Guilty by association: Using word embeddings to measure ethnic stereotypes in news coverage

Kroon, A.C.; Trilling, D.; Raats, T.

DOI
10.1177/1077699020932304

Publication date
2021

Document Version
Final published version

Published in
Journalism and Mass Communication Quarterly

License
CC BY

Link to publication

Citation for published version (APA):
Kroon, A. C., Trilling, D., & Raats, T. (2021). Guilty by association: Using word embeddings to measure ethnic stereotypes in news coverage. *Journalism and Mass Communication Quarterly*, *98*(2), 451-477 . https://doi.org/10.1177/1077699020932304

Journalism & Mass Communication Quarterly
2021, Vol. 98(2) 451-477
© 2020 AEJMC
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1077699020932304
http://journals.sagepub.com/home/jmq
SAGE

Guilty by Association: Using Word Embeddings to Measure Ethnic Stereotypes in News Coverage

Anne C. Kroon[1], Damian Trilling[1], and Tamara Raats[1]

Abstract

The current study provides a new level of empirical evidence for the nature of ethnic stereotypes in news content by drawing on a sample of more than 3 million Dutch news items. The study's findings demonstrate that universally accepted dimensions of stereotype content (i.e., low-status and high-threat attributes) can be replicated in news media content across a diverse set of ingroup and outgroup categories. Representations of minorities in newspapers have become progressively remote from factual integration outcomes, and are therefore rather an artifact of news production processes than a true reflection of what is actually happening in society.

Keywords

bias, stereotypes, automated content analysis, time-series analysis

News media have been accused of spreading stereotypes by repeatedly associating racial/ethnic outgroups with one-sided, biased attributes such as criminality and unemployment. Exposure to such stereotypical associations can contribute to the development of stereotypical beliefs by strengthening mental linkages between social groups and biased attributes (Arendt & Northup, 2015). Once established, these mental linkages can be activated by subsequent exposure to stereotypical media cues, and feed into racial/ethnic prejudice and discrimination on interpersonal and intergroup levels (Atwell Seate & Mastro, 2017; Matthes & Schmuck, 2017; Schieferdecker & Wessler, 2017). Hence, as mediated stereotypes may serve to justify today's troubling

[1] University of Amsterdam, The Netherlands

Corresponding Author:
Anne C. Kroon, Amsterdam School of Communication Research (ASCoR), University of Amsterdam, 1001 NG Amsterdam, The Netherlands.
Email: A.C.Kroon@uva.nl

levels of anti-immigrant attitudes and antipathy toward “others,” it is important to study the nature and strength of stereotypical associations in the media and trace the factors that account for variations in these representations.

Studies analyzing distorted media portrayals typically focus on a selective set of marginalized minorities as represented by a limited number of media outlets over a short period of time (Mutz & Goldman, 2016). Yet, the nature and strength of stereotypical associations in newspaper content are not without change. Given that the rise of polarizing political beliefs appears to have created a media climate where unfavorable beliefs toward “others” can be openly expressed and discussed Schemer, 2012), critical questions remain regarding the extent to which media stereotypes in diverse outlets are becoming more biased over time.

The large-scale investigation of over-time variation in mediated stereotypes is becoming accessible to communication scholars due to advances in the domain of computer-assisted analyses and the availability of so-called big data samples. For example, among researchers interested in the agenda-setting theory, computational approaches are rapidly becoming common (e.g., Guo & Vargo, 2015). Computational approaches are at the same time almost non-existent in the media-stereotyping literature (cf. Arendt & Karadas, 2017). An obvious explanation for this is that, until recently, such approaches have felt short in identifying stereotypes of social groups in mass-mediated content, as stereotypes are often subtle and therefore difficult to detect by computers. However, the use of shallow or deep neural networks for natural language processing, and in particular the use of word embeddings, have made it easier to accurately detect implicit bias in media texts, as is illustrated by recent empirical examples in the field of computational science (Bolukbasi et al., 2016a; Caliskan et al., 2017; Garg et al., 2018).

This study provides a new level of empirical evidence for the nature of ethnic stereotypes in newspaper content due to the combination of advanced automated methods and the analysis of all newspaper content in Dutch leading newspapers between 2000 and 2015 ($N = 3{,}316{,}494$ newspaper texts). The study’s approach is innovative in that it introduces the use of word embeddings to the media-stereotyping domain, herewith illustrating how this scholarship can move beyond frequently employed, but overly simplistic, bag-of-words approaches. The aim of our analysis is twofold. First, the study critically tests whether media depictions are in line with key theoretical notions of prejudice following predictions of the stereotype content model (SCM; Fiske et al., 2002). By testing universal dimensions of stereotype content in newspaper content among a large set of ingroup and outgroup categories, the study’s findings allow drawing more generalizable conclusions about the nature of stereotypical news content. Second, using time-series analysis, the study contrasts over-time variation in stereotype associations with real-world integration outcomes. As such, the study provides a critical test of the question of whether the content of news media has, over time, become more biased. In sum, the study allows sustaining claims regarding the nature of stereotypes in news media content while offering novel insights into the over-time development of news bias at a previously impossible scale.

The Measurement of Representations of Ethnic Groups in News Content

Previous empirical work in the media-stereotyping domain typically relies on human coding, which demands high financial and human resources and therefore does not easily scale across diverse ethnic groups, media outlets, and time periods. Computer-assisted approaches to content analyses, allowing for the fast and affordable analyses of large-scale samples, could potentially overcome these limitations. Yet, the analysis of ethnic bias in news messages requires a challenging level of complexity: To grasp subtle nuances in stereotypical news messages, fine-grained analyses are required. As a consequence, human coding has remained the dominant method to analyze stereotypes in media content.

The few studies that do employ automated methods typically rely on the top-down, relatively easy-to-apply, dictionary approach to measuring the co-occurrence of targets and attributes (Jacobs et al., 2018; Kroon et al., 2018; Ruigrok & van Atteveldt, 2007) such as the within-sentence or within-article co-occurrence of references to Muslims and terror. Among others, this approach has been used to identify linkages between immigration news content and crime, terrorism, and socioeconomic issues (Jacobs et al., 2018). An obvious limitation of such bag-of-words approaches is that contextual semantic information is not taken into consideration, herewith introducing limitations such as a loss of domain-contingent word meanings and grammatical functions (Grimmer & Stewart, 2013).

The use of word embeddings, which is rapidly becoming among the most popular methods in natural language processing, can overcome these limitations. Shortly put, word embeddings are based on the idea that linguistic terms can be accurately represented by contextual information. Word embeddings represent words in a vector space, where words are mapped to numeric vectors. Words with similar meanings are closer to each other in this vector space. Simply put, for each word its relationship to all other surrounding words is summed (Caliskan et al., 2017), the distance (i.e., cosine similarity scores) between vectors can be measured. This idea of distributional similarity is used to predict surrounding words, based on the thought that "[y]ou get to know a word by the company it keeps" (Firth, 1957, p. 11). The embedding models that result from this training algorithm can, for example, predict that *man* is to *king* as *woman* is to *queen*.

Yet, by capturing semantics, word embeddings inevitably reveal human bias (Bolukbasi et al., 2016a; Garg et al., 2018). By learning the meaning of words based on large training corpora of human communication, the resulting models inherently reflect implicit cultural dispositions, among which some are prejudiced, as "language itself contains recoverable and accurate imprints of our historic biases" (Caliskan et al., 2017). To illustrate, the close proximity of female to homemaker and man to programmer in a vector space reveals implicit gender bias (Bolukbasi et al., 2016a). As a consequence, in the general public discourse, the blind down-stream application of such machine learning techniques is often blamed for being biased, leading to headlines such as "Google's Sentiment Analyzer Thinks Being Gay Is Bad" (Thompson, 2017). Yet, it is not Google who "thinks" being gay is bad, but—unfortunately—large parts of society: The algorithm has learned from billions of texts that gay and bad are associated in today's discourse, no matter how offensive or unjust this is.

In this article, we turn around this criticism and make use of the fact that word embeddings pick up linguistic implicit biases. This idea is not new: Studies in the field of computational linguistics and artificial intelligence have successfully employed word embeddings to detect gender and ethnic bias in large samples of texts such as Web data and Google News (Bolukbasi et al., 2016a). More specifically, these studies document that embeddings accurately reflect human bias as measured by implicit association tests (Caliskan et al., 2017), and accurately capture sociological trends (Garg et al., 2018). Within the field of communication science, word embeddings have been used to measure sentiment (Rudkowsky et al., 2018). Yet, its application to detecting ethnic bias is virtually absent (cf. Arendt & Karadas, 2017; Leschke & Schwemmer, 2019). Accordingly, the current study introduces word embeddings as a much-promising method to measure the stereotypicality of media content to the field of communication science and illustrates its use by investigating the representation of diverse ethnic ingroups and outgroups in Dutch news articles.

