

Why don't you like me?

A quantitative study into group exclusion

Adam Kaplan

May 9, 2018

Abstract

In this paper, I will explore norms of difference in group and the prisoner's dilemma fringe introduced by Russel Hardin empirically. [1] Using a large Reddit data set [4], I will define formal theoretic definitions with the data and then estimate models that verify Hardin's theory as well as identify a relationship between commitment of members in a group and the group's strength of norms of difference. [1] This paper's approach to group conflict is novel due to its reliance on large numbers of observations and summary statistics of distributions as independent variables.

1 Introduction

Group conflict is a very important field of study in international relations. Russel Hardin, with his book *One for All*, helped solidify a sociological and formal theoretic explanation of why conflict occurs among groups and why different groups even exist. [1] Due to the sociological nature of group conflict, up until now, empirical verification of the theory was not possible, due to the lack of data. As social media is becoming more and more important in the way groups are built and communicate, it suggests itself to use social media data to look into sociological theories.

In this paper I will focus on the fourth chapter of Hardin's book: "Norms of Exclusion." [1] In addition to trying to confirm Hardin's theory of a prisoner's dilemma fringe, I will also explore the connection between the strength of norms of difference and the commitment of individual members

in different groups. I am going to be using a large Reddit comment data set, which contains over 2 billion posts of users in different subreddits. [4] I will explain why this data set is well-suited for analyzing group identities and how I define Hardin's concept in terms of the data. After running regressions, I will spend some time reflecting on the results and what future research can be done with this data set in terms of group conflict.

2 Related Literature

This paper is heavily based on Russel Hardin's book *One for All, The Logic of Group Conflict*. [1] In particular, it focuses on its fourth chapter "Norms of exclusion. In this book, Hardin studies group identity as strategic interactions, coordination and cooperation. Some of the core questions Hardin tries to answer in this book are: [1]

1. Why do group identities exist? (Chapter 1)
2. Why do people participate in groups? (Chapters 2,3,4)
3. Do group identities differ and if so, how? (Chapters 4,5)

I will quickly address each of these questions in this section, but the remainder of the paper will focus heavily on the last question whilst touching on the second. I want to address all of these questions, because it creates the necessary background for this paper. To answer the first question, we need to consider the power of coordination through the coordination game. [1] I will use the stag hunt game as the example. In this case, hunters can either coordinate and successfully hunt for the stag or not coordinate and hunt individually for the less satisfactory hare. Clearly, the stag is preferred by every hunter, but it requires all of the hunters to coordinate. The hare is less preferred, but does not require any coordination at all. Group identities try to help with this problem. If a group identity somehow can ensure that everyone coordinates, being a member of the group is beneficial. So, the reason group identities exist, according to Hardin, is because they help coordinate a group of people. [1] The last two questions are all about how a group can achieve that.

We have identified a reason for a group identity to exist, but that is not enough to explain why people are contributing to the group through coordination. An easy explanation is that there are no costs associated with joining and coordinating and some benefit can be reaped from it. [1] Then, people are always better off joining this group, since they will always benefit from it. There are also groups which are costly to enter, for instance fraternities/sororities, what makes people join these? Here, the benefits reaped from membership need to be greater than the cost of joining and coordination. [1] A cynic, would still say, what guarantee do members have that no one will cheat? Only if the coordination problem has been overcome, that is members coordinate, the rewards can be reaped. Therefore, punishment needs to come into play. In other words, we are at a Prisoner's Dilemma situation. [1] A group identity needs to implement a deterrent for defection, so cooperation becomes an attractive choice. For example, ethnic group identities or national group identities have a natural punishment: exclusion. Imagine you live in a ethnic community and you are considering defecting. The potential outcome of exclusion from the community would be disastrous and rarely is there a benefit that would outweigh it, hence defection is successfully deterred. This sort of punishment seems to work well if members are heavily involved in the group. [1]

One way this heavy involvement can occur is through what Hardin coins *norms of difference*. [1] In particular, groups will try to differentiate themselves from society. The best example is nationalism. Each country has their own language, way of life, etc. and nationalism can therefore naturally thrive, because people who understand all of these norms of difference, differentiate themselves from other countries. New languages, are an extreme example, but for instance different majors in one university differentiate themselves through different colored hoods during graduation, which is less extreme. This gives rise to the notion of strength of norms of difference. [1]

