, and Conservatives

In this section we examine political feelings towards the Republican and Democratic parties in general. This analysis follows the same route as our candidate analysis in our October 20 update. We target both specific mentions of each party as well as the generic classes of “liberal” and “conservative” that are often associated with each party. We generated FeelScores for each class on the author level as described in our October 20 update.

## 4.1 Overall FeelScores

Figure 4.1a shows the distribution of authors with positive/negative chatter about at least one of the four classes. The sum of the left-leaning classes (46%) and the sum of the right-leaning classes (54%) is similar to the distribution of candidate mentions in sentences from Figure 2.2a of our October 20 update, which had 45% for Obama/Biden and 55% for McCain/Palin. The mention of Liberals vs. Conservatives is relatively even, but the Republican party is mentioned much more often than the Democratic party.

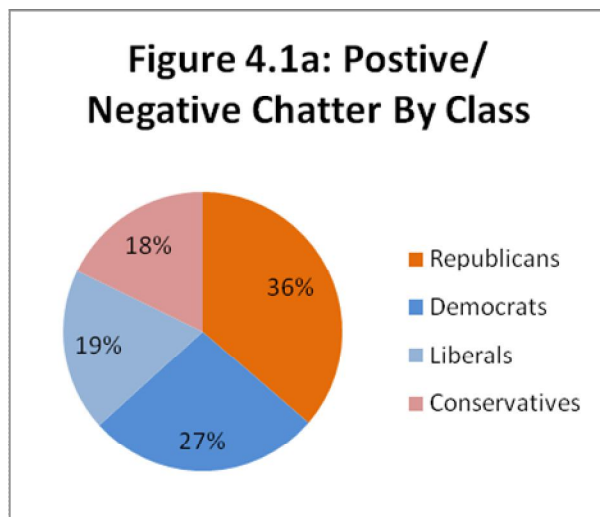
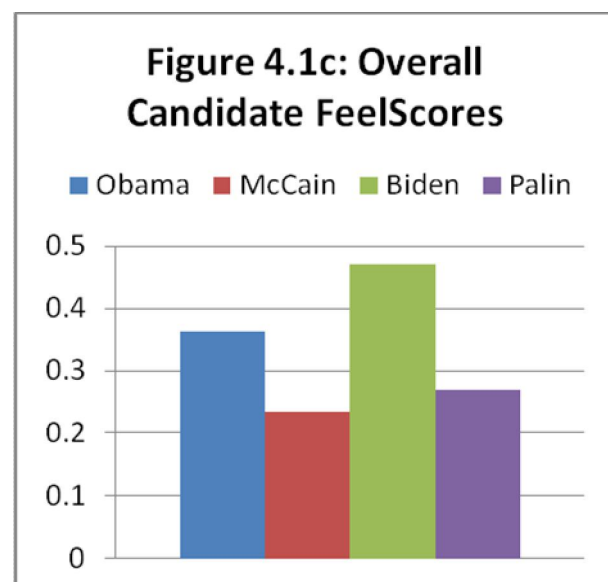
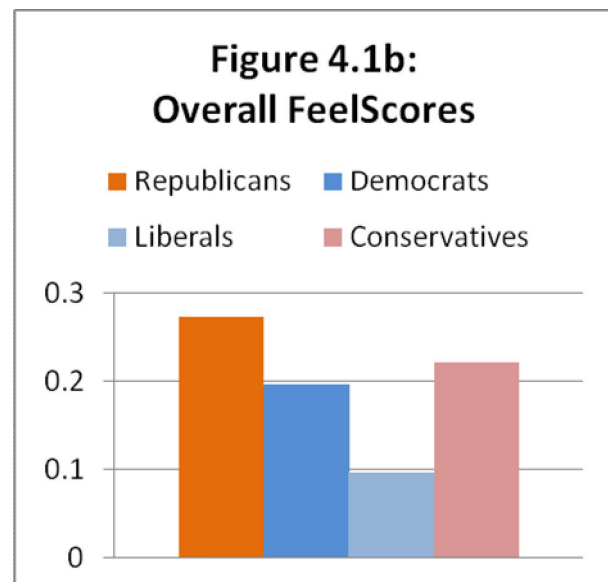


