Analysing Migrant categorizations and European public opinion: Diverging attitudes towards immigrants and refugees

Srivatsa B : bmat2030
Ramkrishna Jyoti Samanta : bmat2020
Srigyan Nandi : bmat2029
Yadvendra Kumar Yadav : bmat2034

Abstract

In this paper, we attempt to reproduce and extend the analysis in Dr David De Coninck's paper - "Migrant categorizations and European public opinion: diverging attitudes towards immigrants and refugees". This paper is a review of the methods we pursued and the sources we consulted to test certain statistical hypotheses related to the paper to bring this project to fruition.

1 Introduction

Since 2011, Europe has been faced with the largest refugee crisis since World War II. As conflicts in the Middle East and Africa have become more intense, refugees from these war torn places have been forced to embark on the long and risky journey to Europe. In the recent years, Europe has been under immense pressure to accept these refugees as public opinion towards such group has become increasingly polarizing. Western Europe has also been accepting an increasing number of immigrants(also termed economic migrants) since the beginning of the 21st century. Recent studies have shown that a significant number of Europeans tend to hold negative attitudes towards these groups of newcomers.

Terms like 'refugees' and 'economic migrants' have been applied to the newcomers since the beginning of the refugee crisis to deny their claims to protection and asylum. As a result, refugees in actual need of protection have been dismissed as economic migrants by several European leaders.Along with this, the newcomers have been subject to heavy media framing which in many cases has suggested that they settle in they first suitable country they come across. Newcomers have mostly been represented in a negative or stereotypical manner by the news media. When portrayed by the media, the newcomers are often linked to criminal activities and are represented as intruders threatening the European way of life. They are also portrayed in a way that suggests that they are a social and economic burden to society. These negative representations by the media are likely to have driven significant shifts in public opinion towards the newcomers as indicated by the increase in xenophobia and Islamophobia in Europe over the past few years.

It is clear that the use of such categorizations presents a lot of legal and social consequences for the newcomers. It is also known that such categorizations and framing tend to shape and influence public opinion towards these groups. There are many cases where some newcomers, depending on the label they are given are entitled to and enjoy certain rights, while others don't. Those labelled 'refugees' are the usual groups who directly benefit from such rights. However in many cases the categorization is very fuzzy and it is unclear as to which group should be put under a certain category.

2 Hypotheses

From the data, the author expects that people hold more positive attitudes towards refugees than they do towards migrants. The reason for this is that migrants have long since been negatively framed as taking advantage of European welfare states' benefits, while the framing of refugees also focuses on their dire situation. This leads to the following hypothesis:

Hypothesis 1: Attitudes towards refugees are more positive than attitudes towards immigrants.

However, the author is also interested in investigating differences within categories. Because of the emphasis on economic elements, the author expect more positive attitudes exist towards immigrants or refugees from wealthier countries, as they may be expected to hold more economic capital than immigrants or refugees from poorer countries. The author also expects people to hold more positive attitudes towards immigrants or refugees with the same ethnicity as most of the native population as opposed to those with a different ethnicity, and from countries within Europe rather than from countries outside Europe. This leads to the following hypotheses:

Hypothesis 2: Attitudes towards immigrants or refugees from wealthy countries are more positive than attitudes towards immigrants or refugees from poorer countries.

Hypothesis 3: Attitudes towards immigrants or refugees with the same ethnicity as most of the population are more positive than attitudes towards immigrants or refugees with a different ethnicity than most of the population.

Hypothesis 4: Attitudes towards immigrants or refugees from European countries are more positive than attitudes towards immigrants or refugees from non-European countries.

Additionally, in an attempt to extend the work of the original paper we are interested in comparing and categorizing attitudes towards immigrants and refugees based on psychological indicators and trust in news media, in an attempt to:

- *Locate differences in favourable and unfavourable attitudes towards immigrants and refugees across different categories of psychological indicators.*

- *Compare differences in attitudes towards immigrants and attitudes towards refugees across different categories of psychological indicators and trust in news media.*

3 Data And Methodology

An online survey was conducted among adults between 18 and 65 years of age, in Belgium, France, Sweden and the Netherlands, in the month of September and October, 2017. This survey lasted for a duration of about three weeks, at the end of which a total sample size of six thousand respondents (1500 per country) was reached. An online survey was chosen (as opposed to a pen and paper survey or a hybrid one) because it was found to be more cost efficient. Also, the polling agency which was responsible for conducting the survey was found to be more active and popular in the four countries chosen, hence the survey was limited to these countries. The response rate was calculated to be around 35% and the responses were weighted by gender and age in order to make sure that the data was representative across each country. The online questionnaire was sent to the respondents using email. The survey was distributed in the official language of the respective country or region that respondents were citizens of. Translations of the survey were carried out by professional translators, and this made sure that the terminology present in the questions is considered 'everyday language' by the respondents. Respondents were unable to skip questions, but some of the questions did have a 'No answer'-option. However, the questions relating to attitudes towards immigrants and refugees did not have a 'No answer' option. Each question in the survey was presented on a different page, and respondents did not have the option to return to previous questions and change their answer.

3.1 Demographics

Respondents had to indicate their gender (1 - male, 2 - female); age is calculated based on their year of birth, religious denomination was categorised in four sections, namely, Christian, Muslim, other denomination and not religious. Educational attainment is measured by the highest level of education attained by the respondent. This variable had four possible categories, namely, no or primary education, secondary education and tertiary education. Migration back-ground was constructed using information collected about the respondent's parents' and grandparents' country of birth. To construct this variable, respondents were first asked to indicate in which country each parent and grandparent was born. If both parents, one parent and at least two grandparents, or more than two grandparents of a respondent were born outside of the country the respondent currently resides in, they were considered to have a migration background. Depending on the country of origin of the respondent, a distinction was made between the respondent having a European migration background (if he/she was born in a European country) vs. having a

Non-European migration background (if he/she was born in a Non-European country).

The plots present in this report are all descriptive. The questionnaire was designed in such a way that many of the answers were to be entered on a Likert scale. A Likert scale is a five (or seven, eleven, etc.) point scale which is used to allow the individual to express how much they agree or disagree with a particular statement. Diverging stacked bar charts and Divided stacked bar charts provide us with an excellent way to visualize responses on a Likert scale. Barplots, diverging stacked bar charts and divided stacked bar charts are used for all the visualizations in this report.
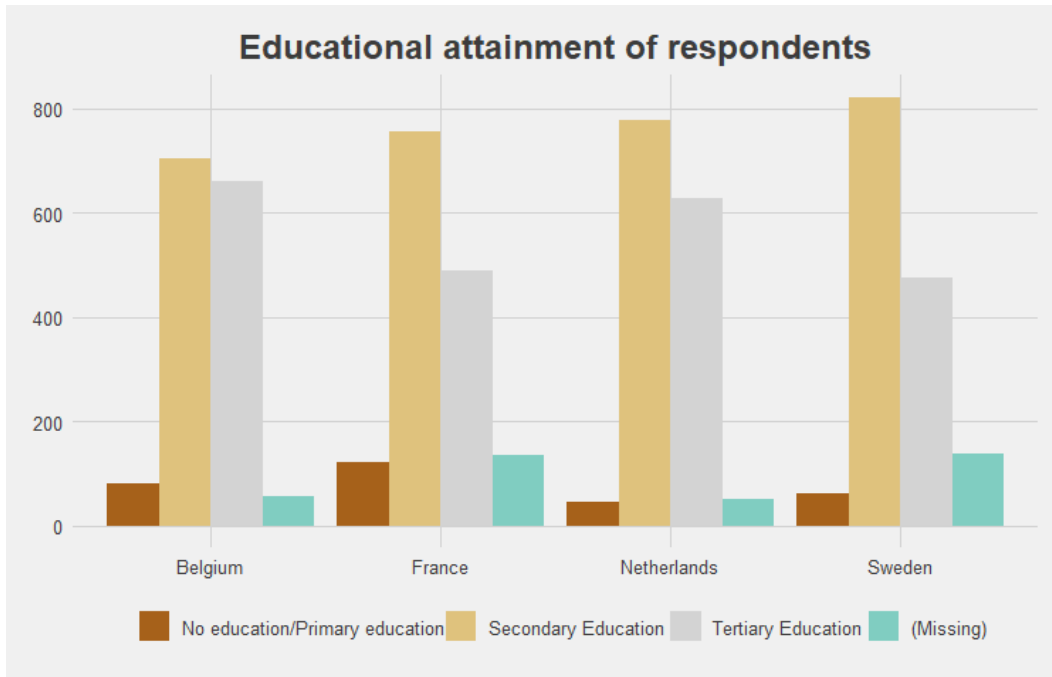

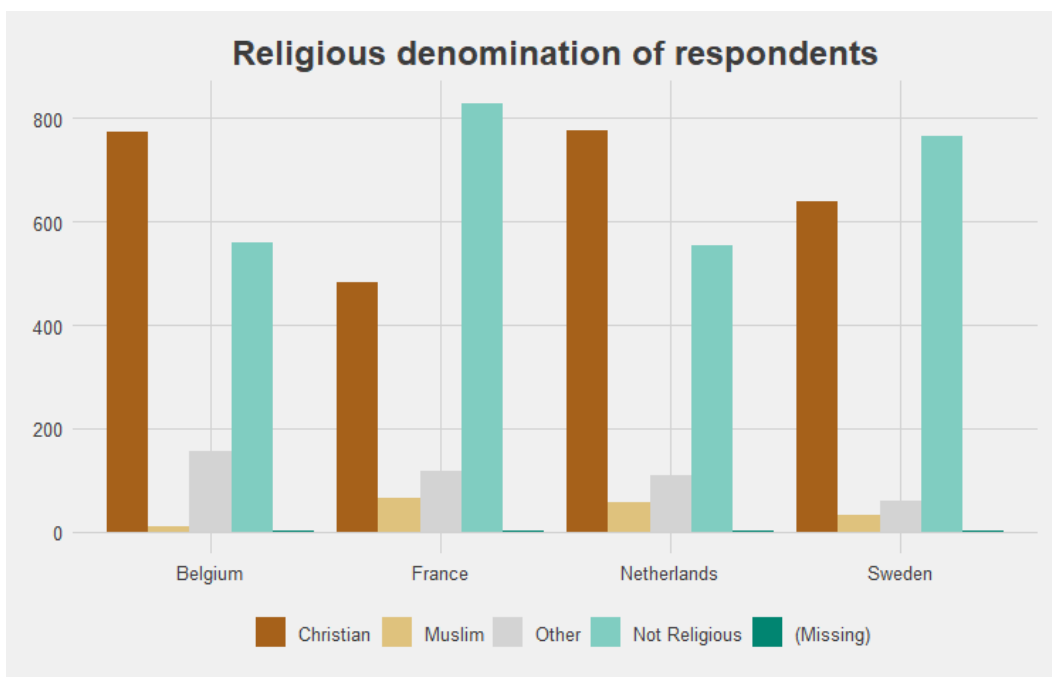Figure 1: Highest level of education attained