Trait Dimensions of Mediated Stereotypical Associations: The SCM

News content of immigration and integration in the Netherlands has been described as turbulent (Bos et al., 2016), and characterized by strong negativity and threats (Vliegenthart & Roggeband, 2007). In particular, the assimilation frame has gained popularity over time in the immigration debate, while socioeconomic emancipation and multiculturalist perspectives have fallen behind (Duyvendak & Scholten, 2012; van Heerden et al., 2014). Overall, Dutch news media have been shown to frequently link immigration to issues of terrorism, crime, and the economy (Jacobs et al., 2018; Vliegenthart & Roggeband, 2007). Other than focusing on the *frames* or *issues* associated with immigration and integration, the current study focuses on stereotypical bias in representations of diverse ethnic minorities.

Drawing generalizable conclusions about news stereotypes is complex: The nature of stereotypical media content varies considerably across research contexts as well as the specific ethnicity under investigation (Mastro, 2009). As a consequence, the formulation of universal conclusions regarding the nature of racial stereotypic news messages remains a challenge. As prior scholarship typically considers media representations of a limited number of social groups, few universal assumptions can be offered regarding the nature of ethnic stereotypes in the media (Mastro, 2009). Regardless, recent evidence suggests that more generalizable assumptions regarding the nature of media stereotypes can be made following the predictions of the SCM (Kroon et al., 2018; Sink et al., 2018). The current study builds on this work to predict implicit stereotypicality in the context of media representations of ethnicity.

The SCM posits that two universal evaluative dimensions organize stereotype content—notably, warmth (e.g., good-hearted, benevolent) and competence (e.g., competent, intelligent; Fiske et al., 2002). The categorization of social groups as relatively high or low on warmth and competence defines how we think, feel, and behave toward "others." The appraisal of social groups along the array of warmth and competence congregates into four distinct quadrants, each of which is associated with different social groups: low warmth and low competence (e.g., poor people, immigrants), low warmth and high competence (e.g., rich, Asians), high warmth and low competence (e.g., elderly, disabled), and high warmth and high competence (e.g., ingroup members and similar others; Fiske et al., 2002).

According to the SCM, warmth and competence attributions are rooted in intergroup relations related to competition and status (Cuddy et al., 2008; Fiske et al., 2007). Whether groups are viewed as cooperative or competitive is largely a question of intent: Are the group's intentions helpful or harmful? Cooperative groups are thought to have helpful intentions, and trigger high-warmth perceptions (e.g., helpful, good-hearted). Competitive groups, on the contrary, are believed to have harmful intentions, which elicits low-warmth perceptions (e.g., harmful, and untrustworthy). The perceived status of social groups evaluates the ability of groups to control resources. High-status groups, viewed as capable of obtaining resources, are seen as competent (e.g., productive, intelligent). Low-status groups, on the contrary, are thought to be incapable of controlling resources and thus receive low-competence judgments (e.g., unproductive, not smart; Cuddy et al., 2008).

As prescribed by the SCM, the origins of warmth (i.e., threat, competitiveness) and competence (i.e., status) judgments are relevant in predicting intergroup bias on cognitive, emotional, and behavioral levels (Fiske et al., 2002). News media messages might be especially informative regarding the threat/competitiveness and status of ethnic ingroups and outgroups. In support of this claim, empirical studies have documented that news media tend to represent ethnic minorities in terms of threats to economic and social resources (Eberl et al., 2018), occupying low-status socioeconomic positions (Kroon et al., 2016). Accordingly, the current study focuses on indicators of threat and status in news media messages. We posit:

Hypothesis 1 (H1): News media implicitly associate ethnic outgroups more strongly with low-status traits than ethnic ingroups.

Hypothesis 2 (H2): News media implicitly associate ethnic outgroups more strongly with high-threat traits than to ethnic ingroups.

Over-Time Variation in Mediated Implicit Stereotypical Associations

Previous research adopting an over-time perspective on the presentation of minorities in the news tends to focus on the volume of coverage, rather than the way minority groups are portrayed. These studies generally conclude that differences in the visibility of migrant groups can be explained by real-world events such as terrorist attacks and elections (Eberl et al., 2018). The few studies that explicitly modeled the influence of time document considerable variation in the stereotypicality of media content. For example, U.S.-based scholarship indicates that Latino characters on prime-time TV are increasingly sexualized over the years (Tukachinsky et al., 2015). In the Netherlands and Flanders, evidence exists for an erratic over-time pattern of news coverage which relates immigration to crime and terrorism (Jacobs et al., 2018).

However, it remains an open empirical question to what extent media stereotypes of diverse ethnic ingroups and outgroups change over time, and if such changes are in sync with real-world developments. We expect that the stereotypicality of news media content about ethnic out-groups increase over time for the following reasons. First, news media content is likely to mirror increasingly unfavorable public perceptions about minority groups: As asserted by diverse sociologists, the shifting equilibrium between the host population and out-group members due to immigration influxes has amplified negative sentiments toward ethnic outgroups across Europe in the past decades (see Gorodzeisky & Semyonov, 2016). Owing in part to increased (perceived) fears of competition over socioeconomic resources, threat perceptions about minority groups have intensified (Erisen & Kentmen-Cin, 2017; Semyonov & Glikman, 2008). Media scholarship documents that these negative sentiments and threat perceptions are reproduced by news content about ethnic outgroups (Eberl et al., 2018). In addition, the rise of popular right across Europe represents a significant and important real-world trend that journalists logically cover. In doing so, however, news media messages may, to an increasing extent, offer a stage for right-wing politicians to voice anti-minority opinions and report on the anti-minority viewpoints put forward by such parties (Vliegenthart & Boomgaarden, 2007).

Following this argumentation, it can be expected that aggregated news media's representation of ethnic groups is rather a product of public opinion and the political climate than a true reflection of actual, numerical integration outcomes. Contrasting the content of media coverage with real-world trends provides insight into the extent to which the unfolding of media representations diverges from factual real-world integration outcomes and herewith provides a true test of media bias. Previous scholarship that put such inter-reality comparisons to the test finds support for the bias hypothesis: Empirical evidence documents that negative media representations of marginalized groups diverge from real-world statistics (Dixon & Linz, 2000; Dixon & Williams, 2015; Jacobs et al., 2018). For example, Jacobs et al. (2018) find that news about immigration is largely unaffected by real-world figures such as crime and socioeconomic issues.

The available inter-reality comparisons tend not to consider over-time variation in media and real-world data, or rely on the involvement of general indicators without explicitly modeling the representation of minority groups among real-life statistics. The complex over-time interaction between news stereotypicality and the representation of diverse ethnic groups among real-world statistics, however, has remained unaddressed by previous media scholarship. The current study includes two real-world indicators that mirror high-threat and low-status stereotypes: criminality rates and the reception of social benefits. The following hypothesis and research question are formulated:

Hypothesis 3 (H3): Over time, the strength of implicit stereotypical associations in news content will increase for ethnic outgroups, while this will not be the case for ethnic ingroups.

Research Question 1 (RQ1): To what extent do real-world integration outcomes (i.e., representation of ethnic groups among criminality rates and the distribution of social benefits) affect the evolution of in implicit stereotypical associations the news media content?

Method

Data preparation, model training, and subsequent analyses were conducted using Python. We used the Word2vec implementation provided by the gensim package (Mikolov, Corrado, et al., 2013; Mikolov, Yih, & Zweig, 2013; Řehůřek & Sojka, 2010). We then used R for hypothesis testing and for creating our final visualizations. Code used to train the models is available here: https://github.com/annekroon/mediabias.

Word embeddings are especially effective at solving word analogies, as they capture semantic word relations with mathematical relationships between vectors. The family of word-embedding techniques called Word2vec has gained popularity after its release in 2013 (Mikolov, Corrado, et al., 2013). This unsupervised machine-learning algorithm takes a large collection of documents as an input (think of hundreds of thousands of news articles, or all articles published on Wikipedia in a given language) and represents a word as a vector of weights across a set of *k* dimensions. In the current study, distributed weights across 100 dimensions are calculated for each word in the training corpus. Word2vec is a shallow neural network, in which the input layer (i.e., the features in the corpus that the model is trained on) is mapped on an intermediate hidden layer, which is then mapped on the output layer, in our case the vector space representations (for an introduction, see, e.g., Goldberg, 2017). We apply a continuous-bag-of-words model (CBOW), meaning that vector representations of each of the target words are learned from its context (i.e., its direct neighboring words). More specifically, in the current study, we look at vocabulary words occurring within five words of each other. In conclude, the final embedding model returns a distribution of weights over 100 dimensions for each word in our training corpus.

Thus, embedding models learn the meaning of words based on the context in which these words occur across its many occurrences in a training corpus (in the current study: newspaper articles). In this process, words that are similar are mapped (i.e., embedded) to nearby points in the vector space, sharing high cosine similarity values. Such neighboring words in a vector space can be synonyms or words that are used in comparable contextual or topical domains (Bolukbasi et al., 2016b). Intuitively, take the example of the word “cereal.” This word will likely occur in similar sentences that refer to “oats,” such as “I like to eat [cereal/oats] in the morning.” As a result, cereal and oats share semantic meaning, and will be close neighbors in the embeddings’ vector space.

