The impact of individual and collective variables on attidues towards homosexuality in collectivist and individualistic societies

This report examines the impact of individual and collective variables on attitudes towards homosexuality in collectivist and individualistic societies (India and the US). The dataset has been derived from a social-attitudes survey conducted by PEW Research Centre in 2019. The dataset is prepared by a rigorous data cleaning strategy which includes using feature selection significance tests such as chi-square. The clean dataset was split into two by country: India and the US. The data was split into train and test sets and then used to train and validate a simple classifier model using the Logistic Regression method. It was concluded that religion, education and income, were the strongest predictors governing attitudes towards homosexuality in both countries. The same variables appear to be operating gin both societies regardless of cultural status (i.e. collectivist vs individualistic). Specific religious groups appear to be more likely to express negative attitude towards homosexuality.

Babak Hessamian
babak.hessamian@city.ac.uk

I. INTRODUCTION

There is much psychological research locating the source of negative attitudes towards minorities (in our case lesbian, gay and bisexual people) within the individual [1]. However, there is also evidence that societal factors, such as political and economic instability, predict negative attitudes. The present research examines the contribution of individual and collective factors to explaining attitudes towards homosexuality in two distinct societies. We have chosen Indian and The United States (US) as our two countries due to their collective and individualistic societies respectively. This can help inform interventions for tackling prejudice in these societies.

II. ANALYTICAL QUESTIONS AND DATA

A. *Data*

The dataset is from the Spring 2019 Global Attitudes Survey by Pew Research Centre. Pew Research Centre is a nonpartisan fact-tank that informs the public about the issues, attitudes and trends shaping the world. The surveys for each individual country were conducted in 2019. There are 607 variables which are multiple choice survey questions. Several of the questions are country-specific, for instance questions about specific political parties in any given country. There are 38,426 data points (respondents) in 34 countries, so the sample size for each country is approximately 1,000 observations.

B. *Analytical Questions*

The following analytical questions are addressed:

1. What effect do collective variables, namely political and economic instability, have on attitudes towards homosexuality?
2. What effect do individual demographic variables, namely gender, age, education, and religiosity, have on attitudes towards homosexuality?

Our hypothesis:

In collectivist societies, such as India, collective variables, such as family values, will be better predictors of attitudes towards homosexuality than individual variables.

In individualist societies, such as the US, individual variables, such as religiosity, will be better predictors of attitudes towards homosexuality than collective variables.

III. ANALYSIS

Our methodology was split in to the following four steps:

A. *Data preparation*

Our dataset came in a .sav (SPSS file) format was read and explored. We used domain knowledge to choose 18 independent variables that are best suited to answer our analytical questions. Our dependent variable is the response to the following survey question:

> Q. Which one of these comes closer to your opinion? Homosexuality should be accepted by society OR Homosexuality should not be accepted by society

The possible answers were:

1.Homosexuality should not be accepted by society
2.Homosexuality should be accepted by society
3.Dont know
4.Refused

We omitted responses with 3 and 4 and any missing responses so we end up with a binary variable. At this stage our dataset has 3286 observations and 18 variables. We also dropped similar non-answers for all other variables, as these would not have contributed to our analysis.

Several of our survey questions has responses that were on a scale. These can be viewed as ordinal

variables. For this reason, we mapped the responses on a numeric scale for these variables.

B. Data derivation

We investigated the distribution of Age grouped by responses to our dependent variable and discovered outliers as well as an unusual values in the lower quartile "Fig. 1". We decided to only include those where the respondent is 18 years and over and then bin the observations into separate age groups. This was considered to help with interpreting the results of our model.

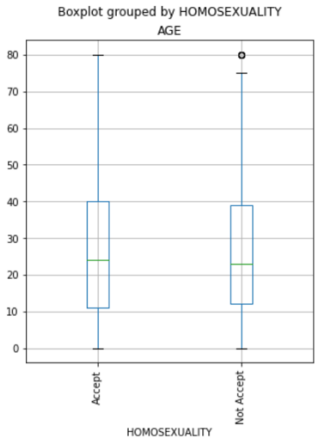

Fig. 1. Boxplot of age disribution.