Figure 2: Religious denomination

3.2 Psychological Indicators

Respondents were asked to indicate how much they identified with certain psychological indicators about themselves, namely:

- Outgoing / Enthusiastic
- Critical / Confrontational
- Reliable / Disciplined
- Tense / Easily in a bad mood
- Open to new experiences / Profound
- Reserved / Quiet
- Empathetic / warm
- Disorganized / sloppy
- Calm / stable
- Conservative / not creative

For each of the categories, the respondents were asked to indicate how much they identified with the category in question. Answer categories ranged from 1 to 5, namely; 1 - Do not agree at all, 2 - Somewhat disagree, 3 - Neutral, 4 - Somewhat agree and 5 - Fully agree.
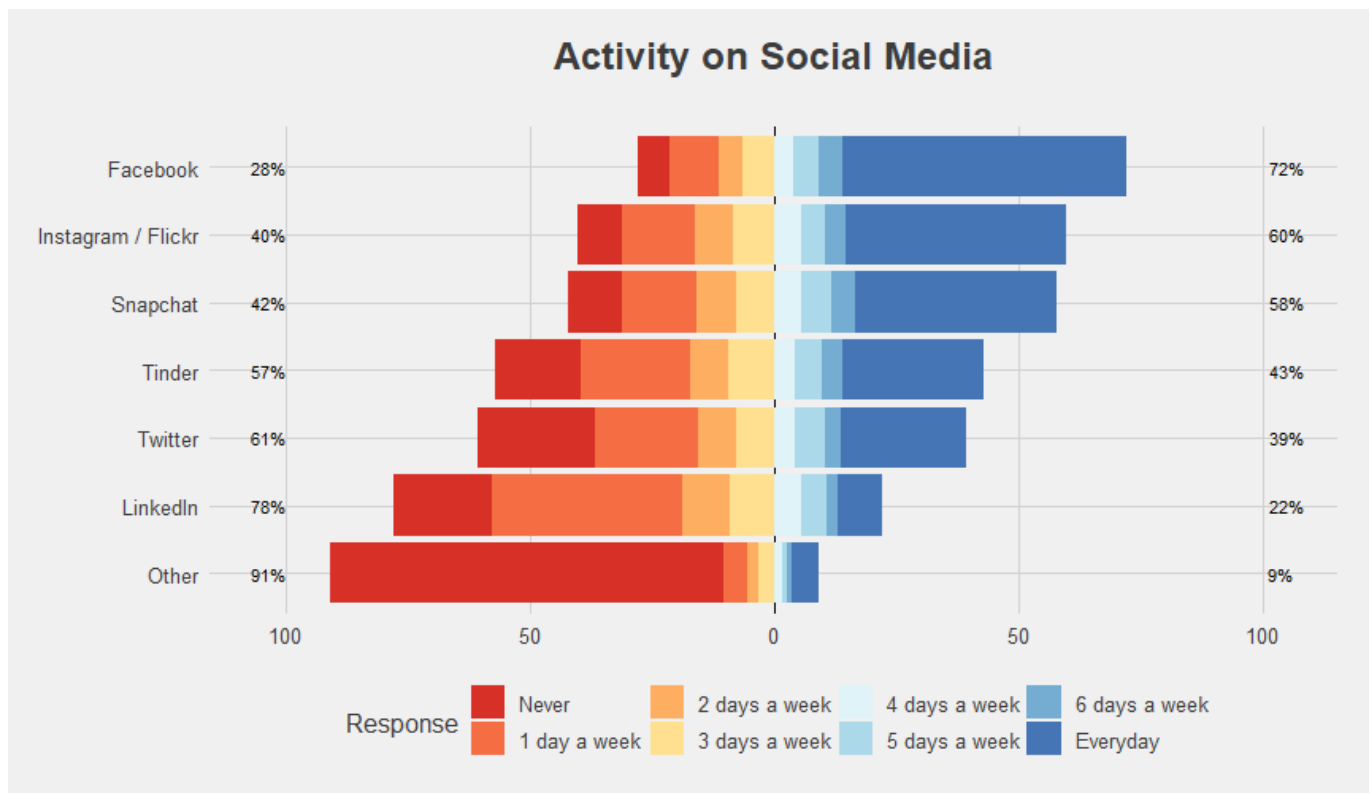

Figure 3: Social Media activity

3.3 Attitudes Towards Immigrants And Refugees

In order to accurately measure the difference in public opinion on immigrants and refugees, a decision to adapt a scale previously used in a rotating module of the European Social Survey had been made. The adapted scale consists of seven items asking which groups of immigrants should be allowed to come and live in Belgium: 'Immigrants of the same race or ethnicity as most of the country's population'; 'Immigrants of a different race or ethnicity as most of Belgium's population'; 'Immigrants of the richer countries in Europe'; 'Immigrants of the poorer countries in Europe'; 'Immigrants of the richer countries outside Europe'; 'Immigrants of the poorer countries outside Europe'; and 'Immigrants coming from Muslim countries who wish to work in Belgium' Answer categories range from 1 - meaning 'Allow none', to 4 - meaning 'Allow many'. To collect the data, the scale was presented in its original form and an extra item concerning immigrants from Muslim countries was added. The reason for the inclusion of this item lies in the fact that a majority of immigrants and refugees entering Europe in the current refugee crisis originate from Syria, Iraq or Afghanistan e predominantly Muslim countries. Prior to completing each block of items, respondents were presented with a definition of immigrants and refugees from the United Nations with the request they keep this definition in mind during completion of the survey.

The definition of immigrants was as follows:

> *An immigrant should be understood as covering all cases where the decision to migrate is taken freely by the individual concerned, for reasons of 'personal convenience' and without intervention of an external compelling reason (e.g., war, natural disaster). (UNESCO, 2017, para. 3)*

The definition of refugees was:

> *A refugee is someone who has been forced to flee his or her country because of persecution, war, or violence. A refugee has a well-founded fear of persecution for reasons of race, religion, nationality, political opinion or membership in a particular social group. (United Nations, 1951, p. 14)*
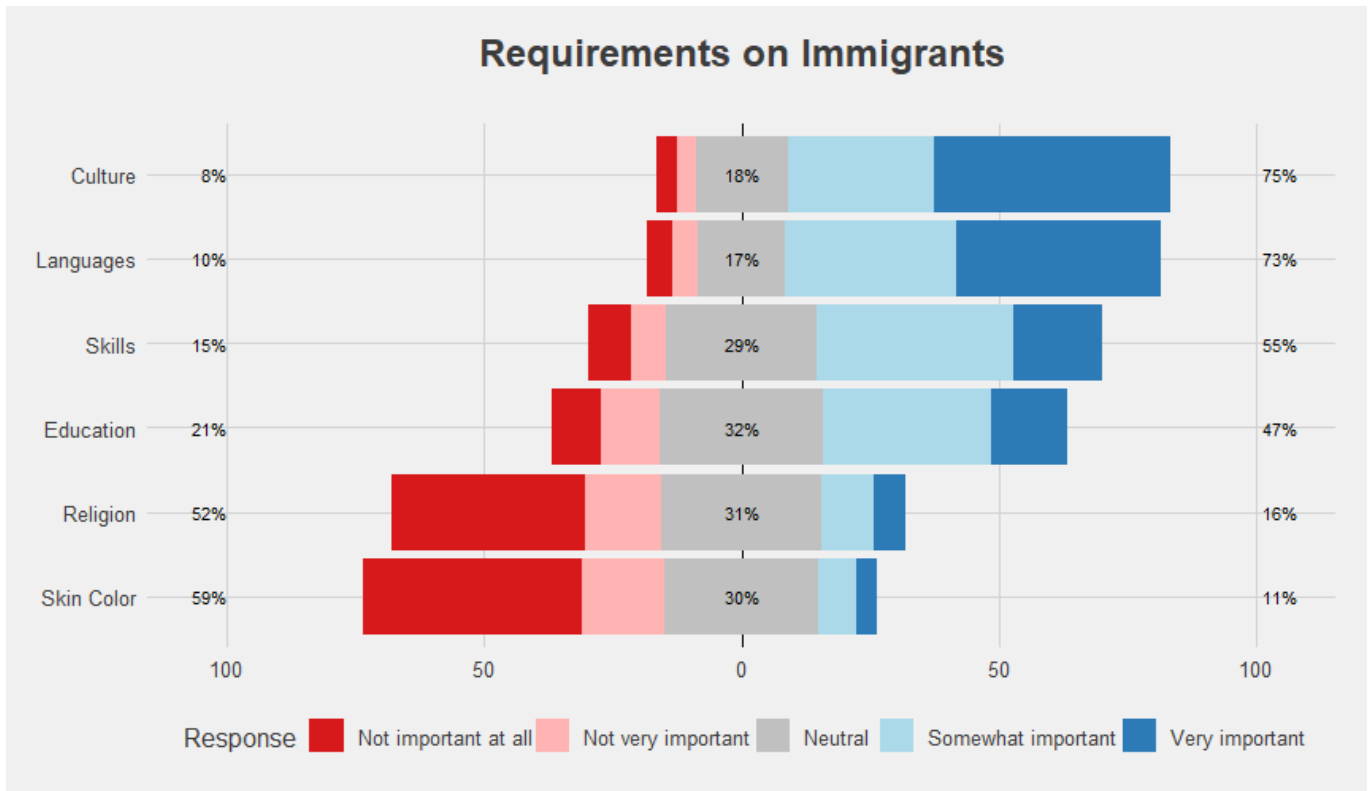