3 Norms of Difference

Each group decides for itself how different it wants to be from society and it can change over time. Hardin shows an example of how African-Americans in the United States have now stronger norms of difference, than say Jews in the United States, which allow more and more marriages to

outsiders. [1] Strong norms of difference might seem unfair to outsiders, but allow the members of a group to coordinate better. In particular, if a group creates their own dialect, which makes in-group communication easier, the members benefit, but at the same time, the group separates itself more and more from society, because for an outsider to join, requires a lot more work and for members, communication with outsiders also seems more work-intensive. [1]

We also know that commitment problems are littered with free rider problems, that is, members do not contribute to goal, but still reap the benefits that come from achieving it. [1] Commitment problems can easily occur with group identities as well, because individuals can join multiple groups, but they have only a finite amount of time they can spend on each group. In other words, there is an incentive for individuals to join a group, not contribute, reap the benefits, but contribute a lot in another and reap the benefits there as well. [1] As discussed in the previous section, that is where punishment comes into play. We noticed that exclusion from a group is a good punishment members which assimilated to a group, but not so much for groups where no assimilation needs to happen to join. Now, strong norms of difference, by definition, force new members to assimilate and groups with weak norms of differences do not. This implies that punishments differ by strength of the norms of differences. [1] Hardin refers to this as the *prisoner's dilemma fringe*. [1] By that he means if a member defected in a group with strong norms of differences, even though the group would be better off with partial exclusion, due to the strength of its norms, full exclusion occurs, thereby reducing the fringe. On the other hand, groups with weak norms do not have the problem of having to raise costs of defection as much, meaning partial exclusion ought to occur more often, thereby increasing the fringe. This is my first (Hardin's) hypothesis.

Hypothesis 1 (H1): *Strong norms of difference increase the number of fully excluded defectors.*
[1]

The connection between the level of commitment and strength of norms is also worthwhile to explore. A group with strong norms has "benefits of comfort, familiarity, and easy communication [...]"¹ [1] So, people who decide to join such groups do it in part to reap these benefits. The only way one can reap these benefits is to actually commit to the group and stay committed.

¹Page 77

Hypothesis 2 (H2): *The distribution of commitments in groups with strong norms of differences is skewed towards high commitment.*

In contrast, groups with weak norms, do not offer benefits whose enjoyment is that directly tied to commitment, and therefore I hypothesize that weak norms lead to a lower proportion of highly committed members.

Hypothesis 3 (H3): *The proportion of highly committed members increases as norms of differences increase in strength. Equivalently, the proportion of highly committed members decreases as norms of differences decrease in strength.*

4 Quantitative Approach

The contribution of this paper is the empirical verification of political theoretic results, mainly by Hardin, in group conflict. It is well-known that group identities and other similar sociological construct are very difficult to quantify. As almost all interaction and gathering happens over social media, more data is starting to become available for researchers to attempt to quantify them.

4.1 Data

I am using a freely available (on Google BigQuery) Reddit comments data set. [4] Reddit.com is a well-known website, which brands itself as the homepage of the internet. The reason it is well-suited for studying group identities is because all of its discussions are bundled into subreddits. Over the years, there is a subreddit for almost anything, ranging from Apple, Android, major religions, nation states, political and non-political movements, etc. Additionally, anyone can sign up for an account and post in any of the subreddits. The data set ranges from 2015 to 2017 and contains every single comment posted on Reddit during that time. The total number of posts in that time period is approximately 2.5 billion. In addition to the comment itself, it includes the username of the user, the subreddit it was posted in as well as a timestamp of when it was posted. It also includes further data points such as the rating of the post and the specific topic it concerned, but for the sake of this paper, I am only going to consider the username, subreddit