Figure 4.1b shows each class' overall FeelScore. The result is surprisingly different from the candidate results of our previous update. While all of the FeelScores are lower than the candidate FeelScores, the Republicans come out on top. This is a stark contrast to the result of our previous analysis, reproduced here as Figure 4.1c, where the Democratic ticket dominated the Republican ticket.

In general, it is expected that the Liberal and Conservative classes have a lower FeelScore than the Republican and Democrat classes. This is because many people tend to use the mere term "liberal" or "conservative" as an insult. In our own experiences, we have found that people trash talking democrats are much more likely to use the term "liberal"

than people trash talking republicans are to use the term "conservative". Additionally, we believe that conservatives are more likely to use the term "liberal" as a negative than liberals are to use the term in a positive manner. This hypothesis explains why the Liberal class has such a low FeelScore in comparison to the other classes.



### 4.1.1 Qualitative Analysis

A qualitative examination of sentences containing “liberal” found few sentences that could be summarized as saying “I am a liberal”. On the other hand, we found many sentences saying “I am a conservative.” We performed this examination by searching for sentences that matched the following regular expressions:

```
I (am|'m) ([^\n]*)liberal
I (am|'m) ([^\n]*)conservative
```

The “liberal” regex returned very few sentences, with a relatively small number of sentences meaning our desired phrase. Some example sentences include:

```
I'm a liberal by politics
I am a liberal here
I'm generally more liberal
```

The “liberal” regex actually returned an equal number of sentences saying something negative about liberals. Examples include:

```
I'm really sick and tired of so-called liberals.
I am home sick today, do I have the flu or am I sick of watching liberals
I'm not a liberal
```

On the other hand, the “conservative” regex returned no negative mentions and returned twice as many sentences meaning our desired phrase. Some examples are:

```
I'm very conservative
I'm fairly conservative
I am conservative
I'm conservative by nature
I am one conservative
```

While our qualitative analysis was not exhaustive, we feel that it adequately supports our hypothesis that the term “liberal” is more often used by conservatives in a negative context than it is used by liberals in a positive way.

### 4.2 FeelScores by Blog Network

Figure 4.2a shows the positive/negative chatter distributions for MySpace, LiveJournal, and Blogger. Just as in our candidate analysis, MySpace’s proportions are significantly different from the other two networks. MySpace has more left-leaning chatter than the other networks. In terms of FeelScores, MySpace is also the oddball.