Figure 4: Requirements on incoming immigrants (Exact questions can be found in the codebook, see References)
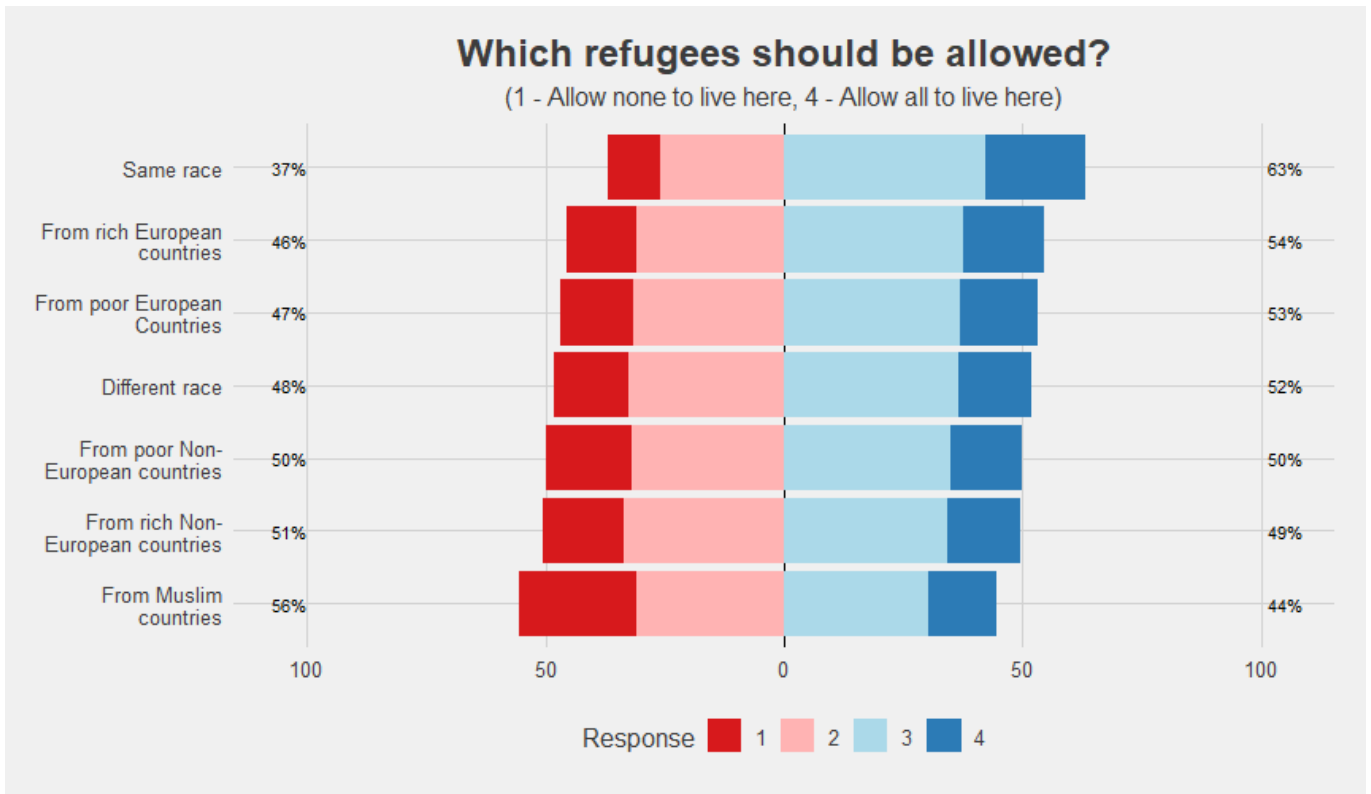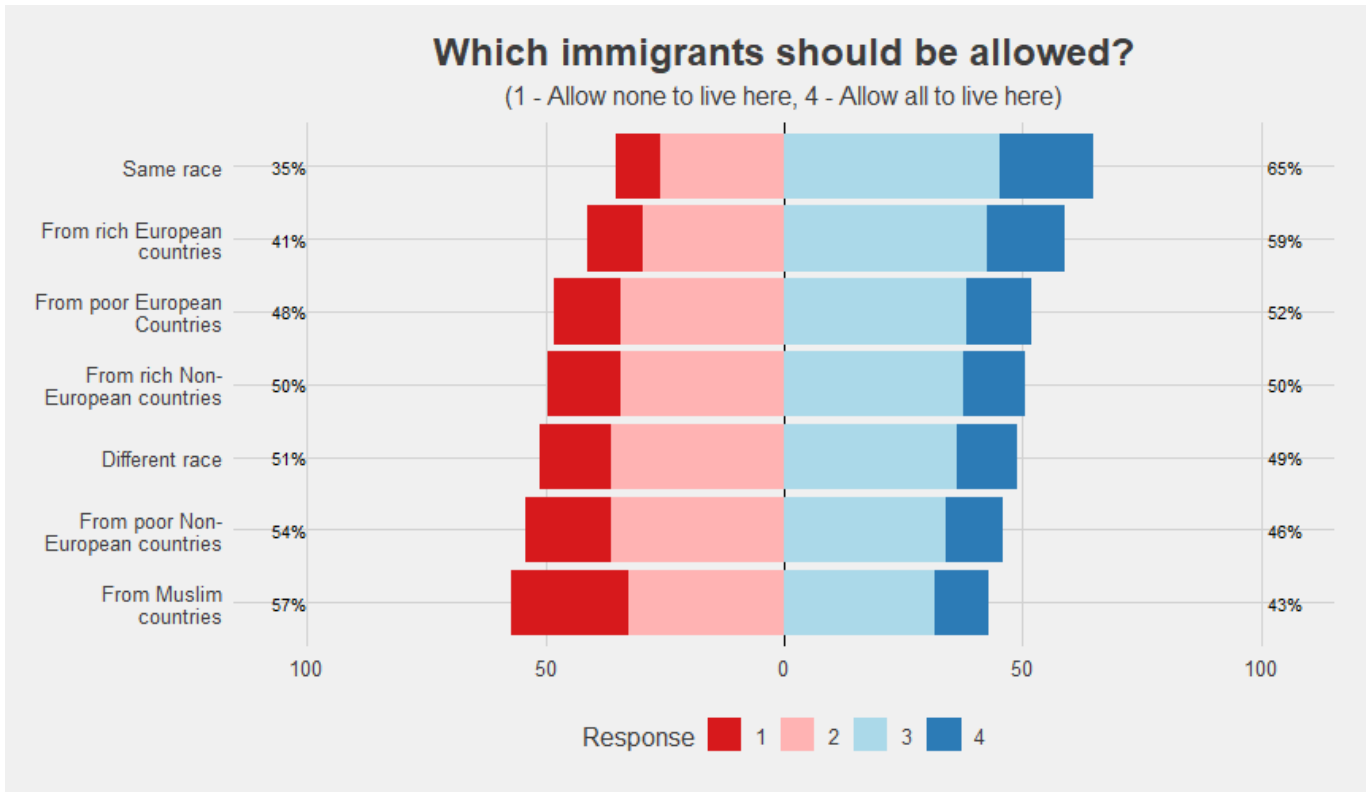
Figure 5: Refugees - Attitude indicators

Figure 6: Immigrants - Attitude indicators

6

3.4 News Media Consumption

Respondents were asked about their media consumption patterns during the past month, with answer categories ranging from 0 (Never) to 7 (Every day). Both television and radio consumption were split into two groups: public and commercial broadcasters. For the classic newspaper and online news consumption, the most commonly read newspapers and commonly visited web-pages in each region were included separately. This selection of newspapers was based on information concerning the circulation of newspapers in each country. Online and social media were also included. For online media, most of the websites of the newspapers that were included were selected, with some additional online-only news outlets. As for social media, the most commonly used media in 2017 were included. Trust in the news media brands was measured by means of a five-point scale with answer categories ranging from 1 (No trust at all) to 5 (A lot of trust). Like the methodology for news media consumption, trust was measured separately for public and commercial television and radio. Trust in newspapers was aggregated into two groups: trust in quality newspapers and trust in popular newspapers to avoid multi-colinearity in the data. Trust in news websites/apps, and trust in social media were also measured on the same scale.