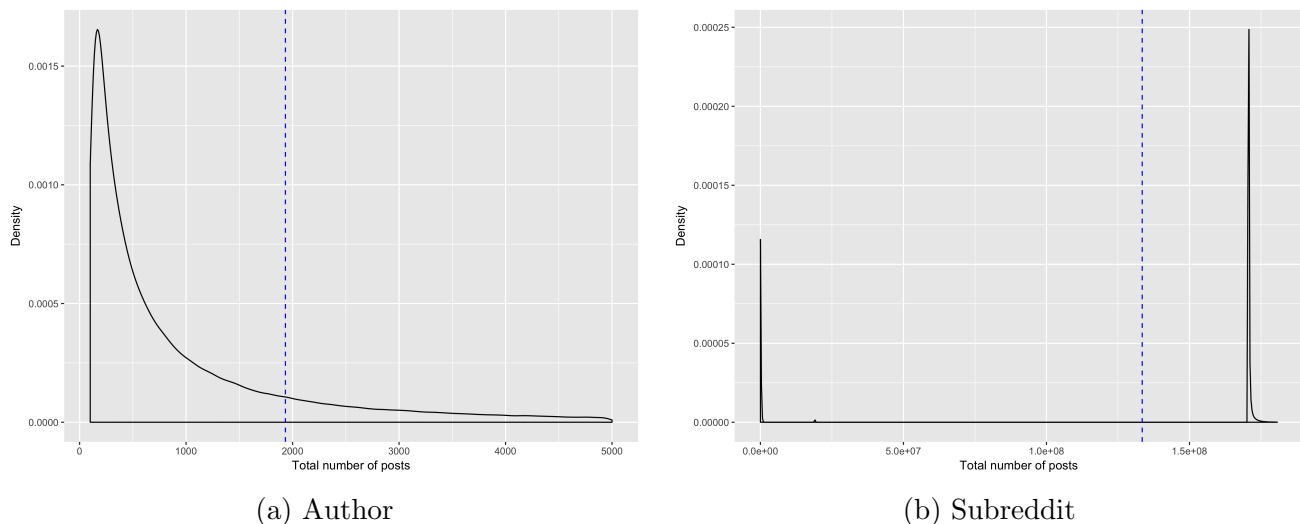


Figure 1: Densities for the number of posts for both authors and subreddits. Blue dashed line indicates the mean.

and timestamp.

Since 2.5 billion comments are far too much to analyze, I had to cut down the sample size. I decided to only consider users that posted in the following subreddits, which are examples of possible groups: Android, apple, Democrat, Republican, islam, Christianity, Buddhism, de, unitedstatesofamerica, Israel, Judaism. I did this to select users which are in at least one subreddit that can represent a traditional group identity. This left me with 580,730 distinct users, which in total posted 728,440,782 times in 31,371,756 different subreddits. To eliminate one-time-posts, I decided to omit all observations where a user posted less than 10 times in a subreddit over the two year period. Similarly, I also removed subreddits with less than 10 users, because on Reddit such a subreddit would be considered "dead." Figure 1a displays the distribution of posts per user over the 2 year period.² We see that most users post about 2000 times. I had to truncate the plot, because the numbers continue well into the tens of thousands of posts over the 2 year period. Similarly, the total number of posts per subreddit over the same time period can be seen in figure 1b.³ We see that there are two big spikes. One, which seems at about 0, but in reality is closer to a couple tens of thousands and the other at around 150 million posts. What we see here is that there is a clear split in the data between "small" and "large" subreddits. In this paper, we focus on the

²Used the following code: <http://www.sthda.com/english/wiki/ggplot2-density-plot-quick-start-guide-r-software>

³Used the following code: <http://www.sthda.com/english/wiki/ggplot2-density-plot-quick-start-guide-r-software>

”smaller” groups due to the fact that in the larger subreddits, most users post only very few times so our measures for skewness and kurtosis would be inaccurate. It would be of interest to compare the results between the two subreddits in future research with different definitions of exclusion.

4.2 From Subreddits to Group Identities

As alluded in the previous section, I am using each subreddit as a group identity. Due to the large number of subreddits in the sample, the group identities can vary from very vague such as ”videos” to a very well known group identity ”Buddhism”. In either case, members of ”videos” are bound together by their enjoyment of funny videos and members of the ”Buddhism” group identity, by their interest in Buddhism. Before going any further, it is important to define what it means to be a member. As mentioned, in Reddit, anyone can join any subreddit, eliminating the possibility of ”costs to enter,” but due to the vast amount of subreddits, many subreddits are as-if hidden to the regular user, requiring directions from current members of a group to join. Therefore, I consider anyone who posted in a subreddit to be a member of the specific group identity. By this definition it is impossible to leave a group, since once a post is made, membership is granted. Since, exclusion is at the core of what this paper is trying to study, it is necessary to define what it means for a member to be excluded from a subreddit.

