Sentiment and hate-speech on social media

An analysis of the relevance of gender and race for the emergence of new communication structures within the YouTube Community

Claudia Buder, Nina-Sophie Fritsch,
Marie-Theres Hesse, Chiara Osorio Krauter,
Aaron Philipp, Roland Verwiebe,
Sarah Weißmann

University of Potsdam

General Online Research Conference 2023

Background

Vital research on sentiment and hate-speech, especially with a focus on X (Twitter), but also on YouTube (YT), FB, Grindr, Instagram, AirBnB etc. (Barbieri et al. 2021, Ghaffari 2022, Kumar & Jaiswal 2020, Latorre & Amores 2021, Ribeiro et al. 2020, Siegel et al. 2021)

- Studies repeatedly show that (negative) sentiment and hate-speech is a key feature of digital societies in the 21st century (strong indication that has to do with the algorithmic structures of major platforms)
- A larger part of theses studies analyze(d) qualitative data.
- Quantitative studies, which use digital traces (via web scraping) and systematically focus on socio-structural differences of the risks of being confronted with sentiment and hate-speech are still quite rare (based on class position, education, gender, age, ethnicity, race etc.)

State of the Art & Hypothesis

- However, existing studies show relatively **stable patterns across platforms** and independent of national context which indicate higher negative sentiment + higher hate-speech risks for women, PoC (and/or PwithMB) ➔ focus of our GOR paper
  - Gender has significant influence on viewer sentiment expressed in (YT)comments with women receiving slightly less positive feedback than men (Amarasekara & Grant 2019, Veletsianos et al. 2018)
  - Research dealing with race in social media suggests that the experience of racial-micro-aggressions are increasing (Sue et al. 2007, Tynes et al. 2018)

- This is the starting point for our empirical analysis & leads to 2 simple hypothesis:
- **H1:** Female YouTuber in Germany receive less positive comments from their audiences (and are confronted with higher level of hate-speech) than male YouTuber.
- **H2:** PoC hosts in Germany receive more negative comments (and are confronted with higher level of hate-speech) than white hosts.

Research Question

Which content patterns of **sentiment** and **hate-speech** can be identified within the communication structure of the YouTube Community?

Background: digital sphere comes with very own properties like anonymity, distance, easy accessibility and asymmetric communication which change the opion expressions and promote discrimination practices (Keum & Miller 2017)

What are the differences in **social structure** (e.g. age, gender, migration background) and **platform characteristics** (e.g., topic, amount of subscribers, amount of videos)?

Background: ‚At-risk' varies with different characteristics (Döring & Mohseni 2019, Thomas et al. 2022); negative Sentiment can cause mental health issues like stress, depression or anxiety; discrimination expected higher than in the offline world

Data & Design of our study

- Basis: full coverage of 115,976 channels in D-A-CH countries (Dec. 2022; min. 10 videos); including YT-specific variables (channel age, subscribers, number of views, video count); today results from a random sample N=4,000 YT-channels
- 07-08/2023: Web scraping of all comments under each video of the random sample (with the YouTube Data API v3)
- **Final sample**: N = 2,305 channels; 382,553 videos; over 29 Mio. comments (Language detection excluded all videos with a non-German community + cleaning the dataset)
- We used a hand-coded classification survey to classify Content Creators' socio-demographic characteristics (sex, age, race, education) and platform related variables of the channel (channel topic/YouTube branch, visibility, economic orientation of the channel etc.)

Variables coded in the Classification Survey

Sociodemographic Variables	N
Sex	
Missing	293
Female	338
Male	1674
Race	
Missing	968
Non white	122
White	1163
Unsure	52
Age	
Missing	663
$\leq$ 20 years	363
21-30 years	492
31-40 years	368
40+ years	419
Education	
Missing	1965
Low	120
High	220
Visibility	
No	1605
Yes	700

Platform Variables	N
Subscriber	
Mean	14,107
Median	429
Channel Age [in years]	
Mean	8.93
Median	8.47
Channel Topic	
Missing	4
Business	24
Conspiracy Theory	62
Culture	293
DIY	189
Education	54
Entertainment	487
Food	36
Gaming	817
Health	52
Lifestyle	69
Society	8
Sport	69
Tourism	94
Other	47
Observations	2,305