News media consumption variables were measured by asking the respondents about the number of days in the past week during which they consumed news via public television broadcasters, commercial television broadcasters and quality newspapers, where the allowed answers were 1 - indicating 'never' and 8 - indicating 'every day'. For print media, potential issues arose when it came to the 'quality' of newspapers, as this is a subjective issue and hence it differs from person to person. To handle this, a country-specific selection of newspapers provided, without labelling them as tabloids or quality newspapers – these categorizations were done after the responses were received. Regarding televised media, relevant country-specific examples of each type of broadcaster were provided. The survey also asked to what extent respondents were exposed to news about refugees in the past year, and the allowed answers were 1 - indicating 'never' and 6 - indicating 'very often'. A summary can be found in Table 1.
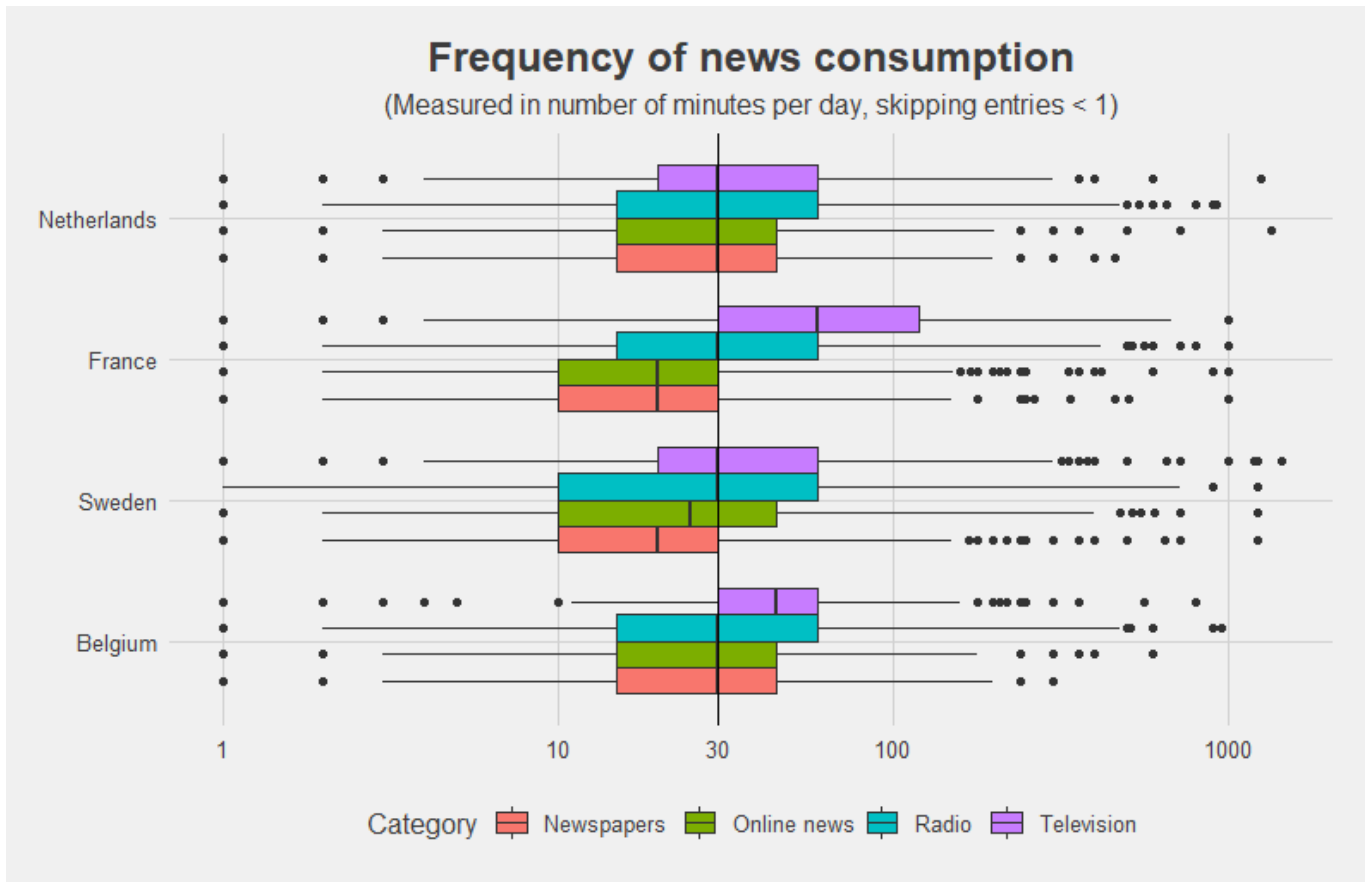

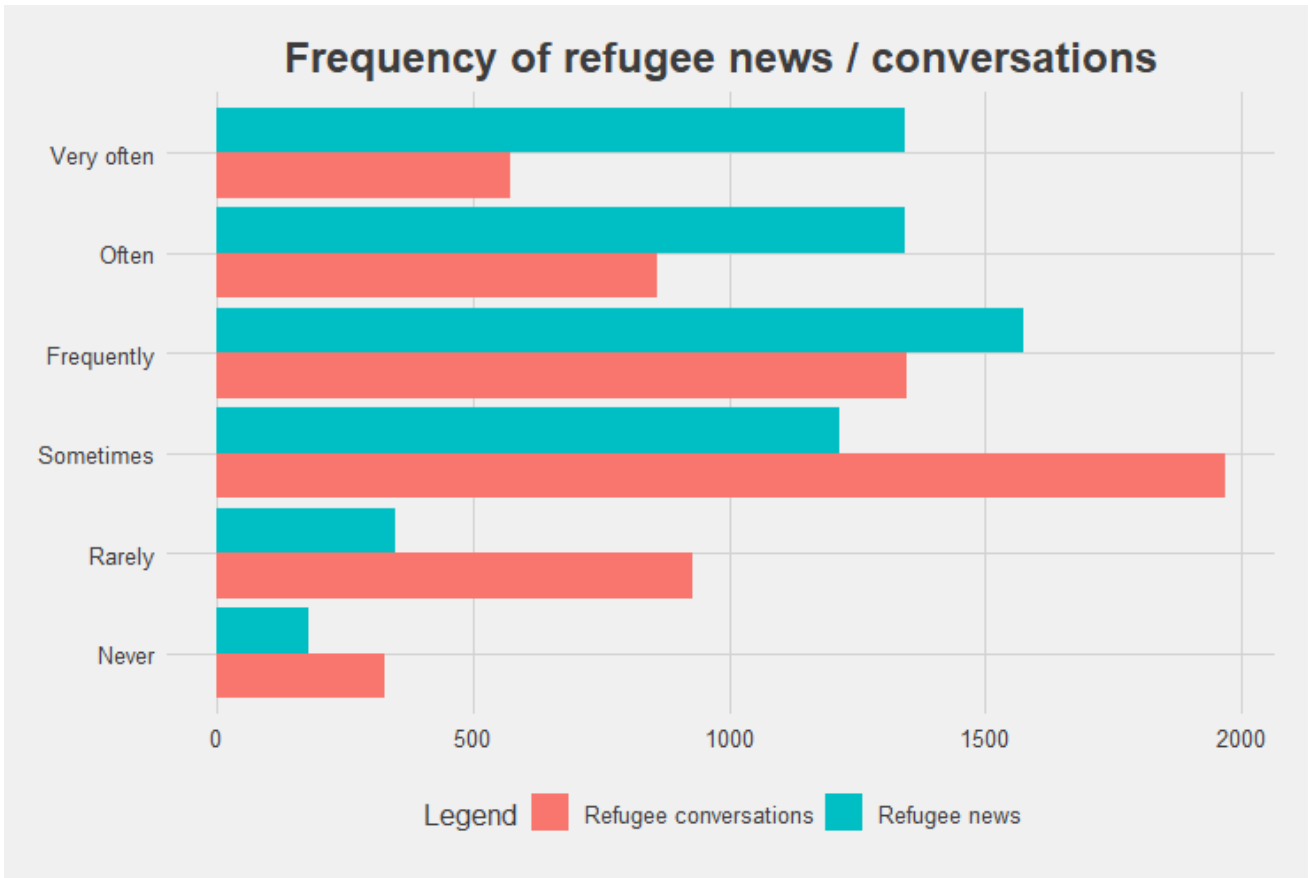Figure 7: Frequency - News consumption

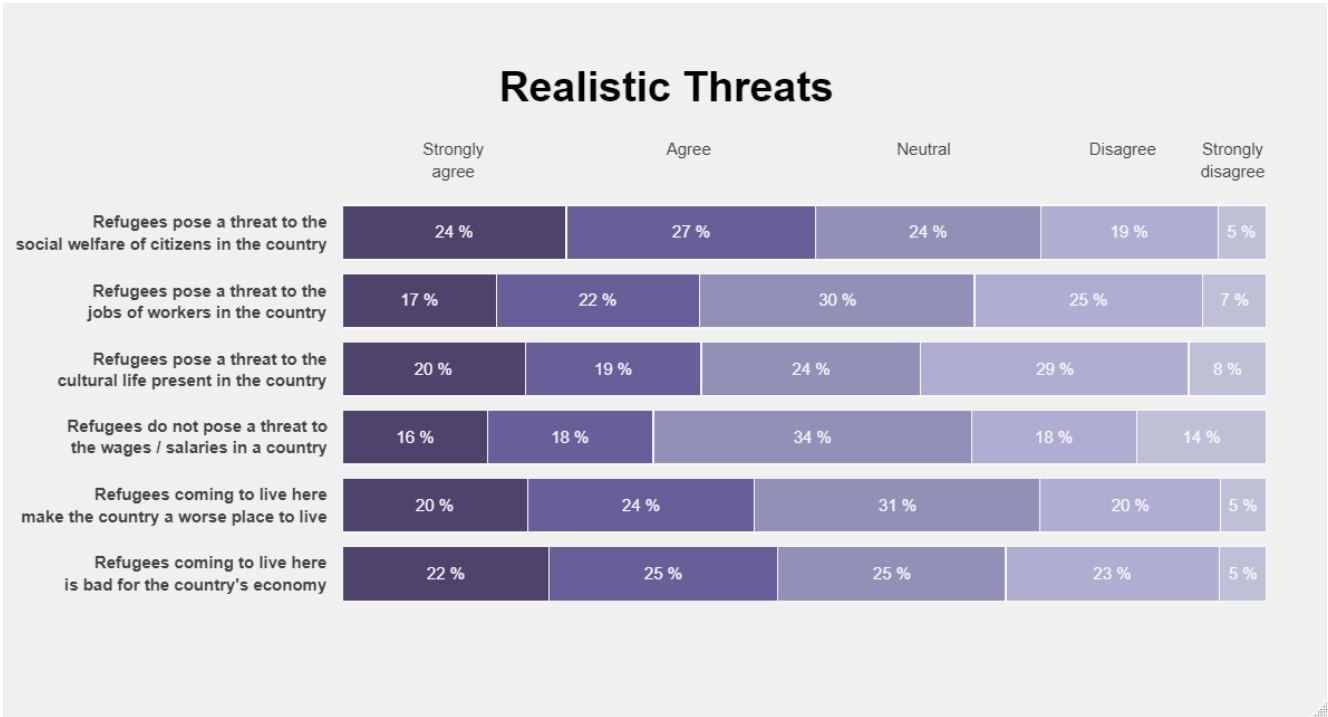Figure 8: Frequency - Refugee news and conversations

Figure 9: Realistic threats - refugees (Exact questions can be found in the codebook, see References)

3.5 Threat Perspectives And Inter-group Contact

In the measurement on threat perspectives, several types of threat were included. A first set of indicators assessed economic (or realistic) threat. This was measured by asking four questions: 'Would you say that refugees who come to live here generally take jobs away from workers in [country], or generally help to create new jobs?', 'Most refugees who come to live here work and pay taxes. They also use health and welfare services. On balance, do you think refugees who come here take out more than they put in or put in more than they take out?', and 'Would you say it is generally bad or good for [country]'s economy that refugees come to live here from other countries?'. One additional item on realistic threat was 'Average wages and salaries are generally brought down by refugees coming to live and work here.' Another threat perspective is on cultural threat and was measured through the following question: 'Would you say that [country]'s cultural life is generally undermined or enriched by refugees coming to live here from other countries?'.