Using word embeddings as a diagnostic tool for bias detection allows us to explicitly move beyond deductive, dictionary approaches that, for example, aim to count how often the word “immigrant” co-occurs with the word “terrorism.” Although such dictionary approaches are informative regarding the co-occurrence of stereotypical terms in specific news articles, they do not consider contextual information and neglect semantic meaning. Alternatively, word-embedding models necessarily encode bias present in the training corpus by learning word meaning based on the semantic context in which social and ethnic groups appear. Word-embedding models have therefore been referred to as an “AI stereotype catcher” (Greenwald, 2017), useful to identify largely implicit forms of bias that may arise even in the absence of blatant and explicit prejudiced accusations. Likewise, influential studies in the field of AI show that semantic nearness of social categories (e.g., man, woman) and attributes (e.g., programmer, nurse) replicate implicit bias as measured by implicit association tests (Caliskan et al., 2017; Garg et al., 2018).

Data and Training

The current study draws on the entire corpus of news articles that appeared in the five Dutch national newspapers with the highest circulation rate (*de Volkskrant*, *NRC Handelsblad*, *Trouw*, *Algemeen Dagblad*, *De Telegraaf*) from January 2000 up to and including December 2015 ($N = 3{,}316{,}494$ newspaper articles). The corpus includes the full range of article types available in these newspapers: news stories, but also op-eds, columns, and editorials. Of the newspapers in our sample, *Algemeen Dagblad* and *De Telegraaf* are often considered as tabloid-like, popular newspapers. On the contrary, *de Volkskrant*, *NRC Handelsblad*, and *Trouw* have generally been considered quality newspapers (Boukes & Vliegenthart, 2020; Roggeband & Vliegenthart, 2007). Especially the inclusion of tabloid-like newspapers might contribute to implicit bias in our corpus, as previous research indicates that European tabloid newspapers are more prone to represent ethnic minorities in stereotypical terms (Arendt, 2010; Kroon et al., 2016; Van Dijk, 2000). Covering several key events such as the aftermath of 9/11, the influx of immigrants and the rise of extreme right across Europe, newspaper content in the time frame under study is likely to resonate with a diverse range of sources, opinions, and arguments regarding ethnic minorities.

Model training was done in two steps. First, and to test our time-invariant hypotheses about the nature of media stereotypes, we train a single embedding model on the entire corpus. Second, and to allow testing of time-varying hypotheses, in a second step we trained word embeddings on consecutive years of the selected newspaper articles. The computer thus “learns” the meaning of words for each year separately, allowing for the detection of over-time variation in stereotypical associations. Thus, for each available year of news content, a single embedding model was trained (2000–2015), resulting in a set of 16 embedding models. To prepare our corpus for model training, we converted the entire text to lowercase sentences and removed numbers. Subsequently, each news article was split into sentences. Algorithmically, during model training, each sentence is processed to predict target words (e.g.,

Table 1. Frequencies, High-Threat and Low-Status Associations Across Ethnic Categories.

Ethnic categories	Group membership	High-threat association	Low-status association	Frequencies
Belgian	Ingroup—ethnicities	0.17	0.07	42,270
German	Ingroup—ethnicities	0.22	0.09	73,981
Dutch	Ingroup—ethnicities	0.24	0.17	242,449
Christian	Ingroup—labels	0.15	0.17	36,025
Western	Ingroup—labels	0.15	0.21	10,936
Civilian	Ingroup—labels	0.24	0.19	161,293
Polish	Outgroup—ethnicities	0.22	0.14	50,390
Surinamese	Outgroup—ethnicities	0.31	0.31	6,482
Turkish	Outgroup—ethnicities	0.34	0.23	32,101
Syrian	Outgroup—ethnicities	0.35	0.26	2,435
Antillean	Outgroup—ethnicities	0.36	0.30	6,221
Moroccan	Outgroup—ethnicities	0.36	0.28	20,106
Somali	Outgroup—ethnicities	0.37	0.34	1,352
Afghan	Outgroup—ethnicities	0.39	0.24	7,729
Iraqi	Outgroup—ethnicities	0.41	0.21	10,779
Foreigner	Outgroup—labels	0.15	0.29	23,877
Migrant	Outgroup—labels	0.28	0.29	19,904
Muslim	Outgroup—labels	0.30	0.26	67,757
Refugee	Outgroup—labels	0.31	0.25	42,117
Immigrant	Outgroup—labels	0.33	0.28	21,292
Arab	Outgroup—labels	0.34	0.26	9,596

Note. Frequencies refer to the total number references to an ethnic category in the total sample of newspaper articles.

"dog") from source context words (i.e., neighboring words, for example, "I walk my ___ everyday").

Accuracy of the embeddings. Several steps were taken to warrant the quality of the baseline embeddings model. First, we verified that references to target categories in our dataset are frequent (> 6K references, see Table 1), warranting the quality of the embeddings we are most interested in (Schnabel et al., 2015). Next, we perform a word analogy task (Mikolov, Yih, & Zweog, 2013; Pennington et al., 2014), which is among the most popular methods to evaluate the quality of word embeddings (Schnabel et al., 2015). This method is based on the idea that humans should be able to predict mathematical operations in a vector space: When given three words (*a*, *b, c*), knowing that *a* is to *b* as *c* is to *d*, one should be able to identify the target word *d* (Schnabel et al., 2015). The most-cited example that illustrates this principle is probably the task to complete the sentence "*Man* (*a*) is to *king* (*b*), as *woman* (*c*) is to _____." The target word (*d*) that we are looking for is *queen*. We use this vector arithmetic to solve 1,424 analogies about common capitals and countries, family relations, and comparisons[1], resulting in a mean accuracy of 66.62%. This is comparable to the quality of previously employed embeddings (Pennington et al., 2014).

Attribute Word Lists Reflecting Low-Status and High-Threat Stereotypes

To measure low-status and high-threat associations, words tapping into these concepts need to be identified. We retrieve words capturing both concepts from the data under investigation, as this is most representative of the language and jargon used by journalists. We use a bottom-up approach to establish word lists using most similar scores retrieved from embeddings trained on the corpora of all news content per news outlet. In this way, we capture inter-media variation in most similar scores while maintaining large-scale accuracy benefits (Mikolov, Corrado, et al., 2013). For each ethnic category formulated in both singular and plural form (e.g., *Moroccan* and *Moroccans*), the 100 most-similar words in the vocabulary were retrieved. Most-similar words are words that are closest to the target word in the vector space.

The resulting word lists were subsequently manually revised and categorized by the authors as follows. We considered only words that carry a negative meaning and could, therefore, reveal a certain stereotype of an ethnic group. Negative words were categorized into two groups: high-threat indicators and low-status indicators, following the conceptualization put forward by Fiske et al. (2002). *High-threat indicators* were defined as words referring to hostility, deviance, threatening behavior or objects, criminal, and/or illegal activities (e.g., stealing, robberies, and murder). We also included words related to judicial authorities (e.g., policeman or cop), as we believe these words evoke associations with hostility and criminality. *Low-status indicators* were defined as words referring to low-social class, (un)intelligence, low education levels, unemployment, addiction, and/or homelessness. Words with a negative connotation that did not fit either of these categories were not included in the analysis. We excluded issue-specific words that are connected to foreign affairs, specific events or issues, and/or a selective set of ethnic categories (such as dictators, child soldiers, and radicalism), as we are explicitly interested in the general dimensions of stereotype content. During several rounds, the selection of words was critically debated and carefully revised by the authors. The final list of words capturing high-threat ($N = 208$ words) and low-status ($N = 95$ words) stereotypes are included in Appendices A and B.

Targets Word Lists Reflecting Ethnic Categories

A target word list is created capturing diverse ethnic categories. We included the eight largest non-western ethnic groups living in the Netherlands (i.e., Surinamese, Turkish, Syrian, Antillean, Moroccan, Somali, Afghan, and Iraqi). In addition, we included three large western ethnic groups living in the Netherlands (i.e., Belgian, German, Polish). Last, the Dutch were included. For reasons of completeness, labels often used in the news media to denote ethnic categories were also included (e.g., immigrants, foreigners—see below). For each ethnicity and each label both the singular and plural forms were included in the target word list (e.g., Moroccan, Moroccans, immigrant, immigrants). See Table 1 for an overview.

Variables Time-Invariant Prediction of Media Stereotypes

Ethnic categories and labels. In addition to the Dutch, we consider Germans and Belgians members of the ethnic ingroup, as they are geographically, linguistically, and culturally strongly related to the Netherlands. All other ethnicities are considered ethnic outgroups. Ethnic labels are also categorized as outgroup (i.e., foreigner, migrant, Muslim, refugee, immigrant, Arab) and ingroup (i.e., Christian, Western, Civilian) categories.

Implicit high-threat and low-status association strength. Using the baseline word-embedding model and the above-defined attribute and target word lists, the association strength between ethnic groups and stereotypical attributes must be computed. To this aim, a Python script was developed to retrieve similarity scores for all combinations of ethnic categories and the words representing respectively high threat and low status using the baseline word-embedding model. The script returns the cosine similarity for each pair. We calculate the average embedding distance between words representing ethnic groups and high-threat attributes. We also calculate the average embedding distance between words representing ethnic groups and words representing low-status stereotypes. Higher scores indicate stronger implicit associations.

