Whose ideas are worth spreading? The representation of women and ethnic groups in TED talks - Supporting Information

Replication material for this article is available at Harvard Dataverse [https://doi.org/10.7910/DVN/EUDWP3](https://doi.org/10.7910/DVN/EUDWP3).

S1  Image recognition algorithm

This section contains metrics on the validation of the image recognition algorithm that we used to capture sex and ethnicity of TED speakers. Numbers are based on a comparison of human and algorithmic coding of 200 random samples speaker images.

Validation

The following tables show the pairwise inter-rater agreements between human coders and the algorithm.

| Rater 1            | Rater 2            | Fleiss’ Kappa |
|--------------------|--------------------|---------------|
| Human coder 1      | Algorithm          | 0.833         |
| Human coder 2      | Algorithm          | 0.888         |
| Human coder 1      | Human coder 2      | 0.825         |

The average inter-rater agreement for ethnicity between human coders and the image recognition algorithm is 0.860.

| Rater 1          | Rater 2          | Fleiss’ Kappa |
|------------------|------------------|---------------|
| Human 1          | Algorithm        | 0.946         |
| Human 2          | Algorithm        | 0.968         |
| Human 1          | Human 2          | 0.957         |

The average inter-rater agreement for sex between human coders and the image recognition algorithm is 0.957.
The following tables show overall inter-rater agreements between both human coders and the algorithm by ethnicity and sex.

| Label    | Fleiss’ Kappa |
|----------|---------------|
| Asian    | 0.932         |
| Black    | 0.965         |
| Hispanic | 0.576         |
| Other    | 0.707         |
| White    | 0.864         |
| Overall  | 0.848         |

| Label | Fleiss’ Kappa |
|-------|---------------|
| Female| 0.957         |
| Male  | 0.957         |
| Overall| 0.957      |

In total, Kappa values for both ethnicity as well as sex are good, with high agreement for the majority of sample images about both attributes. The agreement for the ethnicity categories Hispanic and Other are worse in comparison to the remaining categories. It is worth highlighting at this point that the lower agreement can not only be ascribed to poor performance by the algorithm. Both, pairwise comparisons between human coders as well as between humans and the algorithm showed a small number of disagreements for these categories. For instance, some speaker images were labeled as White by one rater and Hispanic by another rater. Overall, our validation results show that annotations from the image recognition algorithm are not perfect, but mostly in line with annotations by human coders and sufficient for our research task.
Another approach for validating the algorithmic performance is to treat results from one human coder as a gold standard and assess the predictive performance of the image recognition algorithm. In doing so, we calculated precision, recall and F1 scores for each ethnicity class, as can be seen in the following table:

| Ethnicity Class | Precision | Recall | F1 Score |
|----------------|-----------|--------|----------|
| Asian          | 1.00      | 1.00   | 1.00     |
| Black          | 0.95      | 0.95   | 0.95     |
| Hispanic       | 0.54      | 0.46   | 0.50     |
| Other          | 0.50      | 0.75   | 0.60     |
| White          | 0.97      | 0.97   | 0.97     |

As for computing these metrics over all classes, several metrics could be used. The overall micro averaged F1 score, which does not take imbalanced class distributions into account, is at 0.96. In comparison, the macro averaged F1 score, which takes into account the sizes of all classes, is at 0.81.

For gender, precision is at 0.95, recall at 0.97 and the F1 score at 0.96.

Overall, these results are in line with our inter-rater metrics from above.

To further compare the performance of the image recognition algorithm with name-based approaches for identifying gender and ethnicity, we again treated human codings as a gold standard for ethnicity and gender. We then used the first names (for gender) and surnames (for ethnicity) to identify the most likely class for each category (see Wais [2016], Imai and Khanna [2016]). Afterwards, we again computed precision, recall and F1 scores:

| Ethnicity Class | Precision | Recall | F1 Score |
|----------------|-----------|--------|----------|
| Asian          | 0.81      | 0.47   | 0.60     |
| Black          | 0.09      | 0.50   | 0.15     |
| Hispanic       | 0.27      | 0.23   | 0.25     |
| Other          | 0.00      | 0.00   | 0.00     |
| White          | 0.90      | 0.83   | 0.87     |

The micro averaged F1 scores for the name-based predictions is at 0.86 and the macro averaged F1 score at 0.41. These results suggest that the image recognition algorithm performs substantially better for predicting ethnicity categories.

For the name-based gender predictions, precision, recall and F1 scores are at 0.94. These scores are also lower than those achieved by the image recognition algorithm, but only by a very small margin.
**Probability distributions**

The following figure shows the probability distributions for ethnicity annotations by the image recognition algorithm. Most of the probability mass is centered around 0% and 100%, which is why a conversion of the continuous measures to a binary non-white versus white indicator does not result in a major loss of information.