Both types of threat were measured on a 11-point Likert scale, with the high end of the scale signifying a low degree of threat. The only exception is the item on wages and salaries being brought down by refugees coming to live in the country, as this is coded on a five-point Likert scale with the high end of the scale indicating strong agreement with this statement.

A third type of threat is on fear of terrorism, which was particularly relevant to include given the events in Europe around the time of the survey. We adapted the fear of terrorism-scale developed by Nellis and Savage. The scale consists of several hypothetical scenarios ('Someday I may witness a terror attack', 'Someday I may be the victim of terror attack', ...), with answer categories ranging from 0 (Not likely at all) to 10 (Very likely). Whereas the original scale measured this fear for several scenarios separately ('I could be on a plane that is hijacked', 'I could be on a subway or bus that is hijacked'), we have aggregated these into a single item ('Someday I may be on a plane/subway/bus that is bombed'). Some of the original wording was also adjusted.

Direct inter-group contact was measured by asking whether respondents have any inter-ethnic friendships (no, some, or many), and how often they have inter-ethnic random contact on the street, at work, in shops... [seven categories, ranging from 0 (Never) to 6 (Every day)]. The valence of direct contact was measured by an 11-point scale, with 0 indicating a negative evaluation of inter-group contact, and 10 indicating a positive evaluation.
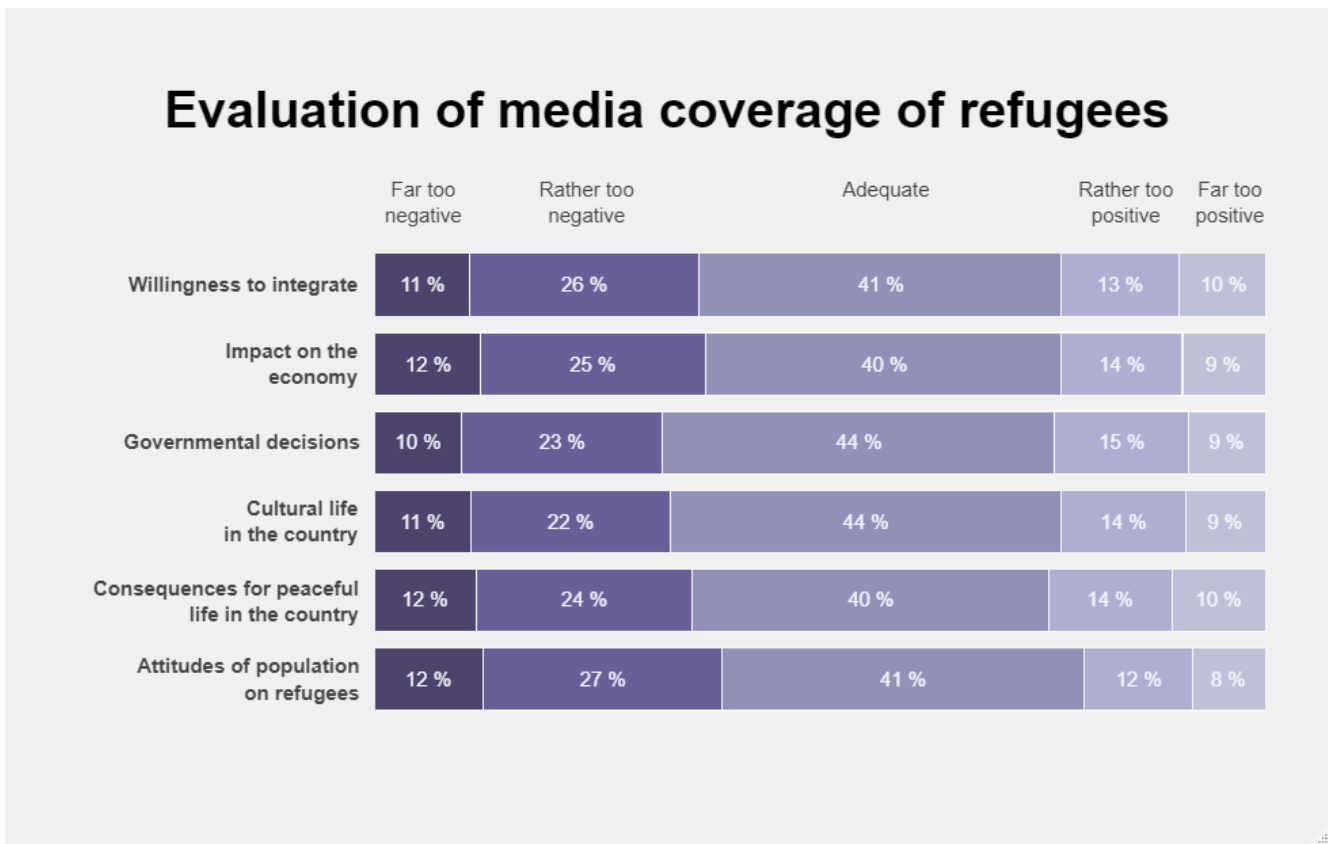
Figure 10: Media coverage on refugees

Fear of terrorism

	Highly likely	Likely	Neutral	Unlikely	Highly unlikely
Someday, I may witness a terror attack	20 %	29 %	18 %	20 %	13 %
Someday, I may be the victim of a terror attack	19 %	27 %	19 %	20 %	15 %
Someday, I may be on a plane/subway/bus that is bombed	19 %	25 %	18 %	20 %	18 %
Someday, family/close friends may witness a terror attack	22 %	30 %	19 %	19 %	11 %
Someday, family/close friends may be the victim of a terror attack	21 %	29 %	18 %	19 %	12 %
Someday, family/close friends may be on a plane/subway/bus that is bombed	21 %	28 %	19 %	20 %	13 %

Figure 11: Culture of fear regarding terrorism

4 Methods And Formulae Used

4.1 Wilcoxon Signed-Rank Sum Test

The Wilcoxon signed-rank sum test allows allows for the comparison of two ordinal measures and is the non-parametric equivalent to the paired samples t-test.When applied to test the location of a set of samples, it serves the same purpose as the one-sample Student's t-test. On a set of matched samples, it is a paired difference test like the paired Student's t-test.Unlike the Student's t-test, the Wilcoxon signed-rank test does not assume that the data is normally distributed. We consider the paired version of the Wilcoxon test.

4.1.1 Data

- The data used typically consists of $2n$ paired observations, with n observations for each subject: X_i & Y_i.

Subject i	X_i	Y_i
1	X_1	Y_1
2	X_2	Y_2
.	.	.
.	.	.
.	.	.
n	X_n	Y_n

Figure 12: Data for the Wilcoxon Signed Rank-Sum test

4.1.2 Assumptions

- Taking $Z_i = X_i - Y_i$, for $i = 1, ..., n$; we assume the differences $Z_1, ..., Z_n$ are mutually independent.

- We assume each Z_i, $i = 1, ..., n$, comes from a continuous population (not necessarily the same one) that is symmetric about a common median θ. If F_i represents the distribution function for Z_i, this assumption requires that
$$F_i(\theta + t) + F_i(\theta - t) = 1$$
, for every t and i = 1, . . . , n.

4.1.3 Null And Alternative Hypotheses

- The null hypothesis of interest here is that of zero shift in location due to the treatment, namely, H_0: $\theta = 0$ This null hypothesis asserts that each of the distributions (not necessarily the same) for the differences is symmetrically distributed about 0.

    - One-sided alternative hypothesis: H_1: $\theta < 0$
    - One-sided alternative hypothesis: H_2: $\theta > 0$
    - Two-sided alternative hypothesis: H_3: $\theta \neq 0$

4.1.4 Procedure

- To compute the Wilcoxon signed rank statistic T^+ , we consider the absolute values $|Z_1|, \ . \ . \ . \ , |Z_n|$ of the differences and order them from least to greatest.

- We let R_i denote the rank of $|Z_i|$, $i = 1, \ . \ . \ . \ , n$, in this ordering.

- We define indicator variables ψ_i , i = 1, . . . , n, where
$$\psi_i = \begin{cases} 1, & \text{if } |Z_i| > 0 \\ 0, & \text{if } |Z_i| < 0 \end{cases}$$

- We consider the n products $R_i\psi_i$. The product $R_i\psi_i$ is known as the positive signed rank of Z_i.

- The Wilcoxon signed rank statistic T^+ is then the sum of the positive ranks, namely,
$$T^+ = \sum_{i=1}^{n} R_i\psi_i$$

- **Tests:**

    - One sided lower tailed test: To test H_0: $\theta = 0$ against H_1: $\theta < 0$ at the α level of significance:
$$\text{Reject } H_0 \text{ if } T^+ \leq \frac{n(n+1)}{2} - t_\alpha \text{ ; otherwise do not reject,}$$

11

– One sided upper tailed test: To test H_0: $\theta = 0$ against H_1: $\theta > 0$ at the α level of significance:

$$\text{Reject } H_0 \text{ if } T^+ \geq t_\alpha \text{ ; otherwise do not reject,}$$

– Two sided test: To test H_0: $\theta = 0$ against H_1: $\theta \neq 0$ at the α level of significance:

$$\text{Reject } H_0 \text{ if } T^+ \geq t_{\alpha/2} \text{ or } T^+ \leq \frac{n(n+1)}{2} - t_{\alpha/2} \text{ ; otherwise do not reject.}$$

The two-sided procedure is the two-sided symmetric test with $\frac{\alpha}{2}$ probability in each tail of the null distribution of T^+.