4.3 Skewness and Kurtosis

Before defining exclusion it is necessary to introduce the concepts of *skewness* and *kurtosis* of a distribution. A very natural definition is the average of the z -score cubed. [3] For any data set with n observations the z -score is defined as follows, [3]

$$z_i = \frac{x_i - \bar{x}}{\sigma}$$

where σ is the standard deviation of the whole data set. Cubing it will yield a greater emphasis on the outliers to the left and right of the mean. [3] Taking the average will yield either a negative or positive number and rarely 0. This sign of the average indicates whether or not the distribution

is negatively or positively skewed. [3] Figure 2 shows an example of what it means to be skewed

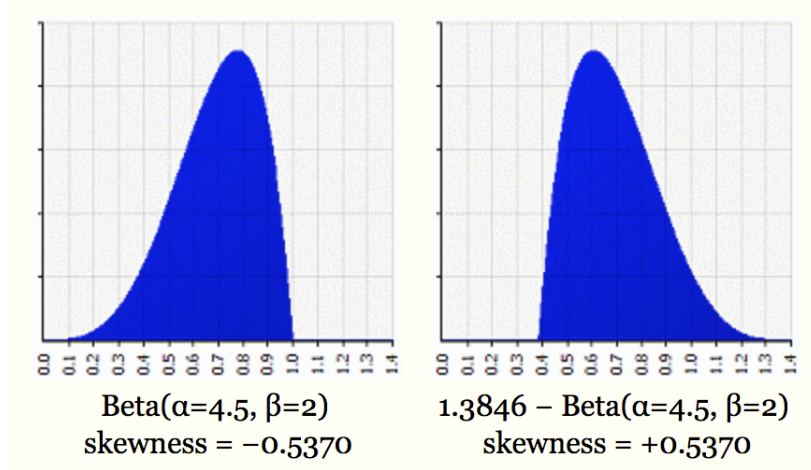


Figure 2: Left: negatively skewed. Right: positively skewed. [3]

in either direction. The last notion required for skewness is the strength of the skewness. A rule of thumb is, [3]

1. Highly skewed: If skewness is less than -1 or greater than 1.
2. Moderately skewed: If skewness is between -1 and $-\frac{1}{2}$ or between $\frac{1}{2}$ and 1.
3. Approximately symmetric: If skewness is between $-\frac{1}{2}$ and $\frac{1}{2}$.

With this in mind we see that the two examples in figure 2 are moderately skewed.

The next important concept is called *kurtosis*. Skewness is the the z -score raised to the third power and kurtosis is the z -score raised to the fourth power. [3] From that follows another good interpretation. Note that for a distribution where a lot of data points are far from a mean the z -score raised to the fourth power will be far greater than 1 and for a distribution where most points are close to the mean, we get a kurtosis coefficient of close to 1. [3] Figure3 displays the difference in kurtosis. In particular we say a distribution is, [3]

1. Mesokurtic: If kurtosis is approx. 3. The normal distribution is mesokurtic.
2. Platykurtic: If kurtosis is less than 3. Tails are shorter and thinner compared to the normal distribution.

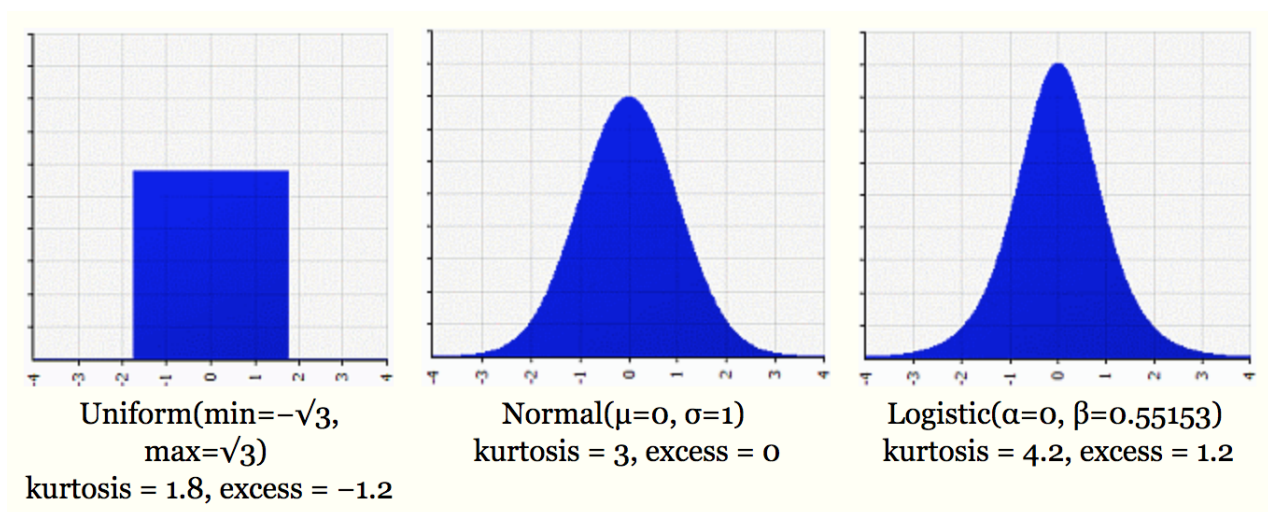


Figure 3: Left: Platykurtic. Middle: Mesokurtic Right: Leptokurtic. [3]