We had two variables where the responses were separate for the two countries: income and education. A decision was made to combine these two columns and re-bin the responses into a scale that is then shared between the countries. In this way it would be possible to compare income and education between the two countries. For instance income was binned into five categories for both countries between 'very low' to 'very high'.

Religion was a separate question for each country as well. So they needed to be combined. We had to then re-bin these to avoid duplicates.

C. Construction of models

After the steps above, the main feature selection was done by performing a chi-square test on the remaining variables. Those variables with a p-value of 0.05 and below were chosen. We used the 2x2 contingency table used for chi-square test to calculated the effect size Cramér's phi (Cramer's V). The most common interpretation of the magnitude of the Cramér's phi is as follows:

Small Effect Size: $V \leq 0.2$
Medium Effect Size: $0.2 < V \leq 0.6$
Large Effect Size: $0.6 < V$

This analysis resulted in us dropping a further five features. Our final list of features is shown in "Fig. 2".

Variable Name	*Survey Question and possible answer*	*p-value*
ECON_SIT	Now thinking about our economic situation, how would you describe the current economic situation in (survey country) – is it very good, somewhat good, somewhat bad, or very bad	0.0001672790977061 0297
COUNTRY_SATIS	Overall, are you satisfied or dissatisfied with the way things are going in our country today?	0.0045
CHILDREN_BETTEROFF2	When children today in (survey country) grow up, do you think they will be better off or worse off financially than their parents?	0.00013
SATISFIED_DEMOCRACY	How satisfied are you with the way democracy is working in our country – very satisfied, somewhat satisfied, not too satisfied, or not at all satisfied?	5.0112259807 20863e-06
RELIG_GOD	Please tell me whether you completely agree, mostly agree, mostly disagree or completely disagree with the following statements. b. God plays an important role in my life	9.1816946130 59763e-25
EDUCATION	What is the highest level of school you have completed or the highest degree you have received? Answer changed to scale from very low to very high.	4.5705854554 43555e-18
INCOME	Would you mind telling me your households total approximate annual income, counting all wages, salaries, pensions and other incomes that come in, before taxes and other deductions? Answer changed to scale from very low to very high.	2.0366947301 234373e-15
RELIGION	What is your current religion, if any? Possible indcluded all majority religions in the two countries.	2.0366947301 234373e-15

Fig. 2. Table of variables used in model building.

Having selected the futures we dummy-coded the categorical variables and fitted a logistic regression model to the whole dataset. We then drew a correlation matrix which revealed a strong correlation between the two countries "Fig.3".

We expected this as it shows there is strong difference in attitudes towards homosexuality in the two countries. This is further supported by the bar chart in "Fig.4".

To address this issue and to work towards answering our analytical question which involves comparing attitudes towards homosexuality, we decided to train two separate models for each country and then identify and compare the strongest predictors in each country. Splitting the dataset on country resulted in two datasets that were similar in size.

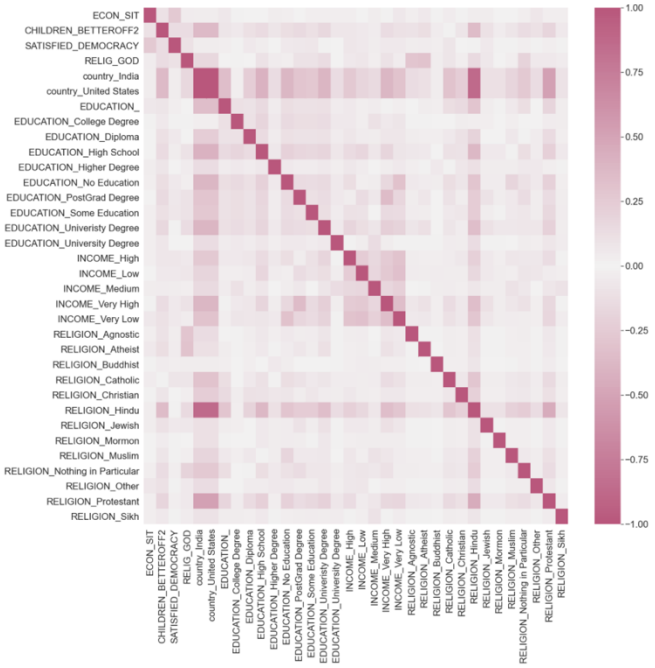

Fig. 3. Correlation matrix of variables used before splitting the data by country to fit our models.