4.1.5 Large-Sample Approximation

- The large-sample approximation is based on the asymptotic normality of T^+, suitably standardized. We first need to know the expected value and variance of T^+ when the null hypothesis is true. When H_0 is true, the expected value and variance of T^+ are:

$$E_0(T^+) = \frac{n(n+1)}{4}$$

$$\text{Var}_0(T^+) = \frac{n(n+1)(2n+1)}{24}$$

- The standardized test statistic T^* of T^+ is:

$$T^* = \frac{T^+ - E_0(T^+)}{\sqrt{\text{var}_0(T^+)}} = \frac{T^+ - \left\{ \frac{n(n+1)}{4} \right\}}{\sqrt{n(n+1)(2n+1)/24}}$$

- When H_0 is true, T^* has, as n tends to infinity, an asymptotic $N(0,1)$ distribution.

- The normal theory approximation for the procedure explained previously for the

  – One sided upper tailed test is: Reject H_0 if $T^* \geq z_\alpha$; otherwise do not reject;
  – One sided lower tailed test is: Reject H_0 if $T^* \leq z_\alpha$; otherwise do not reject;
  – Two sided test is: Reject H_0 if $|T^*| \geq z_{\frac{\alpha}{2}}$; otherwise do not reject.

4.1.6 Zeroes And Ties

- If there are zero values among the Z_i's, we discard the zeros and redefine n to be the number of non-zero Z_i's.

- If there are ties among the (non-zero) Z_i's, we assign each of the observations in a tied group, the average of the integer ranks that are associated with the tied group. After computing T^+ with these average ranks for nonzero Z_i's, we use the procedure explained previously. However, this test associated with the tied Z_i's is only approximately, and not exactly, of significance level α.

- When applying the large-sample approximation, an additional factor must be taken into account. Although the ties in the non-zero Z_i's do not affect the null expected value of T^+ , it's null variance is reduced to:

$$\text{var}_0(T^+) = \frac{1}{24} \left[n(n+1)(2n+1) - \frac{1}{2} \sum_{j=1}^{g} t_j(t_j - 1)(t_j + 1) \right],$$

where g denotes the number of tied groups of non-zero Z_i's and t_j is the size of the tied group j.

- We note that an untied observation is considered to be a tied "group" of size 1. In particular, if there are no ties among the Z_i's, then $g = n$ and $t_j = 1$ for all j. So each term $t_j(t_j - 1)(t_j + 1)$ is zero and the variance is unchanged.

- The variance in the expression for T^* needs to be replaced with the modified variance shown previously, when using the large-sample approximation.

- With this modified value of T^* , the normal theory approximation can be applied.

4.1.7 Comments

There is no requirement that the individual X_i and Y_i are independent, only that the pairs (X_i, Y_i) are, and therefore the resulting differences Z_i be mutually independent. Indeed, in most applications, the individual X_i and Y_i are dependent.

It may appear that some of the information in the ranking of the sample Z - differences is being lost by using only the positive signed ranks to compute the test statistic T^+. Such is not the case. If we define T^- to be the sum of negative ranks, then

$$T^- = \sum_{i=1}^{n}(1 - \psi_i)R_i.$$

It follows that $T^+ + T^- = \dfrac{n(n+1)}{2}$. So the test procedure could equivalently be based on T^-, or even on $min(T^+, T^-)$.

4.2 Ordinal Regression

Ordinal regression (also called "ordinal classification") is a type of regression analysis used for predicting an ordinal variable, i.e. a variable whose value exists on an arbitrary scale where only the relative ordering between different values is significant. Ordinal logistic regression is one example of an Ordinal regression model, another example would be the Ordered Probit model. Ordinal regression turns up often in the social sciences, for example in the modeling of human levels of preference (on a scale from, say, 1–5 for "very poor" through "excellent").

4.2.1 Definitions

Let Y be an ordinal variable with J categories (From 1 upto J). Then $P(Y \leq j)$ is the cumulative probability of Y being less than or equal to a specific category $j = 1, \ldots, J$. Note that $P(Y \leq J) = 1$. The odds of being less than or equal to a particular category j $(j = 1, \ldots, J-1)$ can be defined as

$$\frac{P(Y \leq j)}{P(Y > j)}$$

The log odds, also known as the *logit* is then defined as

$$\mathrm{logit}[P(Y \leq j)] = \log \frac{P(Y \leq j)}{P(Y > j)}$$

4.2.2 Model

The ordinal logistic regression model can be defined as

$$\mathrm{logit}[P(Y \leq j)] = \beta_{j0} + \beta_{j1}x_1 + \ldots + \beta_{jp}x_p$$

Due to the *parallel lines assumption*, the intercepts β_{j0} are different for each category, but the slopes β_{ji}, $i = 1, \ldots, p$ are constant across categories. So the model simplifies to

$$\mathrm{logit}[P(Y \leq j)] = \beta_{j0} + \beta_1 x_1 + \ldots + \beta_p x_p$$

Therefore, the model assumes that the effect of x_i is identical for all $J - 1$ cumulative logits, so it requires a single parameter rather than $J - 1$ parameters to describe the effect of x_i on Y.

4.2.3 Assumptions

- The response variable is ordinal.

- The explanatory variables are continuous or categorical. You can include ordinal variables, but you need to treat them either as continuous or categorical.

- There is no *multicollinearity*. Multicollinearity occurs when there are high correlations among predictor variables, leading to unreliable and unstable estimates of regression coefficients.

- The *parallel lines assumption*, meaning each predictor variable has an identical effect at each of the cumulative logits.

4.2.4 Checking assumptions

- When fitting a proportional odds model, it's a good idea to check the assumption of proportional odds.

- One way to do this is by comparing the proportional odds model with a multinomial logit model, also called an unconstrained baseline logit model.

- The multinomial logit model is typically used to model unordered responses and fits a slope to each level of the J–1 responses.

- This suggests the proportional odds model is nested in the multinomial model, and that we could perform a likelihood ratio test to see if the models are statistically different.

4.2.5 Interpretation

For simplicity, from now on, let us suppose that we have one predictor variable and one response variable (with J categories). So our model looks like

$$\text{logit}[P(Y \leq j)] = \beta_{j0} + \beta x$$

The logistic regression model indicates that the logit increases by β for every unit increase in x. Most of us do not think naturally on a logit scale, so we need to consider alternative interpretations. The above equation implies the following formula for the probability $P(Y \leq j)$, using the exponential function.

$$P(Y \leq j) = \frac{\exp(\beta_{j0} + \beta x)}{1 + \exp(\beta_{j0} + \beta x)}$$

The figure below depicts this model for a four category response and quantitative X (i.e. a single predictor variable).
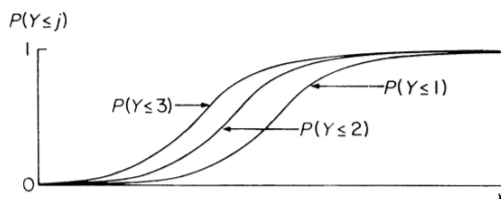

Figure 13: Cumulative probabilities against x, $\beta > 0$

The curve for $P(Y \leq j)$ looks like a logistic regression curve for a binary response with pair of outcomes $P(Y \leq j)$ and $P(Y > j)$. The *parallel lines assumption* or the common effect β for each j implies that the three curves have the same shape, as $|\beta|$ determines how quickly the curve climbs or drops. At any fixed x value, the curves have the same ordering as the cumulative probabilities, the one for $P(Y \leq 1)$ being lowest.
Now, $P(Y = j) = P(Y \leq j) - P(Y \leq j - 1)$. When $\beta > 0$, as x increases, the response on Y is more likely to fall at the low end of the ordinal scale.
As a sidenote, if the predictor variable is categorical, then the *baseline level* is taken to be $x = 0$, i.e. the level $x = 0$ is used as a reference while interpreting the slopes in the model output.

Model interpretations can use odds ratios for the cumulative probabilities and their complements. For two values x_1 and x_2 of x, an odds ratio comparing the cumulative probabilities is

$$\frac{P(Y \leq j \mid x = x_2)/P(Y > j \mid x = x_2)}{P(Y \leq j \mid x = x_1)/P(Y > j \mid x = x_1)}$$

The log of this odds ratio is the difference between the cumulative logits at those two values of x. This equals $\beta(x_2 - x_1)$, proportional to the distance between the x values. In particular, for $x_2 - x_1 = 1$, the odds of response below any given category multiply by $\exp(\beta)$ for each unit increase in x. For this log odds ratio $\beta(x_2 - x_1)$, the same proportionality constant β applies for each cumulative probability. Thus, this property is also called the proportional odds assumption of model.

4.2.6 Ordinal logistic regression in R

- In R, ordinal logistic regression is conducted using the `polr()` function present in the MASS library.

- The syntax for conducting an ordinal regression is:
  ```
  m <- polr(y ~ x1 + x2 + ...  + xp, data = df)
  summarize(m)
  ```

- Here, `df` is the dataframe, `y` is the response column, which has to be an ordinal categorical variable, a.k.a a *factor*, `x1, ... ,xp` are the predictor variables and `m` is the model.

- Note that in the `polr()` function, the ordinal logit model is parametrized as

$$\text{logit}[P(Y \leq j)] = \beta_{j0} - \eta_1 x_1 - \ldots - \eta_p x_p$$

5   Analytic Strategy

Hypothesis 1 is requires us to investigate differences in public opinion between both categories of newcomers(immigrants and refugees). Hypotheses 2, 3 and 4 require us to investigate differences in public opinion between opposite sub-categories of both immigrants and refugees.