3. Leptokurtic: If kurtosis is greater than 3. Tails are longer and fatter compared to the normal distribution.

4.4 Exclusion from a group

Even though Reddit allows for users to be banned from subreddits, by a chosen member of the subreddit, called "moderator," the data set did not contain data on whether or not a user was banned from a subreddit. Therefore, I decided to define exclusion as a rapid fall in posts for a particular subreddit. Since I had the timestamps for every user's post in every subreddit, I created a distribution of when a user posted in a subreddit. Then, I defined a rapid decline in posts in a subreddit for a user as a distribution that is both negatively skewed and platykurtic. This describes a distribution such as the left one in figure 2. In figure 4, we see that by our definition only very few members are being fully excluded overall.⁴ In fact, only a bit over 1000 subreddits out of approximately 70,000 excluded a member at all.

4.5 Strength of Norms of Difference

The same definition also works on a group level. I calculated the average skewness and average kurtosis for a group by simply finding the mean of the skewness and kurtosis coefficients of all

⁴Used the following code: <http://www.sthda.com/english/wiki/ggplot2-density-plot-quick-start-guide-r-software>

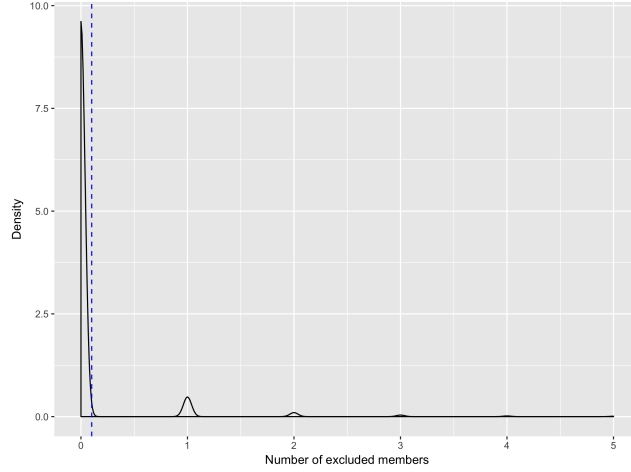


Figure 4: Density plot for the number of excluded members from groups.

members in a group. Then I created a dummy *strongnorms*, it is 1 if the average skewness and kurtosis indicate a negative skewed and platykurtic distribution. A similar dummy *weaknorms* requires a symmetric or positively skewed distribution and only looks at whether the average kurtosis indicates a leptokurtic distribution. These dummies represent strong and weak norms of difference of a group respectively. One can see that if, on average, the distribution of times of posts is platykurtic and negatively skewed, it indicates, that on average more members were fully excluded from the group. I also thought about simply measuring the proportion of members that were excluded and use that as a continuous variable of strength of norms of differences. The issue is that the proportion of excluded users is very small and therefore defining what it means to have strong and weak norms of differences is very difficult. My approach, on the other hand, is more granular and captures exclusion as well, since exclusion itself is defined in terms of skewness and kurtosis. The reason weak norms of difference are defined as having fatter tails, implies that the fall off of posts is more gradual and either partial or no exclusion at all occurred. [3]

4.6 Commitment

The final thing we want to measure is commitment of a member to a group. This measure is the most straight-forward of all. It is simply the number of posts to a specific subreddit divided by the total number of posts. For this paper, we ignore the temporal aspect of commitment and assume that commitment is constant. The reason for that is due to computational and time limitations.