**Figure 4.2a: Positive/Negative Chatter By Blog Network**

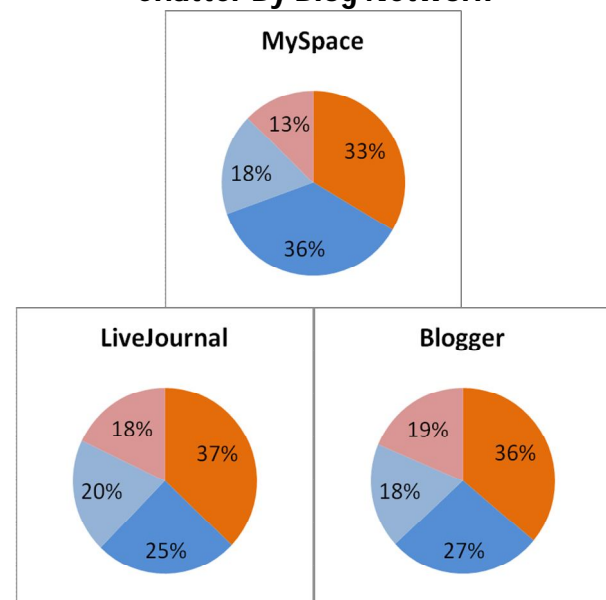
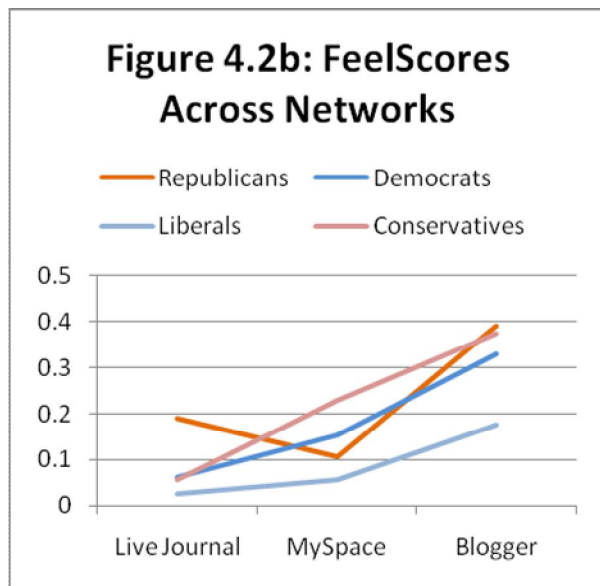


Figure 4.2b shows the FeelScores for each class across the three networks. The graph shows a general trend in the averages of positive and negative mentions moving from left to right. LiveJournal seems to be the harshest network while Blogger is overwhelmingly more positive about all of the candidates than the other networks. Strangely, MySpace has a much lower FeelScore for the Republican class than the other networks yet the Conservative class has MySpace’s highest FeelScore.

The FeelScores in Figure 4.2b are opposite to what would be expected from looking at the candidate analysis from our previous update. In our candidate analysis the Obama/Biden ticket had a significant majority on all net-

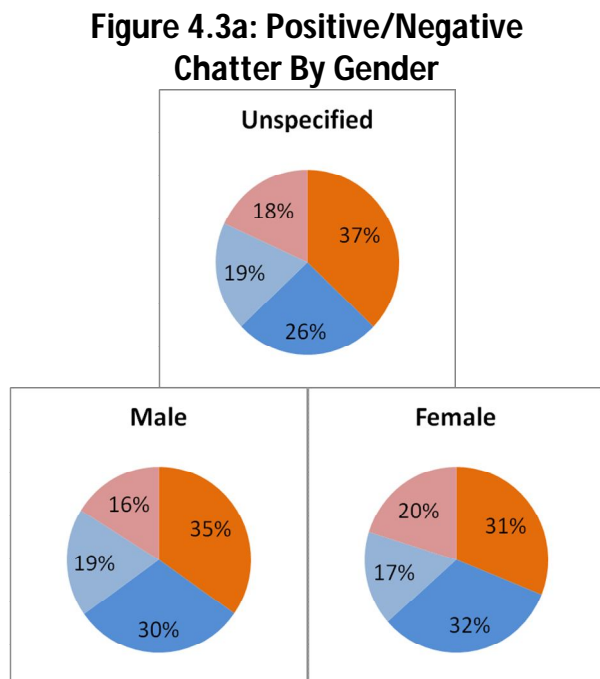


works yet the Republican/Conservative classes have a majority overall in this case.