Since the relevant data was measured on a Likert scale, its ordinal nature leads to the violation of a common statistical assumption, that being: The data is not necessarily normally distributed. Taking this into consideration, we make use of the Wilcoxon Signed Rank-Sum Test in our analyses. Additionally, to further elaborate on the results obtained using the Wilcoxon Signed Rank-Sum Test, we conduct ordinal regressions with the response variables as: Mean scores on attitudes towards immigrants and refugees and the difference between these attitudes(taking mean score across all categories). We use multiple demographic variables as our predictor variables for this regression(using one predictor variable at a time). A positive score for variable indicates that people prefer refugees, and a negative one indicates that people prefer immigrants.

For the additional hypotheses, we run further ordinal regressions with the same response variables and several trust in media and psychological indicator variables as the predictor variables.
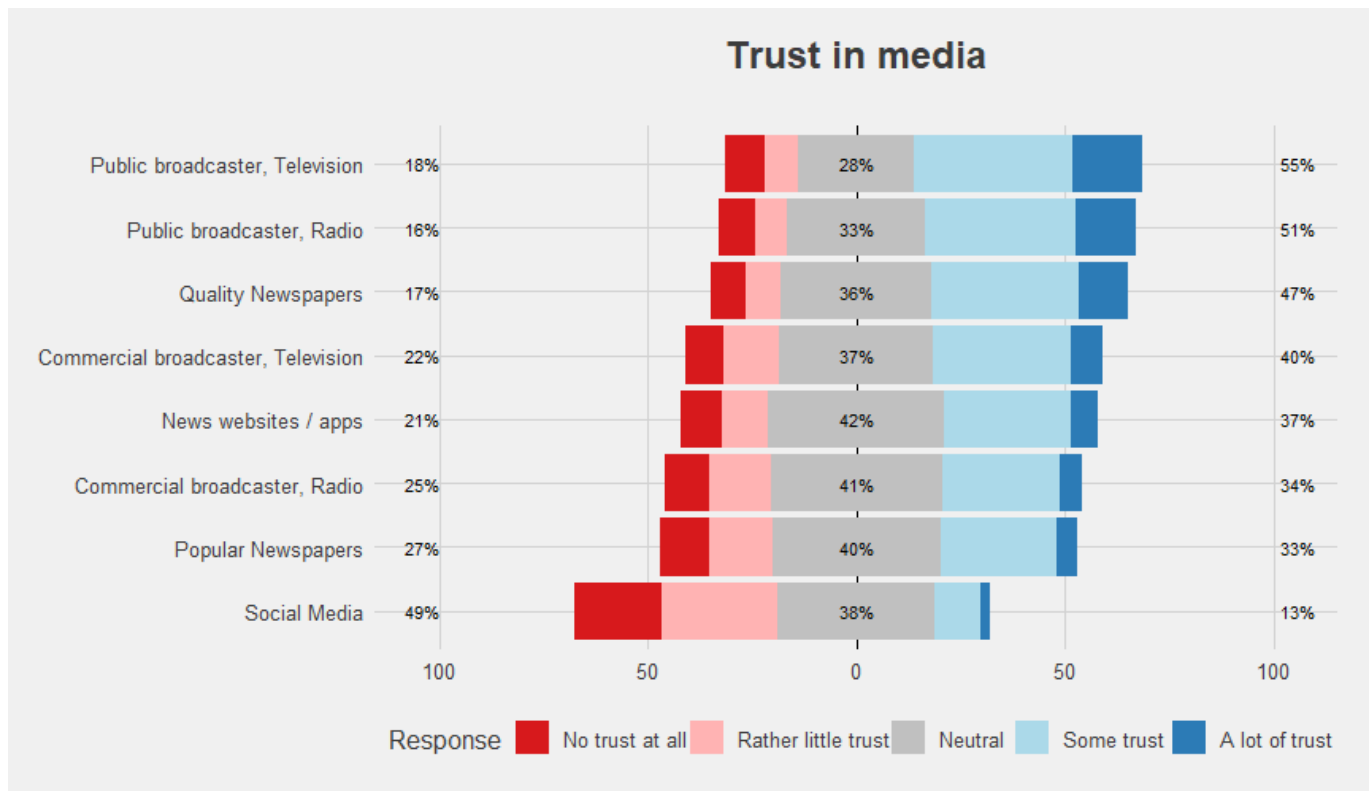

Figure 14: Trust in different media sources

6 Results

	Frequency	Percentage	Mean (SD)
Age	-	-	44.96 (13.11)
Gender (N = 6000)			
Male	2844	47.4	0.47 (0.01)
Female	3156	52.6	0.53 (0.01)
Migration background (N = 5406)			
No migration background	4521	75.35	0.75 (0.01)
European migration background	732	12.2	0.12 (0)
Non-European migration background	153	2.55	0.02 (0)
Educational attainment (N = 5619)			
No / Primary education	310	5.17	0.05 (0)
Secondary education	3059	50.98	0.51 (0.01)
Tertiary education	2250	37.5	0.38 (0.01)
Religious denomination (N = 5991)			
Christian	2674	44.57	0.44 (0.01)
Muslim	163	2.72	0.03 (0)
Other	443	7.38	0.07 (0)
Not religious	2711	45.18	0.45 (0.01)
News media consumption (N = 6000)			
Public broadcaster	-	-	3.6 (2.21)
Commercial broadcaster	-	-	3.6 (2.11)
*Quality newspapers	-	-	1.52 (1.22)
*Tabloids	-	-	2.03 (1.64)
Exposure to refugee news (N = 6000)	-	-	4.26 (1.32)

Figure 15: Table 1

The entries in Table 2 describe the results obtained from applying the Wilcoxon Signed Rank-Sum test while investigating differences in attitudes between immigrants and refugees across a particular indicator (eg. Same race vs. Different race). The first four columns indicate the mean and the standard deviation of the responses given by the respondents, indicating their preferences towards immigrants and refugees. The fifth column gives the 95% confidence interval for the difference of means (Mean response for immigrants - Mean response for refugees). The last column gives the p-value obtained from the two-sided Wilcoxon Signed Rank-Sum test. If this p-value is below 0.05, it indicates that there is a significant difference in attitudes towards immigrants and refugees in the corresponding indicator (pair). If the p-value is below 0.05, then the higher of the two mean values would indicate whether immigrants or refugees are preferred.

Looking at Table 2, using our interpretation as described above, we conclude that significant differences in attitudes can be found in all but one pair (Pair 6) of indicators. Refugees are preferred over immigrants when the newcomers belong to a different race as most of the country's population (Pair 1), are from poor European / Non-European countries (Pairs 4, 6), or they are from Muslim countries (Pair 7). The opposite preference, that is, immigrants are preferred over refugees when the newcomers belong to the same race as most of the country's population (Pair 1), or they are from rich European countries (Pair 3). However, no significant difference in attitudes can be found when the newcomers are from rich Non-European countries (Pair 6) which is indicated by the corresponding p-value (0.48).

Our interpretation for the results displayed in tables 4 and 5 is exactly the same as in table 2, except for the fact that in tables 4 and 5, inter-group attitude differences are measured across subcategories of immigrants and refugees respectively (as opposed to differences in attitudes between immigrants and refugees).

	M	SD	M	SD	95% CI	p-value
Pair 1	Immigrants of the same race		Refugees of the same race			
	2.75	0.87	2.73	0.91	(-0.01 , 0.05)	< 0.05
Pair 2	Immigrants of different race		Refugees of different race			
	2.46	0.89	2.51	0.93	(-0.08 , -0.01)	< 0.001
Pair 3	Immigrants of rich European countries		Refugees of rich European countries			
	2.63	0.88	2.56	0.93	(0.03 , 0.11)	< 0.001
Pair 4	Immigrants of poor European countries		Refugees of poor European countries			
	2.51	0.94	2.54	0.35	(-0.05 , -0.02)	< 0.001
Pair 5	Immigrants of rich Non-European countries		Refugees of rich Non-European countries			
	2.48	0.9	2.47	0.94	(-0.02 , 0.04)	0.48
Pair 6	Immigrants of poor Non-European countries		Refugees of poor Non-European countries			
	2.39	0.91	2.47	0.95	(-0.11 , -0.04)	< 0.001
Pair 7	Immigrants of Muslim countries		Refugees of Muslim countries			
	2.29	0.96	2.33	1.12	(-0.07 , -0.06)	< 0.001

Figure 16: Table 2

Indicator	Attitudes towards refugees	Attitudes towards immigrants	Difference
Age	-0.015	-0.015	-0.001
Gender			
Female (Ref.)			
Male	-0.09	-0.074	-0.02
Educational Attainment			
No / Primary education (Ref.)			
Secondary Education	0.465	0.457	0.074
Tertiary Education	1.063	1.032	0.023
Migration Background			
No migration background (Ref.)			
European migration background	0.304	0.346	-0.051
Non-European migration background	0.486	0.5	0.009
Religious denomination			
Christian (Ref.)			
Muslim	0.985	0.886	0.189
Other	0.427	0.356	0.139
Not religious	0.195	0.115	0.151
**News media consumption			
Public broadcaster	0.08	0.07	0.02
Commercial broadcaster	-0.1	-0.08	-0.05
Quality newspapers	0.13	0.14	0.02
Tabloids	-0.04	-0.02	-0.04
Exposure to Refugee news	0.132	0.105	0.055
Country of residence			
Belgium (Ref.)			
France	-0.258	-0.295	0.045
Netherlands	0.043	-0.027	0.14
Sweden	0.43	0.458	-0.012

Figure 17: Table 3

17

	M	SD	M	SD	95% CI	p-value
Pair 1	Immigrants of the same race		Immigrants of different race			
	2.75	0.87	2.47	0.89	(0.26 , 0.32)	< 0.001
Pair 2	Immigrants of rich European countries		Immigrants of poor European countries			
	2.64	0.89	2.51	0.89	(0.09 , 0.16)	< 0.001
Pair 3	Immigrants of rich Non-European countries		Immigrants of poor Non-European countries			
	2.48	0.9	2.4	0.91	(0.05 , 0.12)	< 0.001
Pair 4	Immigrants of rich European countries		Immigrants of rich Non-European countries			
	2.64	0.89	2.48	0.9	(0.12 , 0.19)	< 0.001
Pair 5	Immigrants of poor European countries		Immigrants of poor Non-European countries			
	2.51	0.89	2.39	0.91	(0.08, 0.15)	< 0.001

Figure 18: Table 4

Table 3 enables us to get a more in-depth look at attitude differences across several demographic / social / news-consumption characteristics. The first column here indicates the mean of the slopes obtained on performing ordinal regressions with attitudes towards refugees as the response variable and indicator in each row of the table as the predictor variable. The second column measures the same metric as the first, but for immigrants. For these columns, a positive number in column 1 and 2 indicates that the corresponding characteristic / predictor variable is positively associated with attitudes towards refugees and immigrants respectively. For the last column, the difference between the responses for each of the seven indicators (same race, different race, etc.) for refugees and immigrants was taken to be the response variable, and the mean of the slopes obtained is given in this column. The predictor variable in this case was again, the characteristic mentioned in the corresponding row in the table. So if this column has a positive number, then it indicates that refugees are preferred over immigrants, on average, otherwise immigrants are preferred. A (Ref.) in the indicator column indicates that this characteristic was taken to be the baseline level while performing the ordinal regression.