Variables Time-Varying Prediction of Media Stereotypes

Group membership. For the dynamic analysis, we focus only on a small subset of the ethnic categories as real-world indicators were only available for these groups. Accordingly, we consider the four largest immigrant groups in the Netherlands as outgroups (i.e., Surinamese, Antillean, Turkish, Moroccan) and the Dutch as ingroup. A dummy variable is created differentiating between outgroup (1) and ingroup (0) category.

Implicit stereotype-association strength. Aiming to explain over-time variation in general patterns of implicit media bias, we do not differentiate between high-threat and low-status dimensions of stereotype content.[2] Instead, we assume that both dimensions represent negative stereotypical associations. More specifically, using the dynamic word-embedding models and the target and attribute word lists, we calculate the mean embedding distance for words representing ethnic groups and both high-threat and low-status attributes.

Year trend. Years received a value ranging from 1 (year 2005) to 11 (year 2015).

Real-life high-threat indicator: Criminality rates. The yearly share of registered suspects of crime by background was used as an indicator of the actual threat posed by different ethnic groups, obtained from Statistics Netherlands ($M = 4.61$, $SD = 2.17$).

Real-life low-status indicator: Unemployment benefits. The yearly share of unemployment benefits by background was used as an indicator of the actual social status of different ethnic groups, obtained from Statistics Netherlands (M = 3.05, SD = 0.95).

Demographic composition. As an indicator of demographic composition, we rely on the yearly share of the total population of the Netherlands by ethnic background, obtained from Statistics Netherlands (M = 9.15, SD = 0.35).

Analysis

To test our time-invariant hypotheses regarding the nature of media stereotypes (**H1**, **H2**), we opt for analyses of variances (ANOVA) to compare means in implicit stereotype-association strength across ethnic categories. Due to the limited availability of data capturing real-life indicators, we consider the embedding models on the years 2005 up to and including 2015 in our analysis predicting over-time variation in general patterns of news media bias. To test our time-variant hypothesis and research question (**H3**, **RQ1**), we aggregate the data to yearly observations for the ethnic groups (11 years × 5 groups = 55 observations). Due to the pooled structure of the data (i.e., yearly observations for ethnic groups), it is important to consider issues related to panel differences and autocorrelation. It was assured that the mean of the dependent series (i.e., implicit stereotype-association strength) was unaffected by changes in time. To this aim, the Levin–Lin–Chu test was used, which corresponds to a pooled Augmented Dickey–Fuller test, and provides an overall assessment of unit root. The result suggests that for each series the null hypothesis of non-stationarity can be rejected, $\chi^2(3) = -7.86$, $p < .001$. Next, patterns of heterogeneity need to be investigated by inspecting fixed effects (Kittel, 1999; Wilson & Butler, 2007). Fixed-effects analysis including all the independent variables indicate significant fixed effects for our dependent variable, $F(3, 46) = 120.8$, $p < .001$. Accordingly, a model capturing the dynamic structure of each target group is desirable. To account for the heterogeneity in our data, we have to choose between random or fixed-effects analysis. Both types of models are compared using the Hausman test, which examines if systematic differences in coefficients can be detected. This test indicates little differences between both models, indicating that both the fixed and random effects yield largely comparable results. As random effect models are slightly more efficient they are preferred. In addition, the fixed-effects model's error structure suggests the presence of panel-level heteroscedasticity, Wald $\chi^2(4) = 50.48$, $p < .001$. This informs us that across ethnic groups (i.e., the panels) the level of variance of the variables differs. In sum, the combination of heterogeneity and heteroscedasticity, as well as the data structure (relatively small N of target groups and T), suggests that the data should be analyzed using ordinary least squares regression with panel corrected standard errors (OLS-PCSE).

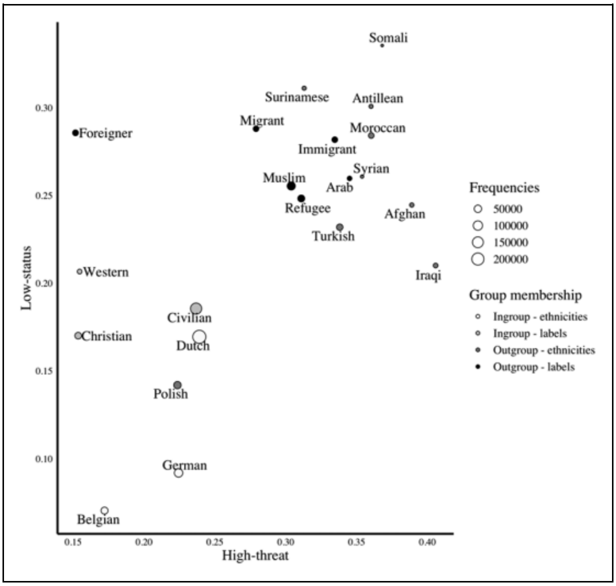

Figure 1. High-threat and low-status associations across ethnic categories and labels.

Results

The Nature of Media Stereotypes: Implicit High-Threat and Low-Status Associations

The word-embedding model trained on the corpus of all news items was used to investigate high-treat and low-status associations across ingroups and outgroups. Figure 1 and Table 1 summarize the descriptive results. As can be seen, ingroup categories (in terms of both labels and ethnicities) are primarily situated in the neutral threat and neutral status quadrant of the figure. Notably, high-threat and low-status associations are relatively weak for Germans, Belgians, Christians, and the Dutch. Outgroup categories (both in terms of labels and ethnicities), on the contrary, are mostly situated in the high-threat, low-status quadrant of Figure 1. High-threat and low-status associations are especially strong for Antilleans, Moroccans, and Somalis. In addition, the label immigrant is strongly associated with high-threat and low-status portrayals. The Polish, however, are an exception—this outgroups scores relatively low on both dimensions of threat and low-status/social inferiority. Table 1 confirms that the association with low-status and high-trait stereotypes is lower for ingroups (i.e., ethnicities and labels) compared with outgroups (i.e., ethnicities and labels). These descriptive findings largely confirm the predictions put forward by SCM scholarship.

Next, we investigate whether descriptive differences in stereotype strength across group membership hold when testing for statistical significance. It was anticipated

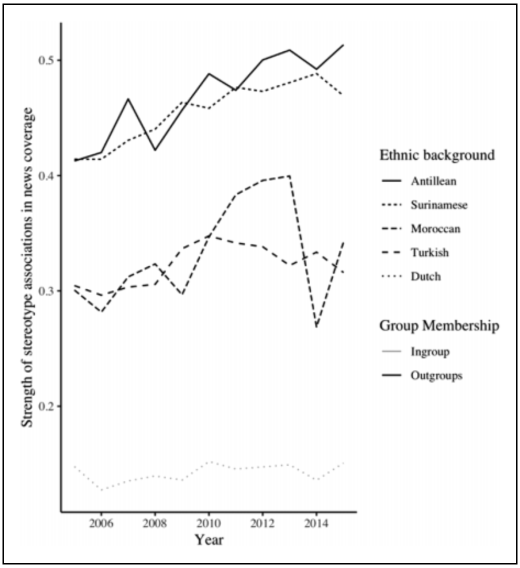

Figure 2. Over-time variation in stereotypical associations across group membership.

that news media implicitly associate outgroup ethnicities compared with ingroup members more strongly with high-threat stereotypes (**H1**) and low-status stereotypes (**H2**). Separate analyses of variances (ANOVA) reveals a stronger association between high-threat stereotypes and outgroup members ($M = 0.35$, $SD = 0.05$) than ingroup members ($M = 0.21$, $SD = 0.04$), $F(1, 10) = 16.15$, $p < .01$, $\omega^2 = .53$. In addition, we also find that outgroup members ($M = 0.26$, $SD = 0.06$) are more strongly associated with low-status stereotypes than ingroup members ($M = 0.11$, $SD = 0.05$), $F(1, 10) = 14.61$, $p < .01$, $\omega^2 = .53$. These findings confirm **H1** and **H2**, and signal the portrayal of outgroup members along the lines the SCM.

Time-Varying Prediction of Implicit Stereotype-Association Strength

Finally, we discuss the analysis explaining over-time variation in implicit stereotypical associations. Here, we use the word embeddings trained on consecutive years of news content (2005–2015). Figure 2 displays the over-time variation in the strength of implicit stereotype associations for the selected outgroup and ingroup categories. In support of our expectations, we see that generally, implicit stereotype-association strength increases over time for most of the ethnic outgroups.[3] Conversely, the association between the ethnic ingroup and implicit stereotype-association strength remains stable across time. Table 2 displays the results of the OLS-PCSE analysis. Model 1 shows the main effects of the predictor variables. The results show that, in agreement with our descriptive analysis, ethnic outgroups receive stronger stereotypical associations than the ethnic ingroups. Time does not exert a significant main

Table 2. Time-Varying Prediction of Stereotype-Association Strength.