![Probability distributions of image recognition algorithm by ethnicity](image)

Figure S1: Probability distributions of image recognition algorithm by ethnicity
S2 Topic modeling

Topic model diagnostics

To find a model that best fits our research purpose, we computed three different models with 20, 30, and 50 topics. For each model, we then calculated semantic coherence and exclusivity, which are measures that quantify necessary statistical properties and are recommended by the authors of the structural topic model Roberts et al. 2014. Semantic coherence is higher when more probable words in a topic frequently co-occur together Mimno et al. 2011. Exclusivity is based on the FREX metric and achieves higher values when more words are exclusive to corresponding topics Lucas et al. 2015. The following figure illustrates average and median values for semantic coherence and exclusivity of all structural topic models that we fitted on the TED talk transcripts.

![Figure S2: Probability distributions of image recognition algorithm by ethnicity](image)

While the figure shows that no model is clearly superior, the model with 30 topics outperforms both other models in terms of exclusivity. In addition to statistical diagnostics, we inspected the topic models by analyzing the most frequent and exclusive words (Lucas et al. 2015, p. 19) - called frex terms - and representative talks with the highest proportions for each topic. Based upon frex terms and highly representative texts we at last assigned labels to each topic. During this evaluation procedure, the model with 30 topics also turned out to be the best model in relation to our research task and was therefore chosen as our final model.
## Topic Labels and Proportions

| Label                              | Proportion | FREX terms                                                                 |
|------------------------------------|------------|-----------------------------------------------------------------------------|
| family                             | 8%         | father, mother, family, felt, mom, fear, love, told, knew, dream, met, son, home, sister, night, friend, dad, brother, die, cri |
| work & misc.                       | 5%         | guy, stuff, phone, somebody, hey, everybody, pick, bear, laugh, email, anyway, morn, oh, sort, cartoon, room, shoe, fun, funni, check |
| computers & technology             | 5%         | comput, machin, design, devic, technolog, softwar, interact, 3d, algorithm, digit, print, interfac, prototyp, manufactur, code, program, sensor, mit, screen, tool |
| countries & poverty                | 4%         | china, africa, india, african, chines, countri, aid, incom, growth, poverti, econom, global, wealth, economi, gdp, inequ, europ, poor, asia, west |
| war & terror                       | 4%         | refuge, muslim, war, afghanistan, militari, peac, soldier, islam, conflict, weapon, arab, bomb, iraq, violenc, iran, terror, kill, gun, attack, secur |
| law & politics                     | 4%         | prison, vote, elect, crimin, polec, legal, citizen, law, democraci, poli, crime, justic, court, jail, govern, democrati, presid, corrupt, congress, lawyer |
| food & lodging                     | 4%         | poet, crit, art, fic, actor, theatri, stage, music, plays, oper, soprano, viol, music, violin, violi, violinc, violoncello, violinista, violoncelli, violoncelli, violoncello, violinista, violoncelli, violoncello, violinista, violoncelli, violoncello, violinista, violoncelli, violoncello |
| money & business                   | 4%         | market, busi, money, dollar, company, financ, invest, sector, profit, entrepreneur, buy, fund, bank, brand, product, employe, capti, custom, pay |
| general science                    | 4%         | choic, science, predict, deci, math, statis, statistic, scientific, wrong, knowledge, solv, answer, model, puzzl, evid, intellig, scientif, theori, bias, mathemat, chaos |
| emotions & psychology              | 4%         | compas, emot, self, happi, moral, psycholog, stress, empathi, relationship, autism, mental, sleep, desir, social, feel, suffer, smile, behavior, love, other |