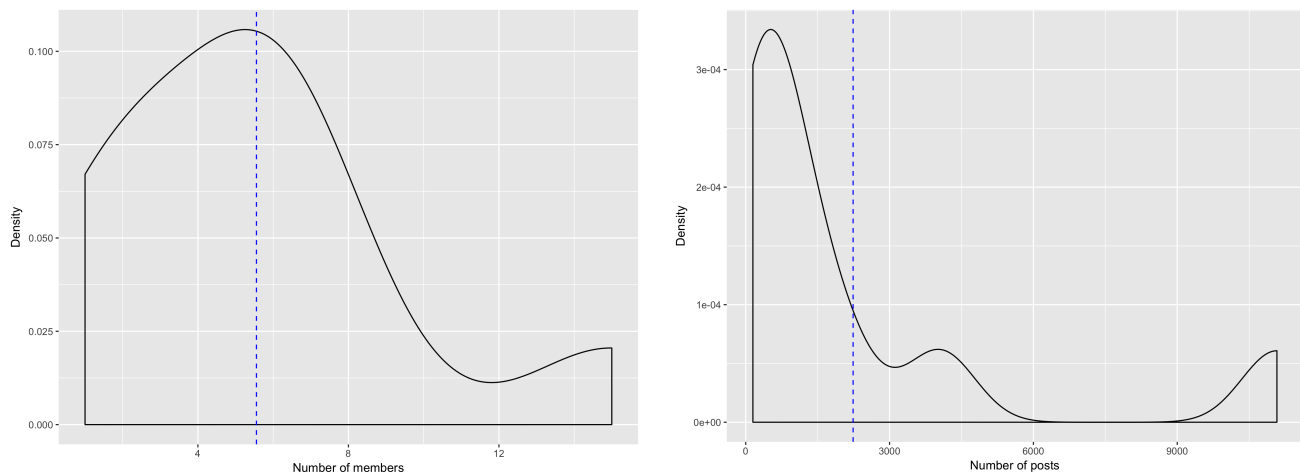


Figure 5: Densities for the number of users and number of posts for subreddits with more than 10% top-committers. Blue dashed line indicates the mean.

The results would be far more accurate and strong if the commitment would vary over time, since once a member gets excluded, commitment will be skewed low overall, even though it might be high and then lower down. Finally, we consider members of groups with over 100 posts over the two year period with a commitment of at 70% to be top-committers. It turns out, that the number of top-committers in subreddits is miniscule. In particular, the average proportion of top-committers among subreddits is 0.0001. Investigating this further yields that most subreddits with more than 10% top-committers have a very small amount of users. As figure 5 shows, the average is about 6 members, yet these members communicate a lot as we can see in the right figure.⁵ This is quite interesting, because it seems to indicate that the smallest groups are formed around highly committed users. Finally, we also care about the distribution of commitment for each individual member of a group. H2 tries to identify a connection between the commitment distribution and strength of norms of difference. To be able to do this, just like with the timestamps, I summarize the distribution of commitments using skewness and kurtosis. [3] Of particular interest to H2 is skewness as we are looking for a positive skewness. It is important to point out that overall, the commitment, as allured by the proportion of top-committers, is skewed towards 0 overall (figure 6), indicating that most Reddit users do not dedicate their time to one specific group.⁶ This is a natural observation as the quantity of groups available to each member is overwhelmingly large.

⁵Used the following code: <http://www.sthda.com/english/wiki/ggplot2-density-plot-quick-start-guide-r-software>

⁶Used the following code: <http://www.sthda.com/english/wiki/ggplot2-density-plot-quick-start-guide-r-software>

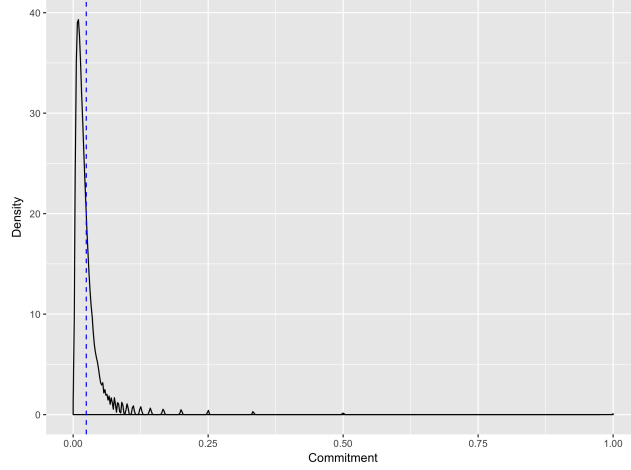


Figure 6: Density plot for commitment. Blue dashed line indicates mean.