	Model 1	Model 2	Model 3
(Intercept)	−1.959 (1.730)	−1.861 (1.771)	−1.283 (1.758)
Group membership	0.135*** (0.023)	0.076** (0.025)	−0.205 (0.132)
Year trend	−0.023 (0.022)	−0.030 (0.022)	−0.027 (0.019)
Criminality	0.019*** (0.003)	0.021*** (0.003)	−0.108 (0.068)
Unemployment benefits	0.040* (0.016)	0.042* (0.016)	−0.003 (0.070)
Demographic composition	0.229 (0.201)	0.222 (0.206)	0.186 (0.196)
Group membership × year trend		0.010*** (0.001)	0.010† (0.005)
Group membership × criminality			0.130† (0.068)
Group membership × unemployment benefits			0.049* (0.019)
N	55	55	55
R^2	.794	.805	.810
Adjusted R^2	.773	.780	.776
Residual *SD*	0.058	0.057	0.058

SD = standard deviation.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.$^{***}p < .001$.

effect on our dependent series. We do find positive effects of both the evolvement of criminality rates and unemployment benefits on the strength of implicit stereotypical associations. Finally, our control variable demographic composition does not significantly affect our outcome variable.

It was anticipated that across time, implicit stereotype-association strength would increase for the ethnic outgroup, but not for the ethnic ingroup (**H3**). To test this assumption, we include the interaction term of group membership (ethnic outgroup vs. ingroup) and the time trend in Model 2. We find a significant interaction between group membership and time, such that when time increases, the implicit stereotype-association strength increases for outgroups, but decreases for ingroups. We accept **H3**.

Last, we asked to what extent real-world integration outcomes (i.e., criminality rates and unemployment benefits) are related to the evolution of stereotype-association strength. To this end, we included interaction terms between group membership and the integration outcomes to our final model. The results indicate that the effects of criminality rates and the division of unemployment benefits on implicit stereotype-association strength in the news media are different for ingroup and outgroup categories. The inspection of the interaction terms reveals that the positive effect of criminality on implicit stereotype-association strength is stronger for the ethnic ingroup compared outgroups. More specifically, the representation among criminality rates matters for the news media portrayal of ethnic ingroups but not for ethnic outgroups: Regardless of the actual decreasing criminality rates among outgroups, evaluations of these groups in the media did not decline. Regarding the interaction between the share of unemployment benefits and group membership, a

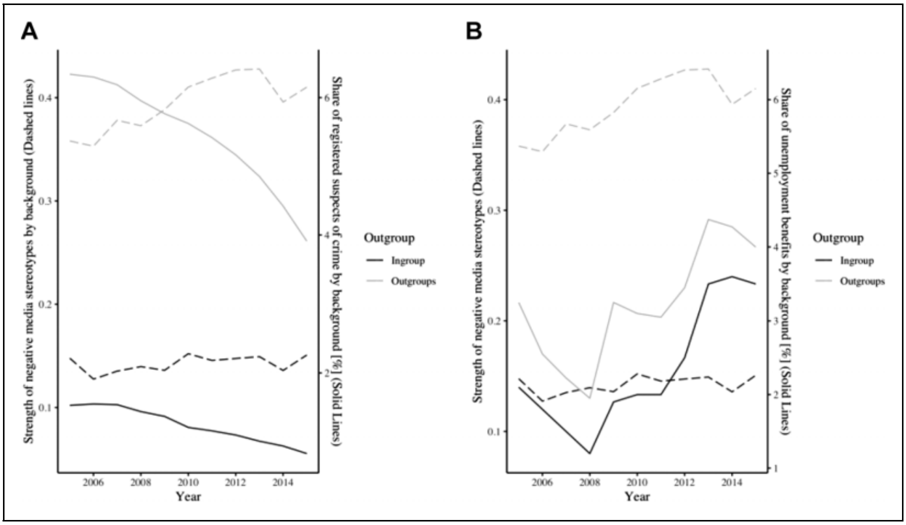

Figure 3. .Inter-reality comparison of integration outcomes and stereotype-association strength: (A) criminality and (B) unemployment benefits.

different pattern emerges: The positive effect of unemployment benefits on implicit stereotype-association strength is stronger for ethnic outgroups than for the ingroup. Raising unemployment benefits among ethnic outgroups results in stronger implicit stereotype associations, while this is not the case for ethnic ingroups. Figure 3 illustrates these findings by juxtaposing the evolution of the factual integration proxies and implicit stereotypical associations in news content across group membership.

Discussion

The twofold aim of this study was to investigate universal dimensions of stereotype content in news content across a diverse set of ethnic ingroups and outgroups, and to trace the extent to which ethnic bias in news coverage has increased over time. To answer these questions, the current study introduces the use of word embeddings to the media-stereotyping scholarship as a promising method for detecting implicit bias in news content. Relying on an analysis of more than 3 million Dutch news articles, the current study finds strong support for the key dimensions of stereotype content as put forward by SCM scholarship (Fiske et al., 2002). In addition, the data show that, across time, content about ethnic outgroups has become progressively negative and remote from factual integration outcomes.

We first discuss our findings regarding the nature of media stereotypes. Importantly, the study shows that universally accepted dimensions of stereotype content as documented by an astounding body of psychological studies (see Cuddy et al., 2008, 2009) can be replicated using news media data. More specifically, the presented results show that news media portray ethnic outgroups in terms of fundamental dimensions explaining shared perceptions of societal groups: Compared with ethnic ingroups, news content proved to implicitly associate ethnic outgroups relatively strongly with low-status and high-threat stereotypes.

These results hold across a diverse set of ethnic subgroups. Implicit associations with low-status and high-threat stereotypes were significantly weaker for ethnic ingroups, defined as native citizens (i.e., the Dutch) as well as residents of immediate neighboring countries (i.e., Germans and Belgians). Ethnic outgroups from non-European soil received uniformly negative evaluations on both dimensions. Particularly, ethnic outgroups like Moroccans, Somali, and Antillean could uniformly be positioned in the high-threat, low-status quadrant of stereotype content. This indicates that news media portray immigrant subgroups largely in terms of dominant social perceptions about them (Fiske, 2012; Lee & Fiske, 2006). In sum, the findings are largely in line with predictions of the SCM (Fiske et al., 2002), and herewith offer strong support for the dual-dimensionality of stereotype content in newspaper content.

A few exceptions emerged. We found that implicit low-status and high-threat associations were weak for the Polish, the single European outgroup in our sample. This finding might be a reflection of the generally less pronounced negative attitudes toward European compared with non-European immigrants (Gorodzeisky & Semyonov, 2009). In addition, we found strong low-status but weak implicit threat associations for the generic label "foreigner," revealing ambivalent stereotype content. Semantically, this generic label might comprise a range of people who are not perceived as threatening such as tourists or expats.

These findings have important societal ramifications. Status and threat indicators are core components of social judgments: Status informs individuals about the ability and competence of others—while perceived threat signals intentions—boiling down to the question whether "they" can be seen as friend or foe (Fiske et al., 2007). Consequently, low-status groups tend to be evaluated as incompetent, while threatening groups are judged as low in warmth. Especially in race-segregated societies, where close inter-racial contacts are the exception rather than the rule (Musterd, 2005), news media content is especially apt to inform individuals about the status and potential threat posed by ethnic outgroups. Our findings indicate that such mediated information is biased, and likely feeds into dominant ethnic stereotypes. This is especially problematic for groups that are frequently covered in Dutch news media such as Moroccans, Muslims, and immigrants in general.

Consequently, the widespread perception of ethnic outgroups in terms of low-status/low-competence and high-threat/low-warmth, which has been convincingly documented across diverse countries and cultures (Cuddy et al., 2008), seems to be—at least partly—rooted in the everyday news we consume. Such stereotypical perceptions are not inconsequential: SCM scholarship indicates that groups that are perceived as threatening and low in status are generally disliked, elicit feelings of pity and disgust, and tend to be excluded from diverse parts of society. Taken together, by spreading low-status and high-threat stereotypes of ethnic outgroups, news media might actively contribute to the maintenance of inequality and race-based exclusion in society.

Next, we discuss our findings regarding the over-time variation in media stereotypes. Our analysis shed light on long-term developments in real-world integration outcomes and implicit stereotype-association strength of the ethnic ingroup and major ethnic outgroups in the Netherlands. The results show that across time, negative associations show a slight upward trend for ethnic outgroups. In contrast, the evaluation of the ingroup remained relatively stable across time. By juxtaposing these findings to numeric integration outcomes, the study finds that across time, media representations of ethnic outgroups are becoming progressively remote from real-world statistics. First, the representation of ethnic outgroups among criminality rates (considered a proxy for the actual threat posed by these groups) decreased substantially during the studied time frame. Conversely, the implicit stereotype-association strength of ethnic outgroups displayed an opposite trend by becoming progressively stronger with time. Second, the over-time development of unemployment benefits (included as a proxy for actual social status) showed similar trends for ingroup and outgroup categories. However, increasing levels of unemployment benefits resulted in stronger implicit stereotype associations for outgroups, while this was not the case for ingroups. These results indicate that ethnic stereotypes in news content progressively diverge from real-world trends.