| games & music                      | 4%         | internet, onlin, twitter, web, data, media, network, googl, facebook, inform, privaci, digit, blog, post, content, website, link, user, phone, site |
| medical                            | 4%         | cancer, patient, medic, clinic, treatment, doctor, tumor, vaccin, disease, surgeri, medicin, drug, breast, health, physician, hospi, hiv, symptom, trial, epidem |
| architecture & design              | 3%         | music, game, song, play, musician, sound, piano, player, sing, orchestra, listen, video, voic, instrument, opera, hear, nois, ear, improvise, concert |
| inequality                         | 3%         | women, men, gender, gay, sexual, sex, girl, woman, rape, feminisit, femal, black, marriag, male, boy, slaveri, marri, equal, abus, violenc |
| universe & physics                 | 3%         | galaxi, partil, telescop, star, quantum, atom, univers, mathemat, hole, theori, dark, dimens, light, graviti, physicist, symmetri, physic, astronom, magnet, sun |
| astronomy                          | 3%         | mar, ice, earth, satellit, atmospher, asteroid, planet, cave, pole, moon, nasa, orbit, mountain, explor, glacier, rock, solar, nission, antarctica, surfac |
| genetics                           | 3%         | dna, gene, genom, cell, tissue, genet, molecul, stem, bacteria, protein, biolog, molecular, sequence, chromosom, mutat, virus, sil, evolut, transplant, bone |
| architecture & design              | 3%         | citi, architectur, urban, build, neighborhood, architect, park, street, mayor, built, hous, york, design, communiti, space, roof, site, town, tower, rio |
| environment                        | 3%         | oil, climat, energi, nuclear, carbon, fuel, emiss, coal, elect, gas, co2, solar, renew, wind, batteri, burn, fusion, effici, fossil, heat |
| language & writing                 | 3%         | languag, book, english, translat, write, text, word, dictionari, letter, spell, sentenc, read, librar, script, page, written, metaphor, writer, publish, editor |
| religious & mystery                | 3%         | god, film, religion, movi, religi, fiction, storytell, character, conscious, realiti, argument, stori, faith, magic, mysteri, truth, christian, tomm, comic, believ |
| children & school                  | 3%         | school, teacher, educ, kid, student, children, classroom, teach, grade, colleg, class, parent, child, graduat, skill, adult, young, villag, lunch, childhood |
| animals & nature                   | 3%         | forest, chimpanze, tree, extinct, speci, eleph, creatur, bat, anim, ancestor, bird, amazon, rainforest, mammal, bear, monkey, soil, beetl, frog, ecosystem |
| body & sports                      | 3%         | leg, arm, finger, limb, disable, breath, knee, foot, climb, danc, bodi, blind, athlet, swim, feet, wheelchair, ball, walk, injuri, jump |
| stopwords                          | 2%         | ca, ok, yeah, la, chris, ted, poem, flag, oh, mr, ah, card, anderson, yes, smell, prize, pleas, audienc, sir, prime |
| materials & sustainabil-ity        | 2%         | plastic, water, sand, wast, oxygen, recyc, materi, bottl, mushroom, toilet, drink, air, wash, clean, chemic, pollut, sanit, river, temperatur, pump |
| food & farming                     | 2%         | bee, food, farmer, crop, mosquito, farm, eat, diet, bread, flower, seed, agriculture, meat, plant, feed, pig, meal, grain, chicken, cow |
| neuroscience & robotics            | 2%         | neuron, brain, robot, ant, memor, cortex, neural, neurosci, activ, signal, pattern, region, task, motor, sensori, coloni, babi, function, rat, electro |
| transportation                     | 2%         | car, vehicl, airplan, driver, crash, traffic, road, flight, drive, gos, fli, aircraft, balloon, seat, highway, mile, wheel, wing, pilot, plane |
| ocean & sea                        | 2%         | shark, fish, ocean, coral, whale, reef, sea, dolphin, boat, underwater, marin, ship, dive, island, shrimp, tag, pacific, swim, coast, jellyfish |
Prevalence effects