4.7 Methodology

To test the first hypothesis I estimated four OLS regressions. In particular, the dependent variable is the number of excluded members and the independent variable of interest is the dummy *strongnorms*. Since the exclusion of a member and the strength of the norms of difference are measured by the skewness and kurtosis of the post timestamps for a particular group, the second OLS includes control variables for the number of users and the number of posts to control for the fact that as the number of users increases, the number of excluded members also increases and similarly, as the number of posts increases, the more precise the measure of skewness and kurtosis is. In addition to simply focusing on *strongnorms*, I also decided to see whether or not *weaknorms* also has an effect on the number of excluded. One can see the last two models as sort of robustness checks of my hypothesis and my definitions. I expect to see no or a negative effect of *weaknorms* on the number of excluded members, due to the fact that according to Hardin's theory, the prisoner's dilemma fringe should increase as the norms of difference weaken and therefore the number of fully excluded members should decrease. [1]

For my second hypothesis, I am interested whether or not, on average, commitment distributions skewed toward the right, that is higher commitments, is in any way related to strong norms of difference. Since my dependent variable is the dummy variable indicating whether or not a group has strong norms of difference, I decided to use a Logit model. In particular, I am estimating two Logit models, where I interact skewness and kurtosis dummies for each members commitment

distributions. These combinations of dummies categorize every possible combination of skewness and kurtosis. One of the models also includes controls for the number of users and the number of posts in the given subreddit, to control for accuracy of the commitment distribution and strength of norms of differences.

Finally, for my last hypothesis I decided to estimate an OLS regression. The dependent variable is the proportion of top-commitors (members with more than 70% commitment) and the independent variables of interest are the dummies for the strength of norms of difference. Just like in the previous models I also control for the number of users and number of posts. Note that, by definition, weak and strong norms of difference are defined disjointly, meaning a subreddit cannot be both weak and strong, therefore I can include both dummies in this regression.

5 Results

As can be seen in table 1, H1 seems to be in fact correct. Thereby, confirming Hardin’s theory that stronger norms of difference decrease the prisoner’s dilemma fringe. [1] An interesting observation comes from the two control models. It seems as if, the weak norms of difference also have a positive effect on the fringe. This goes against Hardin’s theory, because groups with weak norms of difference do not have an incentive to fully exclude members after an act of defection. This effect is only statistically significant in the model where I do not control for the number of posts and users, and even though the effect is positive in both models, we also see the coefficient get closer to 0. In addition to *strongnorms* being statistically significant in all models, the number of posts and users is too. This is expected since the definition of being excluded depends on the number of posts and also the number of users in a subreddit.

The second hypothesis (H2) has been shown to not hold. In particular, we expected a positive statistically significant effect of a positive skew on *strongnorms*, but instead table 2 shows a negative statistically not significant effect of positive skew. In fact, barely anything is statistically significant except the distribution of commitments having fat tails, but this effect also becomes not significant once we control for the number of posts and number of users. All in all, it means that using the commitment distribution we cannot make any statements regarding the strength of

<i>Dependent variable:</i>				
	Number of excluded members			
	(1)	(2)	(3)	(4)
strongnorms	0.036*** (0.008)	0.123*** (0.006)		
Number of posts		0.00003*** (0.00000)		0.00003*** (0.00000)
Number of Users		0.001*** (0.00001)		0.001*** (0.00001)
weaknorms			0.031*** (0.004)	0.005 (0.003)
Constant	0.099*** (0.002)	−0.003** (0.002)	0.092*** (0.002)	0.004** (0.002)
Observations	69,920	69,920	69,920	69,920
R ²	0.0003	0.439	0.001	0.436
Adjusted R ²	0.0003	0.439	0.001	0.435
Residual Std. Error	0.529 (df = 69918)	0.396 (df = 69916)	0.529 (df = 69918)	0.397 (df = 69916)
F Statistic	19.507*** (df = 1; 69918)	18,223.370*** (df = 3; 69916)	48.376*** (df = 1; 69918)	17,980.260*** (df = 3; 69916)
<i>Note:</i>				
*p<0.1; **p<0.05; ***p<0.01				

Table 1: Results from OLS regression for the first hypothesis. [5]

	<i>Dependent variable:</i>	
	strongnorms	
	(1)	(2)
Mesokurtic	−9.355 (139.277)	−11.319 (378.468)
Leptokurtic	−0.727*** (0.165)	−0.136 (0.166)
Moderate positive skew	−0.117 (0.236)	−0.077 (0.236)
High positive skew	−0.235 (0.270)	−0.181 (0.270)
Number of users		−0.001 (0.001)
Number of posts		−0.002*** (0.0002)
Mesokurtic, Moderate positive skew	9.068 (139.278)	11.053 (378.468)
Leptokurtic, Moderate positive skew	1.109 (0.780)	0.523 (0.780)
Mesokurtic, High positive skew	9.274 (139.278)	11.250 (378.468)
Constant	−2.211*** (0.215)	−2.054*** (0.215)
Observations	38,023	38,023
Log Likelihood	−6,754.520	−6,264.619
Akaike Inf. Crit.	13,525.040	12,549.240
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