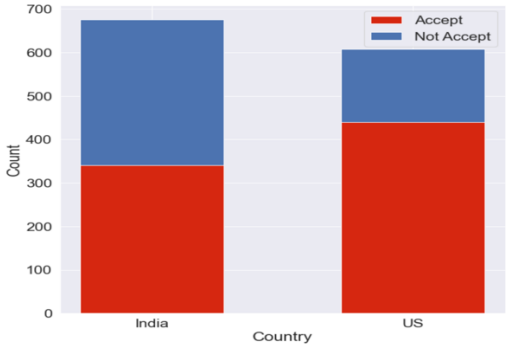

Fig. 4. The differnce in attitudes towards homosexuality in the US and India.

D. Validation of results

For each country we split our data into input and target sets. We then split each set randomly into a training (80%) and testing (20%) set to be able to validate the performance of the models we build.

"Fig.5" shows the confusion matrix for our Indian subset, and "Fig.6" shows the confusion matrix for our US subset. These two models are trained on data and then validated on unseen data, then their performance measured and quantified.

Judging by the accuracy measure it may appear that the model fit for the US data, "Fig. 6", is better at 76% accuracy. However in looking at the confusion matrices we may infer that the the model has a high potential to misclassify observations as false positives, while missing true positives. In other words the US model is more likely to classify observations as negative. This may be due to multicollinearity of some of our variables in the RELIGION column.

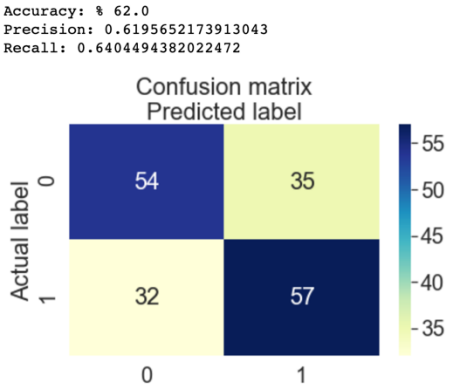

Fig. 5. Confusion matrix and preformance metrics for the Logistic

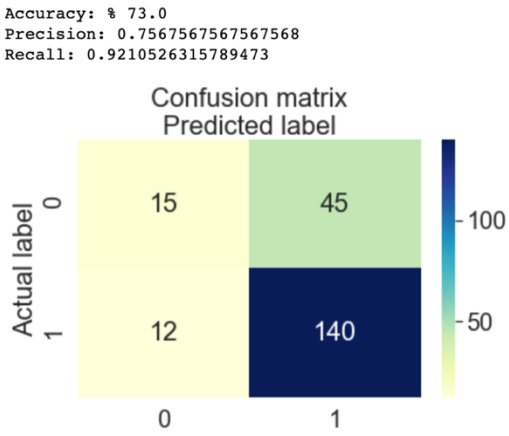

Fig. 6. Confusion matrix and performance metrics for the Logistic Regression model fitted to our US subset.

IV. FIDINGS AND REFLECTIONS

Once the models were fitted and validated we observed the coefficients for each of the variables for each country. These are available in the python notebook that accompanies this report. In comparing the coefficients we found the following. We will refer to 'Accept' responses as having positive attitude towards homosexuality and 'Not Accept' responses as having negative attitude towards homosexuality.