The following figures includes prevalence effects from the structural topic model. Effects are illustrated for speaker ethnicity (A), speaker gender (B) and publication date of talks (C) with 95% confidence intervals.

Figure S3: Prevalence effects of speaker ethnicity (A), speaker gender (B) and publication date of a talk (C) on inequality topic with 95% confidence intervals.
S3 Regressions

Coefficient table

The following table includes unstandardized regression coefficients for the comment sentiment and the number of dislikes of YouTube videos. The *work and misc.* topic was excluded from both models to avoid collinearity issues.

Table S8: Regression models for comment sentiment and number of dislikes

| Comment sentiment | Video dislikes |
|-------------------|----------------|
| Intercept         | 3.26 (0.94)*** | −338.79 (13.82)*** |
| Video date        | −0.00 (0.00)*** | 0.17 (0.01)*** |
| Video views       | 0.00 (0.00)     | 0.00 (0.00)*** |
| Gender: female    | −0.01 (0.00)**  | 0.28 (0.04)*** |
| Ethnicity: non-white | 0.02 (0.00)*** | −0.17 (0.05)*** |
| Topic: games and music | 0.03 (0.02) | −0.02 (0.31) |
| Topic: stopwords  | −0.02 (0.02)    | −0.09 (0.36) |
| Topic: inequality | −0.19 (0.02)*** | 3.53 (0.34)*** |
| Topic: countries and poverty | −0.07 (0.02)*** | 0.39 (0.28) |
| Topic: universe and physics | −0.03 (0.02) | −0.18 (0.28) |
| Topic: family     | 0.00 (0.02)     | −0.36 (0.29) |
| Topic: war and terror | −0.16 (0.02)*** | 0.37 (0.28) |
| Topic: astronauts | −0.06 (0.02)**  | −1.05 (0.30)*** |
| Topic: genetics   | −0.04 (0.02)    | −0.91 (0.30)*** |
| Topic: architecture and design | −0.01 (0.02) | −1.16 (0.30)*** |
| Topic: environment | −0.09 (0.02)*** | 0.42 (0.30) |
| Topic: materials and sustainability | −0.02 (0.02) | −1.33 (0.34)*** |
| Topic: law and politics | −0.17 (0.02)*** | 1.01 (0.30)*** |
| Topic: art        | 0.05 (0.02)*    | −0.71 (0.31)* |
| Topic: food and farming | −0.05 (0.02)* | −0.09 (0.33) |
| Topic: language and writing | 0.01 (0.02) | −0.30 (0.35) |
| Topic: religion and mystery | −0.07 (0.02)** | 2.39 (0.36)*** |
| Topic: money and business | −0.03 (0.02) | −0.87 (0.31)** |
| Topic: neuroscience and robotics | −0.04 (0.02) | −1.31 (0.31)*** |
| Topic: children and school | 0.03 (0.02) | −1.07 (0.32)*** |
| Topic: computers and technology | −0.00 (0.02) | −0.42 (0.29) |
| Topic: general science | −0.02 (0.02) | −0.15 (0.31) |
| Topic: emotions and psychology | −0.01 (0.02) | 0.19 (0.29) |
| Topic: data and internet | −0.05 (0.02)* | −0.39 (0.30) |
| Topic: transportation | −0.09 (0.02)*** | −0.53 (0.35) |
| Topic: animals and nature | −0.02 (0.02) | −1.43 (0.31)*** |
| Topic: body and sports | −0.01 (0.02) | −0.62 (0.32) |
| Topic: medical     | −0.09 (0.02)*** | −1.50 (0.28)*** |
| Topic: ocean and sea | −0.03 (0.02) | −2.05 (0.31)*** |

Log Likelihood 2913.20 −13709.98
Observations 2310 2323

***p < 0.001, **p < 0.01, *p < 0.05; standard errors in parentheses.
**Interaction between date and ethnicity**

To further examine the relation between time and ethnicity regarding the YouTube sentiment, we computed an additional regression which includes an interaction between date and a dummy variable for white versus non-white speakers. The following visualization shows the corresponding effects while holding all other covariates at their observed values.

![Figure S4: Predicted sentiment score for ethnicity and date.](image)

It becomes apparent that, similar to our model without an interaction term, TED talks by non-white speakers receive a more positive sentiment on YouTube. However, over time, differences between non-white and white speakers vanish.
Dislikes on YouTube

The following figure includes regression estimates for the number of dislikes for each TED talk in our sample. Subfigure A shows standardized coefficients of generalized linear models with 95% confidence intervals. Subfigures B.1-B.3 show predicted values for descriptive and substantive representation covariates.

Figure S5: Regression estimates for the number of dislikes of TED YouTube videos. Sub-figure A: standardized coefficients of generalized linear model with 95% confidence intervals. Sub-figures B.1 - B.3: predicted sentiment values for descriptive and substantive representation covariates.
References

Imai, Kosuke and Kabir Khanna (2016). “Improving ecological inference by predicting individual ethnicity from voter registration records”. In: *Political Analysis* 24.2, pp. 263–272. DOI: 10.1093/pan/mpw001

Lucas, Christopher, Richard A Nielsen, Margaret E Roberts, Brandon M Stewart, Alex Storer, and Dustin Tingley (2015). “Computer-Assisted Text Analysis for Comparative Politics”. In: *Political Analysis* 23.2, pp. 254–277. ISSN: 1047-1987. DOI: 10.1093/pan/mpu019 URL: http://pan.oxfordjournals.org/cgi/doi/10.1093/pan/mpu019%5Cnpapers3://publication/doi/10.1093/pan/mpu019

Mimno, David, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum (2011). “Optimizing semantic coherence in topic models”. In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. 2, pp. 262–272. ISBN: 9781937284114.

Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-luis, Shana Kushner Gadarian, Bethany Albertson, and David G Rand (2014). “Structural Topic Models for Open-Ended Survey Responses”. In: *American Journal of Political Science* 58.4, pp. 1064–1082. DOI: 10.1111/ajps.12103

Wais, Kamil (2016). “Gender Prediction Methods Based on First Names with genderizeR”. In: *The R Journal* 8.1, pp. 17–37. DOI: 10.32614/RJ-2016-002 URL: https://doi.org/10.32614/RJ-2016-002