Importantly, these results indicate that the biased portrayal of ethnic minorities is rather an artifact of the news production process than a true reflection of what is actually happening in society. This conclusion is in line with previous studies that show that media content of ethnic minorities diverges from real-world data (Dixon & Linz, 2000; Jacobs et al., 2018). Based on a large-scale content analysis of Dutch media coverage on immigration, Jacobs et al. (2018) conclude that trends in immigration news develop remotely from trends in society. Our findings are in line with these conclusions. Several explanations for such a disconnected media reality can be offered. First, it might result from journalists' tendency to report on specific newsworthy events, clearly demarcated in space and time, rather than general trends, which is evidenced by journalists' reliance on episodic rather than thematic frames (Papacharissi & de Fatima Oliveira, 2008). Second, intensified negative sentiments regarding ethnic outgroups (Eberl et al., 2018), accelerated by the rise of the extreme right across Europe, seem to have permeated in the news content. Likewise, prevailing news values might have encouraged the production of news stories that focus on conflict: Reporters often find themselves writing about conflict when covering the immigration beat—especially regarding stories outside European/U.S. borders. Third and last, news media messages may, to an increasing extent, offer a stage for right-wing politicians to voice anti-minority opinions and report on the anti-minority viewpoints put forward by such parties (Boomgaarden & Vliegenthart, 2007). This means that, in their pursuit of offering balance news, journalists may find themselves chronicling the voice of those who might perpetuate stereotypes.

Following these considerations, an important question revolves around the notion of objectivity. More specifically, one may wonder *how* reporters can counteract biased news content. Although this question is not easily answered, we argue that it is important for reporters to be aware of structural inequalities and stereotypes on the societal level—and the possible impact hereof on the viewpoints of sources and the presumed newsworthiness of events. In this light, it is also important to consider which stories are *not* being told and which sources might not be cited to understand how bias seeps into the news.

The study makes the following methodological contributions. Importantly, the current study demonstrates the use of word embeddings to analyze subtle biases in large-scale mass-mediated texts. We have shown that word-embedding models, which are increasingly used in other fields, but new to communication science, are a promising opportunity for the detection of biases and stereotypes. In particular, they allow us to move beyond simplistic dictionary approaches that can measure if two words co-occur together, but not how similar they are. In sum, the here-reported findings are in line with recent scholarship that finds that human bias, as reflected in human language, can be accurately picked put up by word embeddings.

It should be noted that the tabloid-like newspapers in our sample could have contributed more strongly to the here-reported disparities in representations. Especially, anti-minority sentiment might resonate stronger in tabloid newspapers due to stronger affiliations with right-wing political parties (see Arendt, 2010; Kroon et al., 2016; Van Dijk, 2000). Finally, it is important to stress that our findings are the likely outcome of forces influencing the media agenda, such as routines, real-world events, and sources (Reese, 2001), rather than merely conscious or unconscious bias on the part of individual journalists.

Like all studies that employ new and innovative methods, ours has some limitations. Most notably, even though word embeddings have been successfully used before to quantify biases in texts (Bolukbasi et al., 2016a; Caliskan et al., 2017; Garg et al., 2018), we need more systematic validation studies. For instance, one could imagine that the ingroup is less often explicitly named but obvious for the reader from the context—an omission that might result in less quality word embeddings. Whether this is the case or not is an empirical question to be addressed in future work. In addition, by relying on large amounts of training data, this study did not answer the question *of how* and *why* bias exactly arises in our news corpora: It remains unclear which source types, topics, and articles mostly contribute to here-reported bias. Scholars are exploring the possibilities of tracing the origins of bias in embedding models back to the article level (Brunet et al., 2019). Future research could further explore such approaches, potentially combined with manual coding, to better understand how bias surfaces at the sentence or article level. Such a detailed, in-depth investigation of bias could be supplemented with journalists' perspective on how and why such disparities in representations are encoded in news content. Last, the data used in this study provide merely insights into general representations in media content. Motives, reasons, and/or the presence of unconscious bias on the part of journalists cannot be inferred from the here-presented data.

In addition, our findings offer only support for the dual-dimensionality of stereotype content among ethnic outgroups, who are evaluated negatively on both dimensions. Future research should investigate the extent to which these dimensions are also useful to understand the portrayal of social groups receiving ambivalent stereotypes such as the elderly and disabled people (see Fiske et al., 2002).

Despite these limitations, the here-reported findings contribute to the formulation of universal dimensions of stereotype content, generalizable beyond specific ethnic categories. Moreover, the findings reveal that the biased portrayal of ethnic minorities is rather an artifact of news production processes than a true reflection of what is actually happening in society. A more thoughtful and accurate representation of minority groups in terms of the issues and topics associated with these groups may promote more favorable attitudes toward ethnic others, and pave the way for more inclusive societies.

Appendix A

Words Capturing High-Threat Stereotypes

Dutch: *afperser, agent, agente, arrestant, arrestanten, autodief, autokraker, bajesklant, bandiet, bandieten, bankovervaller, bankrover, bedelaar, bedreiger, bende, bendeleden, bendeleider, bendelid, benden, bendes, beroepscrimineel, berovingen, beschieting, beul, boef, bolletjesslikker, bolletjesslikkers, bommenmaker, bordelen, brandstichter, brandstichters, corrupt, criminaliteit, crimineel, criminelen, cyberpesten, dader, daders, delinquent, delinquenten, dief, draaideurcrimineel, drugsbaas, drugsbaron, drugsbende, drugsbendes, drugscrimineel, drugsdealer, drugsdealers, drugsgebruikers, drugshandelaar, drugshandelaars, drugssmokkelaar, dubbelagent, ftetsendief, gangster, gangsterbende, gedetineerde, gedetineerden, gegijzelden, gevangenbewaarders, gevangene, gevangenen, gevangenisbewaarder, gevangenissen, geweldsman, gijzelaar, gijzelaars, gijzelnemer, gijzelnemers, handlanger, hardrijder, hoofdagent, hoofdagente, hoofddader, hoofdverdachte, hooligan, hooligans, huurmoord, huurmoordenaar, illegalen, inbreker, indringer, jeugdbende, jeugdbendes, jeugddelinquent, kaper, kapers, kidnapper, kidnappers, kinderlokker, kindermisbruiker, kindermoordenaar, kindslaven, krijgsgevangenen, kruimeldief, kunstdief, ladykiller, lastpak, lastpost, liquidatie, loverboy, lovergirls, lustmoordenaar, maffia, maffiabaas, maffiosi, maffioso, maftabaas, massamoordenaar, massamoordenaars, mededader, medegedetineerde, medegevangene, medeplichtige, medeverdachte, mensenhandelaren, mensensmokkelaar, mensensmokkelaars, messentrekker, misdaden, misdadig, misdadiger, misdadigers, misdadigerwapenhandelaar, moordenaar, moordenaars, moordernaar, moordernaars, moordmachine, moordverdachte, motoragent, neerstak, neersteken, ontvoeringen, oplichter, overvaller, pedoftel, pedoftelen, piraten, plunderaar, plunderaars, politieagent, politieagente, politieagenten, politiecommandant, politiegeneraal, politiegewonde, politieman, politiemannen, politiemensen, politieofficier, politiepost, politierechercheur, politiestaat, politievrouw, poltiemensen, pyromaan, recidivist, relschopper, relschoppers, roofmoord, roofoverval, scherpschutter, schutter, seriemoordenaar, skinheads, slaaf, slachtoffers, slaven, sluipschutter, sluipschutters, smokkelaar, smokkelaars, snelheidsduivel, souteneur, stalker, straatbende, strafbaar, strafklacht, struikrover, tasjesdief, terreurgroep,* *terreurverdachte*, *terrorist*, *uitbuiting*, *vechtersbaas*, *veelpleger*, *veelplegers*, *veiligheidsagent*, *veiligheidsagenten*, *veiligheidspolitie*, *verdachte*, *verkrachter*, *vermisten*, *voortvluchtige*, *vreemdeling*, *vrouwenhandelaar*, *wapenhandelaar*, *winkeldief*, *winkeldievegge*, *wreker*, *wurgmoord*, *zakkenrollers*, *zedendelinquent zedendelinquent*, *zedendeliquent*, *zelfmoordenaar*, *zwartrijder*.

English translation: a better, accomplice, arms dealer, arrested, arrestees, arsonist, arsonists, art thief, avenger, bandit, bandits, bank robber, beggar, bicycle thief, body packer, body packers, bomb maker, brothels, burglar, carjacker, car thief, chief officer, child abuser, child murderer, child predator, child slaves, co-suspect, contract killer, contract killing, cop, corrupt, crime, crimes, criminal, criminal complaint, criminal gun dealer, criminals, crook, cyber bullying, delinquent, delinquents, detainee, detainees, double agent, drug baron, drug criminal, drug dealer, drug dealers, drug gang, drug gangs, drug lord, drug smuggler, drug traffickers, drug users, executioner, exploitation, extortioner, fellow prisoner, female cop, female head agent, female trafficker, fighting man, freeloader, frequent offender, frequent offenders, fugitive, gang, gang leader, gang member, gang members, gangs, gangster, gangster gang, head agent, highwayman, hijacker, hijackers, hit man, hitman, hooligan, hooligans, hostage, hostages, hostage taker, hostage takers, human trafficker, human traffickers, illegals, intruder, juvenile delinquent, juvenile gang, kidnapper, kidnappers, kidnappings, killer, killing machine, knife puller, lady killer, liquidation, looters, lover boy, lover girls, mafta, mafta boss, maftosi, maftoso, main perpetrator, main suspect, mass murderer, mass murderers, missing, motor agent, mugger, murderers, murder suspect, pedophile, pedophiles, perpetrator, perpetrators, petty thief, pick pocketers, pimp, pirates, plunderer, police commander, police detective, police general, policeman, policemen, police officer, police officers, police state, police station, policewoman, police woman, police wound, prisoner, prisoners, prisoners of war, prisons, professional criminal, punishable, purse snatcher, pyromaniac, rapist, recidivist, rioter, rioters, robberies, robbery, robbery homicide, security agent, security agents, security police, serial killer, sex murderer, sex offender, sharpshooter, shooter, shooting, shoplifter, skinheads, slave, slaves, smuggler, smugglers, sniper, snipers, someone who threatens, speed devil, speeder, stabbed, stabbing, stalker, stranger, strangulation, street gang, suicide, suicide bomber, superintendent, suspect, terror group, terrorist, terror suspect, thief, troublemaker, victims, violent man, warden, wardens, youth gangs.