Table 2: Logit results for the second hypothesis. Note that if a combination of skewness or kurtosis is not presented, it means it was not present in the data. The skewness and kurtosis dummies describe the commitment distributions of individuals in the subreddit. [5]

norms of difference. In other words, just because a group has strong norms of difference, it does not mean that it necessarily has, on average, higher committed members. This is very interesting, because we can rule out that highly committed members are the root cause of stronger norms of differences.

Finally, the last hypothesis tries to tackle H2 from the other way. Does strength of norms of differences have an effect on the proportion of highly committed members. The answer is yes and no. Table 3 shows a very interesting result. H3 is shown to be wrong, insofar it is flipped.

	<i>Dependent variable:</i>
	Proportion of top-commitors
Number of users	−0.00000*** (0.00000)
Number of posts	0.00000*** (0.000)
weaknorms	0.0001*** (0.00004)
strongnorms	−0.0001 (0.0001)
Constant	0.0001*** (0.00002)
Observations	69,920
R ²	0.005
Adjusted R ²	0.005
Residual Std. Error	0.005 (df = 69915)
F Statistic	84.807*** (df = 4; 69915)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

Table 3: Results from OLS for the third hypothesis. A top-commitor is a member of a subreddit with a commitment of greater or equal to 70%. [5]

We see that a group with strong norms of differences has no statistically significant effect on the proportion of top-commitors. On the other hand, groups with weak norms of differences have a positive, statistically significant effect. The logic behind it could be as follows: Since the norms of

difference are weak, there is less pressure to assimilate and therefore less rules to follow and more people are willing to join and spend a lot of time in that group. It is crucial to point out that Reddit is filled with such groups. For example, take a "videos" subreddit. Here, there is no other rule than to just post and comment on videos one enjoys. Clearly, a lot of people would be very active on this page by commenting and sharing their opinions on the videos. In other words, I am questioning the strength of this result insofar, the Reddit data is biased towards weak norms of difference groups with high proportions of highly committed members.

6 Future Research

In this paper, I focused solely on the distribution of posts among users in different subreddits, but the data set is far richer. In addition to the information that I am using, one also has access to the "score" of a comment (meaning the sum of +1 and -1 votes) as well as the actual text of the comment. This offers itself for text analysis and in particular sentiment analysis. With this, one could try to identify words or phrases that lead to exclusion in specific groups. Looking at it from another perspective, one could in fact try to identify using text analysis whether or not someone was in fact banned or excluded from a group. Finally, since everyone can respond to everyone in Reddit, a network analysis would offer an even more detailed approach to identifying and distinguishing groups from each other. For instance, one could look at a dynamic social network analysis to identify subgroups and the core members of each subgroup. [2] Once these core members are identified, some sort of inference of leadership can be done and then measure authority and see what properties a leader of a certain groups needs to have.

7 Conclusion

In this paper, I have shown that Hardin's theory of a prisoner's dilemma fringe does in fact exist. I identified a data set that can be used in future group conflict literature as it is immense in size and detailed. More importantly, this paper, empirically verified formal theoretic results in the group conflict literature and identified that commitment in groups does not have an effect on the

strength of norms of difference. In that, it showed an approach to quantify strength of norms of difference using skewness and kurtosis of commitment and timestamp distributions.

In this paper, I struggled the most with actually finding proper definitions for the concepts Hardin described in his book. In addition, due to the large data set, I had to cut the sample size down multiple times and I fear that due to the initial selection, the sample might be biased.

References

- [1] Hardin, Russell. One for all: The logic of group conflict. Princeton University Press, 1997.
- [2] Sekara, Vedran, Arkadiusz Stopczynski, and Sune Lehmann. "Fundamental structures of dynamic social networks." *Proceedings of the national academy of sciences* 113, no. 36 (2016): 9977-9982.
- [3] Brown, Stan. Measures of Shape: Skewness and Kurtosis. 2008. <https://brownmath.com/stat/shape.htm#WhatsNew>
- [4] Google BigQuery. Reddit Comments Dataset. https://www.reddit.com/r/bigquery/comments/3cej2b/17_billion_reddit_comments_loaded_on_bigquery/
- [5] Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.