### 4.3 FeelScores by Gender

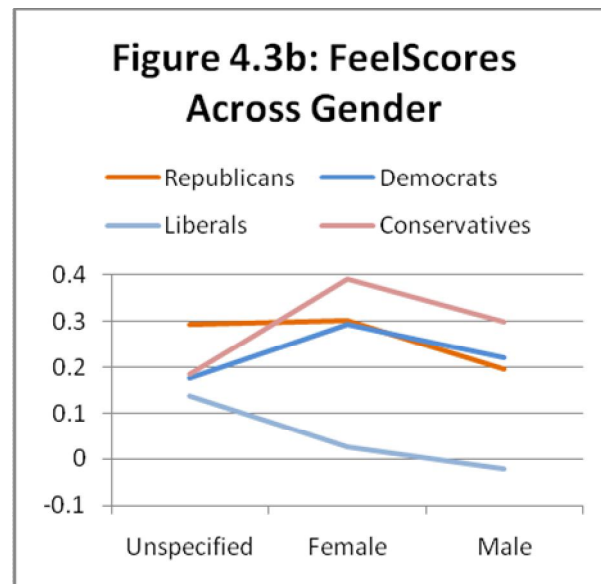
Figure 4.3a shows the positive/negative chatter distributions by gender. The data does not contain any noteworthy items. Most of the differences between genders are too close to definitely say anything.



The FeelScores on the other hand are quite different. Figure 4.3b shows the FeelScores of each class across gender. When comparing males and females, males tend to be harsher across the board: all classes have a lower FeelScore from males than they do from females. Compared to authors that do not specify their gender, males and females have a much lower opinion of the Liberal class with the male FeelScore even going negative.

This result is heavily influenced by the fact that the LiveJournal network does not contain gender information. A glance back at Figure 4.2b shows that LiveJournal was the harshest network for all classes except Republican. This exposes a flaw in our previous report's examination of candidate FeelScores, which we will update when doing a final overview after the election.

Since we are examining FeelScores by gender, we should only include networks that contain gender information. Figure 4.3c presents a corrected graph that only uses authors from MySpace and Blogger. Obviously, the FeelScores for females and males remain the same because removing LiveJournal data did not affect any author that specified his or her gender. The differences in FeelScores for the unspecified gender class is quite large.



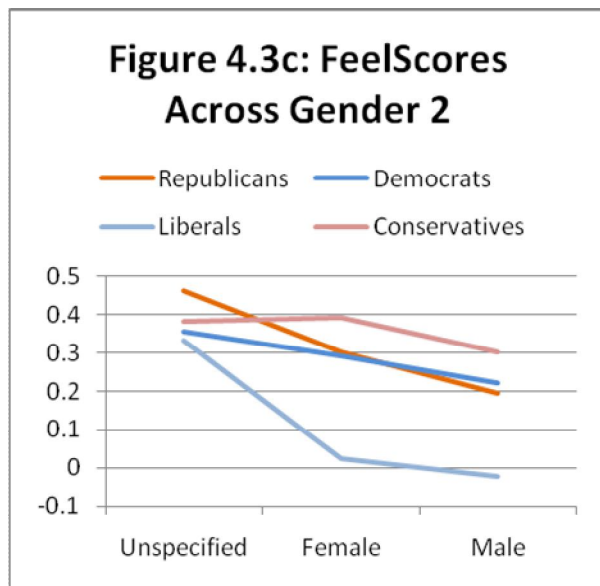


Figure 4.3c shows that for all classes except Conservative, authors that specify their gender are more likely to rate the class lower than their unspecified gender counterparts. This difference is quite dramatic for the Liberal class while less dramatic for the Republican and Democrat classes. The result is counterintuitive because in general one would believe that anonymity enables bloggers to say more drastic and harsh comments. LiveJournal, which does not have a gender specification field in profiles, is definitely more negative than the other networks. Yet here, the act of specifying gender is strongly correlated with negativity (especially for the Liberal class).

## 5 Future Work

Our future work includes incorporating interest and community information into each author's FeelScore, as well as creating a FeelScore for each author over time. We will examine each class' FeelScore over time and look for changes in FeelScore that can be aligned with important events such as each party's convention, the debates, and other news items. We will also separate FeelScores over time by gender and other demographics when enough information is available. Another interesting avenue will be to examine the link between authors' Republican, Democrat, Liberal, and Conservative FeelScores and their candidate FeelScores. This might help us understand why on the candidate side the left-leaning classes are rated highly while on the generic term side the left-leaning classes are rated much lower than their competition.