Appendix B

Words Capturing Low-Status Stereotypes

Dutch: *achterlijk*, *achterlijke*, *achterstanden*, *achterstandskinderen*, *achterstandsleerling*, *achterstandsleerlingen*, *achterstandswijken*, *achterstandwijken*, *achterstelling*, *alcoholicus*, *alcoholist*, *alcoholiste*, *alcoholisten*, *analfabeet*, *analfabete*, *analfabeten*, *armoedig*, *barbaars*, *bastaardzoon*, *bedelaar*, *bedelaars*, *bijstandsgerechtigden*, *bijstandsgerechtigen*, *boerenlul*, *dakloze*, *daklozen*, *dronkelap*, *druggebruikers*, *drugsgebruiker*, *drugsgebruikers*, *drugsrunners*, *drugstoeristen*, *drugsverslaafde*, *drugsverslaafden*, *hangjongere*, *hangjongeren*, *hoer*, *hoerenlopers*, *hulpbehoevend*, *hulpbehoevende*, *idiot*, *junk*, *junkers*, *junks*, *kansarme*, *kansarmen*, *kindertehuizen*, *krottenwijk*, *laaggeschoolde*, *laaggeschoolden*, *laagopgeleide*, *laagopgeleiden*, *loser*, *malloot*, *minderwaardig*, *nestbevuiler*, *nietsnut*, *onderklasse*, *onderontwikkeld*, *ongeletterde*, *ongeschoolde*, *overlastgevende*, *pooier*, *primitief*, *probleemjongeren*, *prostituee*, *prostituees*, *prostitutiebedrijven*, *reljongeren*, *schoolverlaters*, *slet*, *sloeber*, *sloebers*, *spijbelaar*, *spijbelen*, *straatarm*, *straatkinderen*, *straatprostitutie*, *sukkel*, *taalachterstand*, *tienermoeder*, *tienermoeders*, *uitkeringgerechtigden*, *uitkeringsgerechtigden*, *uitwas*, *verschoppelingen*, *verslaafde*, *verslaafden*, *weeskinderen*, *werkloos*, *werkloze*, *werklozen*, *werkschuwe*, *zwerver*, *zwervers.*

English translation: addict, addicts, alcoholic, alcoholic, alcoholics, backlogs, backward, bad neighborhoods, barbaric, bastard son, beggar, beggars, beneficiaries, benefit recipients, children's homes, disadvantaged, disadvantaged children, disadvantaged neighborhoods, disadvantaged people, disadvantaged pupil, disadvantaged students, drug addict, drug addicts, drug runners, drug tourists, drug user, drug users, drunk, drunken, fool, half-timer, hick, hobo, hobos, homeless, homeless people, idiot, illiteracy, illiterate, illiterate people, inferior, invalid, junk, junkers, junks, language deficiency, loitering, loser, low-educated, low-skilled people, low-skilled person, low-educated, low-educated people, malolly, most miserable, needy, nuisance, orphans, outcasts, penniless, pimp, primitive, problem youth, prostitute, prostitutes, prostitution companies, rejuvenate, retarded, school drop-outs, school leavers, shabby, slob, slobs, slum, slut, softy, street children, street prostitution, street youth, street youths, subordination, teen mom, traitor, tramps, truancy, truant, underclass, underdeveloped, unemployed, unemployed people, unemployed person, unskilled, wanderer, welfare recipients, whore, whore hoppers, work shy, wretch.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Damian Trilling https://orcid.org/0000-0002-2586-0352

Notes

1. Analogies were tested relating to common capitals and countries (n = 306 combinations), such as "Bangkok is to Thailand as Beijing is to_____," family relations (n = 306 combinations), such as "Grandpa is to grandma as son is to_____," and comparisons (n = 812 combinations), such as "bad is to worse as is big to_____." The model should uniquely identify the target word for a correct match (e.g., China, daughter, and bigger).

2. This decision was partly informed by the strong correlations between both stereotype-dimensions for the selected outgroup categories, indicating that these groups score relatively high on both high-threat and low-status dimensions. These finding supports predictions put forward by the SCM (Fiske et al., 2002,).
3. As can be seen, between 2009 and 2013 a sharp increase in the strength of stereotypical associations of Moroccans was witnessed, afterward it dropped significantly in 2014. Presumably, this sharp decline can be explained by the response to the controversial speech of radical right-wing politician Geert Wilders: During the 2014 regional elections, Wilders asked during a TV-broadcasted speech if people wanted "more or less Moroccans" afterward the audience replied with "less, less, less." This provoked severe public outcry and led several media personalities and politicians to express their sympathy with the Moroccan community.

References

Arendt, F. (2010). Cultivation effects of a newspaper on reality estimates and explicit and implicit attitudes. *Journal of Media Psychology*, *22*(4), 147–159. https://doi.org/10.1027/1864-1105/a000020

Arendt, F., & Karadas, N. (2017). Content analysis of mediated associations: An automated text-analytic approach. *Communication Methods and Measures*, *11*(2), 105–120. https://doi.org/10.1080/19312458.2016.1276894

Arendt, F., & Northup, T. (2015). Effects of long-term exposure to news stereotypes on implicit and explicit attitudes. *International Journal of Communication*, *9*, 732–751.

Atwell Seate, A., & Mastro, D. (2017). Exposure to immigration in the news: The impact of group-level emotions on intergroup behavior. *Communication Research*, *44*(6), 817–840. https://doi.org/10.1177/0093650215570654

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Advances in neural information processing systems* (pp. 4349–4357).

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016b). *Quantifying and reducing stereotypes in word embeddings*. https://arxiv.org/abs/1606.06121

Boomgaarden, H. G., & Vliegenthart, R. (2007). Explaining the rise of anti-immigrant parties: The role of news media content. *Electoral Studies*, *26*(2), 404–417. https://doi.org/10.1016/j.electstud.2006.10.018

Bos, L., Lecheler, S., Mewafi, M., & Vliegenthart, R. (2016). It's the frame that matters: Immigrant integration and media framing effects in the Netherlands. *International Journal of Intercultural Relations*, *55*, 97–108. https://doi.org/10.1016/j.ijintrel.2016.10.002

Boukes, M., & Vliegenthart, R. (2020). A general pattern in the construction of economic newsworthiness? Analyzing news factors in popular, quality, regional, and financial newspapers. *Journalism: Theory, Practice & Criticism*, *21*, 279–300. https://doi.org/10.1177/1464884917725989

Brunet, M.-E., Alkalay-Houlihan, C., Anderson, A., & Zemel, R. (2019). Understanding the origins of bias in word embeddings. In Proceedings of the 36th International Conference on Machine Learning. http://arxiv.org/abs/1810.03611

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora necessarily contain human biases. *Science*, *6334*(356), 183–186.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, *40*, 61–149. https://doi.org/10.1016/S0065-2601(07)00002-0

Cuddy, A. J. C., Fiske, S. T., Kwan, V. S. Y., Glick, P., Demoulin, S., Leyens, J.-P., & Ziegler, R. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *The British Journal of Social Psychology*, *48*, 1–33. https://doi.org/10.1348/014466608X314935

Dixon, T. L., & Linz, D. (2000). Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television news. *Journal of Communication*, *50*(2), 131–154.

Dixon, T. L., & Williams, C. L. (2015). The changing misrepresentation of race and crime on network and cable news. *Journal of Communication*, *65*(1), 24–39. https://doi.org/10.1111/jcom.12133

Duyvendak, J. W., & Scholten, P. (2012). Deconstructing the Dutch multicultural model: A frame perspective on Dutch immigrant integration policymaking. *Comparative European Politics*, *10*(3), 266–282. https://doi.org/10.1057/cep.2012.9

Eberl, J., Meltzer, C. E., Heidenreich, T., Theorin, N., Lind, F., Berganza, R., & Schemer, C. (2018). The European media discourse on immigration and its effects: A literature review. *Annals of the International Communication Association*, *842*(3), 207–223. https://doi.org/10.1080/23808985.2018.1497452

Erisen, C., & Kentmen-Cin, C. (2017). Tolerance and perceived threat toward Muslim immigrants in Germany and the Netherlands. *European Union Politics*, *18*(1), 73–97. https://doi.org/10.1177/1465116516675979

Firth, J. R. (1957). *Papers in linguistics, 1934-1951*. Oxford University Press.