Looking at Table 3, using the interpretation described above, we see that educational attainment, having a migration background and belonging to a non-Christian community is positively associated with attitudes towards immigrants and refugees. We also observe that those who tend to prefer public broadcasters (on television / radio) have tend to have positive attitudes towards immigrants and refugees, whereas the opposite is true for those who tend to prefer commercial broadcasters, in which case they tend to have negative attitudes towards both categories of newcomers. Coming the geographic characteristics, when compared to the Belgians, we see that the French have more negative attitudes towards both immigrants and refugees, the Dutch have positive attitudes towards refugees but not immigrants, and the Swedish citizens have more positive attitudes towards both categories of newcomers.

Coming to differences in attitudes, we see that educational attainment, having a Non-European migration background and belonging to a non-Christian community is positively associated with a higher preference towards refugees (over immigrants), but the opposite conclusion is made when people have a European migration background and tend to prefer commercial news broadcasters.

The double asterisk in the News media consumption indicators indicates that these numbers were obtained from the results in the original paper (by David De Coninck), as we were unable to make out which variables in the data were used for conducting these ordinal regressions.

Looking at Table 4, we conclude that significant differences in attitudes can be found in all pairs indicators, as all of the p-values are less than 0.001. We observe that immigrants of the same race as most of the country's population, from rich European countries, and from rich Non-European countries are preferred over immigrants of a different race as most of the country's population, from poor European countries, and from poor Non-European countries respectively (Pairs 1, 2 and 3). Also, immigrants from rich European countries and poor European countries are preferred over immigrants from poor European countries and poor Non-European countries (Pairs 4, 5).

	M	SD	M	SD	95% CI	p-value
Pair 1	Refugees of the same race		Refugees of different race			
	2.73	0.91	2.51	0.93	(0.18 , 0.24)	< 0.001
Pair 2	Refugees of rich European countries		Refugees of poor European countries			
	2.56	0.93	2.54	0.93	(-0.01 , 0.05)	< 0.01
Pair 3	Refugees of rich Non-European countries		Refugees of poor Non-European countries			
	2.47	0.94	2.47	0.96	(-0.03 , 0.04)	0.94
Pair 4	Refugees of rich European countries		Refugees of rich Non-European countries			
	2.56	0.93	2.48	0.91	(0.06 , 0.13)	< 0.001
Pair 5	Refugees of poor European countries		Refugees of poor Non-European countries			
	2.54	0.93	2.47	0.95	(0.04 , 0.11)	< 0.001

Figure 19: Table 5

	Attitude towards refugees	Attitude towards immigrants	Difference in attitudes
Trust in public broadcaster on television	0.41	0.387	0.069
Trust in commercial broadcaster on television	0.18	0.19	-0.001
Trust in public broadcaster on radio	0.434	0.408	0.07
Trust in commercial broadcaster on radio	0.111	0.106	0.012
Trust in quality newspaper	0.442	0.414	0.087
Trust in popular newspaper	0.086	0.115	-0.039
Trust in newswebsites/apps	0.266	0.255	0.039
Trust in social media	-0.048	-0.015	-0.059

Figure 20: Table 6

	Attitude towards Refugee	Attitude towards Immigrants	Difference between attitudes
Outgoing ,Enthusiastic	0.173	0.192	-0.015
Critical , Confrontational	-0.031	-0.004	-0.057
Reliable , Disciplined	0.234	0.176	0.103
Tense , Easily in a bad mood	-0.025	-0.021	-0.021
Open to new experiences , profound	0.545	0.459	0.174
Reserved , Quiet	0.07	0.053	0.029
Emphatic , Warm	0.493	0.407	0.177
Disorganized , Sloppy	0.113	0.119	-0.001
Calm , Emotionally stable	0.151	0.143	0.025
Conservative ,not creative	-0.143	-0.0117	-0.063

Figure 21: Table 7

Looking at Table 5, we conclude that significant differences in attitudes can be found in all but one pair (Pair 3) of indicators. We observe that refugees of the same race as most of the country's population and from rich European countries are preferred over refugees of a different race as most of the country's population and from poor European countries respectively (Pairs 1, 2). Also, refugees from rich European countries and poor European countries are preferred over refugees from poor European countries and poor Non-European countries (Pairs 4, 5). However, no significant difference in attitudes can be found between refugees from rich Non-European countries and refugees from poor Non-European countries (Pair 3), because the corresponding p-value is 0.94.

The results in Table 6 indicate that people who watch public or commercial television or listen to public or commercial radio generally have a favourable attitude towards both migrants and refugees, while preferring refugees over migrants, except in the case of commercial television, where migrants are preferred over refugees. Trust in quality/popular newspapers and news websites is correlated to a more favourable attitude towards immigrants and refugees, with trust in popular newspaper being the only case where immigrants are preferred over refugees. Trust in social media seems to be correlated to a general unfavourable attitude towards both immigrants and refugees, with a preference for immigrants over refugees.

The results in Table 7 indicate a strike contrast in attitudes towards immigrants and refugees based on categories of psychological indicators. The results indicate that the categories that hold favourable attitudes towards immigrants, also hold favourable attitudes towards refugees and vice-versa.

The results also indicate that those who identify themselves more as critical, conservative or tense have a more unfavourable attitude towards refugees and immigrants alike, while the other categories seem to have more favourable attitudes towards them. On the other hand, among the categories, there isn't sufficient evidence to conclude a specific trend on the difference between immigrant and refugee attitudes, since a significant number of categories view immigrants as more favourable and a significant number of categories also view refugees as more favourable.

7   Limitations and Conclusion

- First, we cannot claim that the results are representative of the European population, because the study involved only four countries in Europe, and in particular, they were Western European countries. The sample would have been representative of the European population if other European countries were included in the survey, such as Germany, the United Kingdom, etc.

- Second, although the framing of migrants in the media is often subtle, our formulations of categories (like immigrants from rich countries outside Europe) is much more explicit. This may have skewed attitudes to favour newcomers who are more similar to the group mentioned.

In conclusion, the author believes that this study provides valuable insights into several new avenues of research that may shed further light on the complex and intricate relationship between migrant categorizations and public opinion preferences or attitudes of the general population. This concludes our report.

8   References

1. Reference paper: https://doi.org/10.1080/1369183X.2019.1694406

2. Dataset (Also contains the codebook with data descriptions): https://data.mendeley.com/datasets/3m439mbrmm/2

3. Data paper: https://www.sciencedirect.com/science/article/pii/S235234091930890X?via%3Dihub

4. Fitting and interpreting a proportional odds model:
   https://data.library.virginia.edu/fitting-and-interpreting-a-proportional-odds-model/

5. Choosing the correct test: https://stats.oarc.ucla.edu/other/mult-pkg/whatstat/

6. Ordinal logistic regression: https://stats.oarc.ucla.edu/r/dae/ordinal-logistic-regression/

7. Ordinal logistic regression:
   https://bookdown.org/mpfoley1973/data-sci/ordinal-logistic-regression.html

8. Interpreting the coefficients of an ordinal logistic regression: https://stats.oarc.ucla.edu/r/faq/ologit-coefficients/

9. Implementing and interpreting ordinal logistic regression:
   https://towardsdatascience.com/implementing-and-interpreting-ordinal-logistic-regression-

10. Likert scale: https://www.simplypsychology.org/likert-scale.html

8.1   Books

1. Peter Dalgarrd - Introductory Statistics with R - 2nd edition - Springer Publications

2. Myles Hollander, Douglas A. Wolfe, Eric Chicken - Nonparametric Statistical Methods - 2nd edition - Wiley Publications

3. Alan Agresti - Introduction to Categorical Data Analysis - 2nd edition - Wiley publications

4. McClave, James T_Sincich, Terry T - Statistics - Chapter 14 - Nonparametric Statistics - Pearson publications

5. John A. Rice - Mathematical Statistics and Data Analysis - 3rd Edition - Cengage publications

8.2   R And R packages

1. Hadley Wickham and Garret Grolemund - R for Data science - https://r4ds.had.co.nz/

2. tidyverse package - https://www.tidyverse.org/packages/

3. plotly package - https://plotly.com/r/

4. janitor package - https://garthtarr.github.io/meatR/janitor.html

5. misty package - https://www.rdocumentation.org/packages/misty/versions/0.4.3

6. MASS package - https://www.rdocumentation.org/packages/MASS/versions/7.3-47

7. plyr package - https://www.rdocumentation.org/packages/plyr/versions/1.8.6

8. ggthemes package - https://mran.microsoft.com/snapshot/2017-02-04/web/packages/ggthemes/vignettes/ggthemes.html

9. RColorBrewer package - https://www.r-graph-gallery.com/38-rcolorbrewers-palettes.html