Believing that God plays an important role in the respondents life had an impact on the outcome for both countries. Religion was the strongest predictors for attitudes towards homosexuality in both countries. In the US the strongest religious predictors for positive attitudes towards homosexuality were Agnostic and an Atheist, while

in India Sikhs and Buddhists were more accepting of homosexuality. This is also consistent with our group-by-religion histogram that we plotted in the exploratory phase of our study, “Fig. 7”.

We found that education level was a stronger predictor in India that in the US. In India survey respondents with higher degrees, such as Postgraduate and PhD, were less accepting of homosexuality and those with some or no education had the most positive attitude towards homosexuality. This is surprising as we expected this to have less of an effect in India since this variable is considered an individual demographics variable and our assumption was that India is a collectivist society.

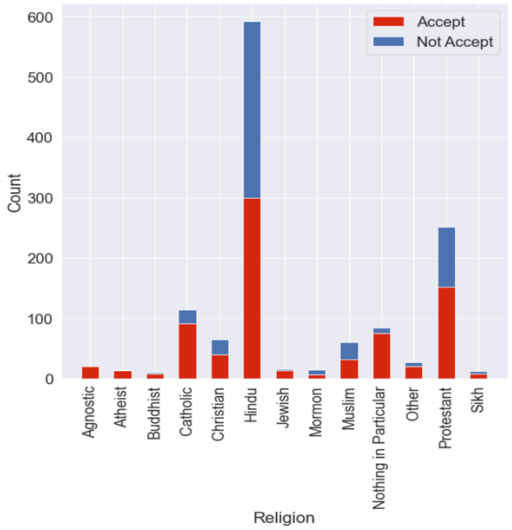

Fig. 7. Group-by-religion histogram.

Income was a relatively strong predictor for both countries, but to a lesser extent than religion or education. In India participants on higher incomes tended to have a more positive attitude towards homosexuality. We plotted a histogram of income across two countries after categorising them into discrete bins which supports this finding, “Fig. 8”.

Looking at our strongest predictor religion, we found that Mormons, Muslims and Protestant in the US and Protestants and Catholics were amongst the respondents who had the least acceptance towards homosexuality.

We also found that those less satisfied with the economic situation of their respective countries and less optimistic about the financial future of their children, are marginally less accepting of homosexuality.

We set out to investigate the effect of collective variables versus individual demographic variables on attitudes towards homosexuality. We have found that religion, education and income, are the strongest predictors governing attitudes towards homosexuality in both countries. These are all individual demographic variables. As such, we found no evidence that there is a difference in predictors in a society considered to be individualistic versus one that is considered to be collectivist.

Work to reduce prejudice should therefore focus on certain religious groups in both countries.

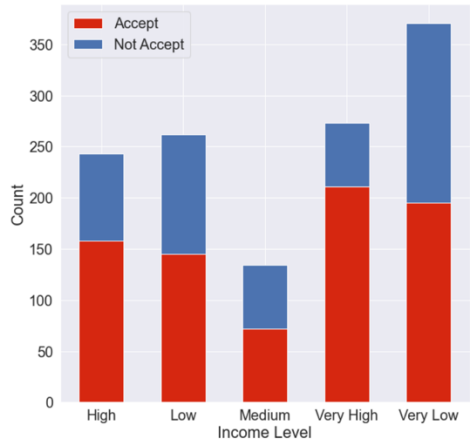

Fig. 8. Group-by-income histogram.

V. FURTHER WORK

There were limitations with our study. Most notably the re-binning of ordinal variables may have cause collinearity. A deeper domain knowledge about the socioeconomics of the two countries would perhaps yield a better result for model accuracy.

For further work it would be interesting to observe the temporal changes in attitudes towards homosexuality by combing this dataset with other social attitudes datasets gather by PEW Research.

References

[1] Herek, GM. “The Psychology of Sexual Prejudice” in Current Directions in Psychological Science. vol. 9, pp 19-22, February 2000.

Section	*Word Count*
Abstract	148
Introduction	94
Analytical Questions and Data	184
Analysis	994
Finding and Reflections	425
Further work	66
Total	1911