Fiske, S. T. (2012). Warmth and competence: Stereotype content issues for clinicians and researchers. *Canadian Psychology*, *53*(1), 14–20. https://doi.org/10.1037/a0026054

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037//0022-3514.82.6.878

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E3635–E3644.

Goldberg, Y. (2017). *Neural network methods for natural language processing*. Morgan & Claypool.

Gorodzeisky, A., & Semyonov, M. (2009). Terms of exclusion: Public views towards admission and allocation of rights to immigrants in European countries. *Ethnic and Racial Studies*, *32*(3), 401–423. https://doi.org/10.1080/01419870802245851

Gorodzeisky, A., & Semyonov, M. (2016). Not only competitive threat but also racial prejudice: Sources of anti-immigrant attitudes in European societies. *International Journal of Public Opinion Research*, *28*(3), 331–354. https://doi.org/10.1093/ijpor/edv024

Greenwald, A. G. (2017). An AI stereotype catcher. *Science*, *356*(6334), 133–134. https://doi.org/10.1126/science.aan0649

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts., *Political Analysis*, *21*, 267–297. https://doi.org/10.1093/pan/mps028

Guo, L., & Vargo, C. (2015). The power of message networks: A big-data analysis of the network agenda setting model and issue ownership. *Mass Communication and Society*, *18*(5), 557–576. https://doi.org/10.1080/15205436.2015.1045300

Jacobs, L., Damstra, A., Boukes, M., & De Swert, K. (2018). Back to reality: The complex relationship between patterns in immigration news coverage and real-world developments in Dutch and Flemish newspapers (1999–2015). *Mass Communication and Society*, *21*(4), 473–497. https://doi.org/10.1080/15205436.2018.1442479

Kittel, B. (1999). Sense and sensitivity in pooled analysis of political data. *European Journal of Political Research*, *35*, 225–253. https://doi.org/10.1111/1475-6765.00448

Kroon, A. C., Kluknavská, A., Vliegenthart, R., & Boomgaarden, H. G. (2016). Victims or perpetrators? Explaining media framing of Roma across Europe. *European Journal of Communication*, *31*(4), 375–392. https://doi.org/10.1177/0267323116647235

Kroon, A. C., Trilling, D., Van Selm, M., & Vliegenthart, R. (2018). Biased media? How news content influences age discrimination claims. European Journal of Ageing, *16*, 109–119. https://doi.org/10.1007/s10433-018-0465-4

Lee, T. L., & Fiske, S. T. (2006). Not an outgroup, not yet an ingroup: Immigrants in the stereotype content model. *International Journal of Intercultural Relations*, *30*(6), 751–768. https://doi.org/10.1016/j.ijintrel.2006.06.005

Leschke, J. C., & Schwemmer, C. (2019). Media bias towards African-Americans before and after the Charlottesville Rally. In *Proceedings of the Weizenbaum Conference*, *2019*, *"Challenges of Digital Inequality—Digital Education*, *Digital Work*, *Digital Life"* (pp. 1–10). https://doi.org/10.34669/wi.cp/2.25

Mastro, D. (2009). Effects of racial and ethnic stereotyping. In J. Bryant & M. B. Oliver (Eds.), *Media effects: Advances in theory and research* (pp. 325–341). Routledge.

Matthes, J., & Schmuck, D. (2017). The effects of anti-immigrant right-wing populist ads on implicit and explicit attitudes: A moderated mediation model. *Communication Research*, *44*(4), 556–581. https://doi.org/10.1177/0093650215577859

Mikolov, T., Corrado, G., Chen, K., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *ICLR:,Proceeding of the International Conference on Learning Representations Workshop Track* (pp. 1301–3781).

Mikolov, T., Yih, W., & Zweig, G. (2013, June). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (pp. 746–751). Association for Computational Linguistics.

Musterd, S. (2005). Social and ethnic segregation in Europe: Levels, causes, and effects. *Journal of Urban Affairs*, *27*(3), 331–348. https://doi.org/10.1111/j.0735-2166.2005.00239.x

Mutz, D. C., & Goldman, S. K. (2016). Mass media. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The SAGE handbook of prejudice, stereotyping and discrimination* (pp. 241–258). SAGE.

Papacharissi, Z., & de Fatima Oliveira, M. (2008). News frames terrorism: A comparative analysis of frames employed in terrorism coverage in U.S. and U.K. newspapers. *International Journal of Press/Politics*, *13*(1), 52–74. https://doi.org/10.1177/1940161207312676

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics.

Reese, S. D. (2001). Understanding the global journalist: A hierarchy-of-influences approach. *Journalism Studies*, *2*(2), 173–187. https://doi.org/10.1080/14616700120042060

Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50)., University of Malta. http://www.muni.cz/research/publications/884893

Roggeband, C., & Vliegenthart, R. (2007). Divergent framing: The public debate on migration in the Dutch parliament and media, 1995-2004. *West European Politics*, *30*(3), 524–548. https://doi.org/10.1080/01402380701276352

Rudkowsky, E., Haselmayer, M., Wastian, M., Jenny, M., Emrich, Š., Sedlmair, M., & Sedlmair, M. (2018). More than bags of words: Sentiment analysis with word embeddings. *Communication Methods and Measures*, *12*(2–3), 140–157. https://doi.org/10.1080/19312458.2018.1455817

Ruigrok, N., & van Atteveldt, W. (2007). Global angling with a local angle: How U.S., British, and Dutch newspapers frame global and local terrorist attacks. *Harvard International Journal of Press/Politics*, *12*(1), 68–90. https://doi.org/10.1177/1081180X06297436

Schemer, C. (2012). The influence of news media on stereotypic attitudes toward immigrants in a political campaign. *Journal of Communication*, *62*(5), 739–757. https://doi.org/10.1111/j.1460-2466.2012.01672.x

Schieferdecker, D., & Wessler, H. (2017). Bridging segregation via media exposure? Ingroup identification, outgroup distance, and low direct contact reduce outgroup appearance in media repertoires. *Journal of Communication*, *67*, 993–1014. https://doi.org/10.1111/jcom.12338

Schnabel, T., Labutov, I., & Mimno, D. (2015). Evaluation methods for unsupervised word embeddings. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, (pp. 298–307). Association for Computational Linguistics.

Semyonov, M., & Glikman, A. (2008). Ethnic residential segregation, social contacts, and anti-minority attitudes in European societies. *European Sociological Review*, *25*(6), 693–708. https://doi.org/10.1093/esr/jcn075

Sink, A., Mastro, D., & Dragojevic, M. (2018). Competent or warm? A stereotype content model approach to understanding perceptions of masculine and effeminate gay television characters. *Journalism & Mass Communication Quarterly*, *95*(3), 588–606. https://doi.org/10.1177/1077699017706483

Thompson, A. (2017). Google's sentiment analyzer thinks being gay is bad. *Vice*. https://www.vice.com/en_us/article/j5jmj8/google-artificial-intelligence-bias.

Tukachinsky, R., Mastro, D., & Yarchi, M. (2015). Documenting portrayals of race/ethnicity on primetime television over a 20-year span and their association with national-level racial/ethnic attitudes. *Journal of Social Issues*, *71*(1), 17–38. https://doi.org/10.1111/josi.12094

Van Dijk, T. A. (2000). New(s) racism: A discourse analytical approach. In S. Cottle (Ed.), *Ethnic minorities and the media* (pp. 33–49). Open University Press.

van Heerden, S., de Lange, S. L., van der Brug, W., & Fennema, M. (2014). The immigration and integration debate in the Netherlands: Discursive and programmatic reactions to the rise of anti-immigration parties. *Journal of Ethnic and Migration Studies*, *40*(1), 119–136. https://doi.org/10.1080/1369183X.2013.830881

Vliegenthart, R., & Boomgaarden, H. G. (2007). Real-world indicators and the coverage of immigration and the integration of minorities in Dutch newspapers. *European Journal of Communication*, *22*(3), 293–314. https://doi.org/10.1177/0267323107079676

Vliegenthart, R., & Roggeband, C. (2007). Framing immigration and integration: Relationships between press and parliament in the Netherlands. *International Communication Gazette*, *69*(3), 295–319. https://doi.org/10.1177/1748048507076582

Wilson, S. E., & Butler, D. M. (2007). A lot more to do: The sensitivity of time-series cross-section analyses to simple alternative specifications. *Political Analysis*, *15*(2), 101–123. https://doi.org/10.1093/pan/mpl012

Author Biographies

Anne C. Kroon is an assistant professor in Corporate Communication at the Amsterdam School of Communication Research (ASCoR), University of Amsterdam. Drawing on computational methods, her research documents representations of minorities and examines the consequences of exposure to such portrayals for interpersonal and labor market outcomes.

Damian Trilling is an associate professor for Political Communication and Journalism and co-director of the Communication in the Digital Society Initiative of the Department of Communication Science at the University of Amsterdam, where he is also affiliated with the Amsterdam School of Communication Research (ASCoR). He is interested in the impact of the ever-changing media environment on news and journalism as well as the use and development of computational research methods.

Tamara Raats is a Communication Science graduate at the University of Amsterdam, where she assisted and carried out several studies and research projects. Implementing computational methods, these projects concentrated on media portrayals of minorities, framing of economic organizations and agenda-setting in political documents.