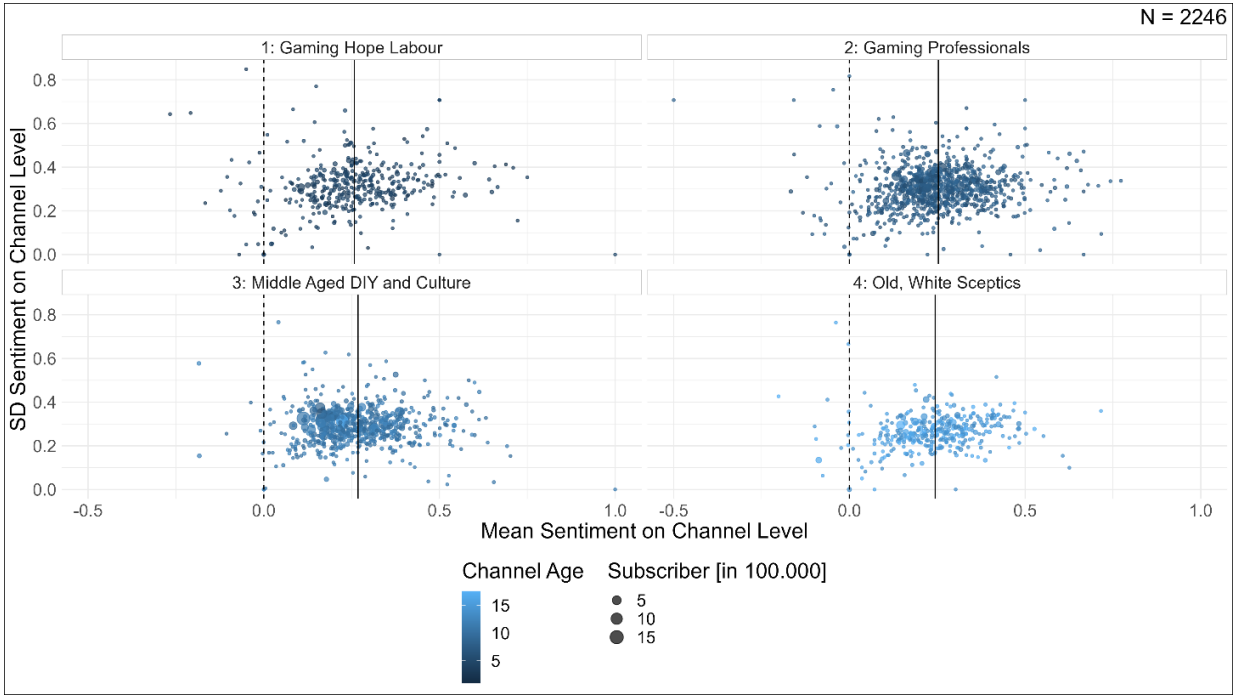
Sentiment Analysis

- Analyses of texts with regard to their emotional message (pos/neu/neg)

1. Dictionary-approach
   - Using SentiWS, German Polarity Cues, AFINN & Emoji Sentiment Ranking v1.0 and calculate the sentiment
2. Machine-learning-approach
   - Using Lasso Regression to label comments (trained with Detox dataset)

- Resulting Variables:
  - Mean Sentiment on channel level to measure the overall sentiment (-1 = only negative sentiment; +1 only positive sentiment)
  - SD Sentiment on channel level (mean of standard deviation from video level) to measure the polarisation (0 = no polarisation of the channel; the higher the value, the more polarising the channel)

Sample description based on main characteristics (Cluster analysis)



Input variables:

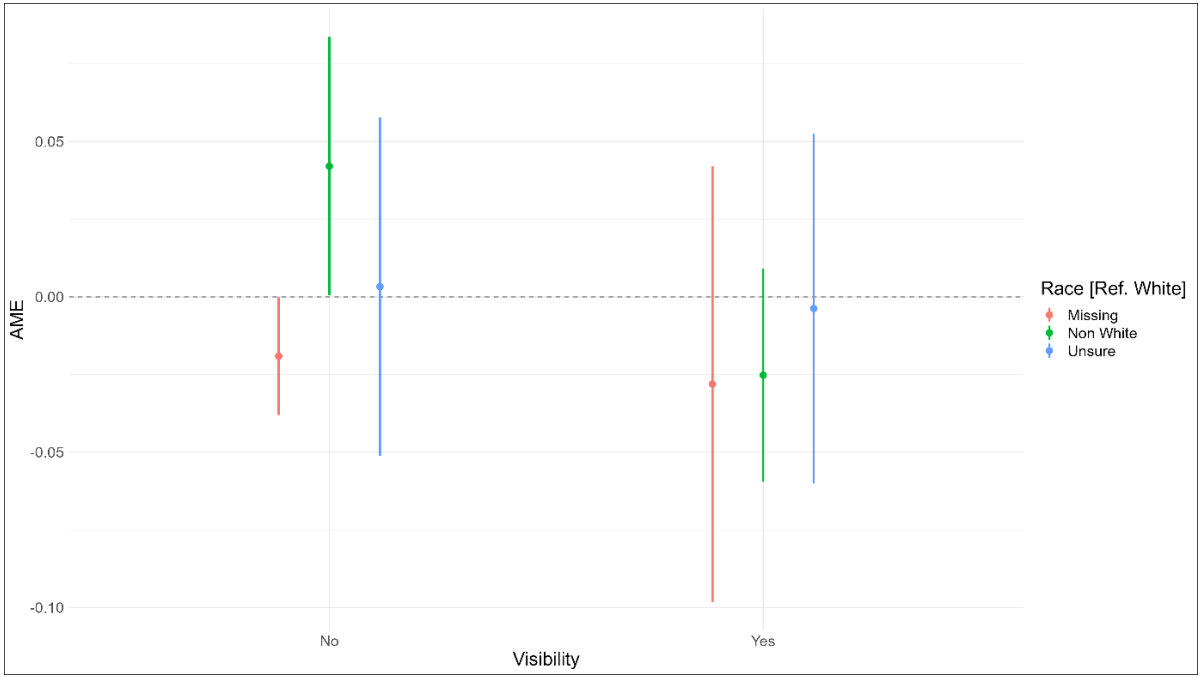
- **Age**
- **Education**
- Race
- Sex

- **Channel Age**
- **Channel Topic**
- Subscriber count
- mean sentiment
- polarisation

Channel Sentiment depending on Gender and Race

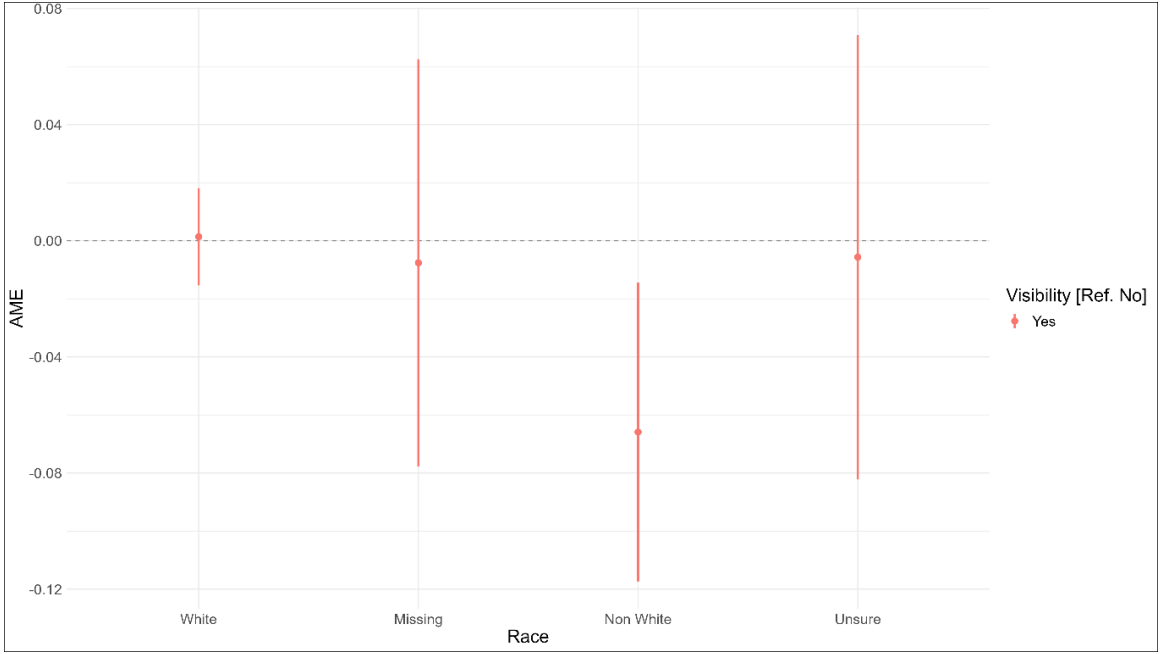
	Dependent Var.: Mean Sentiment on Channel Level
Sex [Ref: Male]	
Missing	−0.019* (0.011)
Female	**0.080*** (0.009)**
Race [Ref.: White]	
Missing	−0.019** (0.010)
Non White	**0.042** (0.021)**
Unsure	0.003 (0.028)
Visibility [Ref.: No]	
Yes	0.001 (0.009)
Interaction Race x Visibility	
Missing x Yes	−0.009 (0.037)
Non White x Yes	**−0.067** (0.027)**
Unsure x Yes	−0.007 (0.040)
Subscriber [in 100.000]	−0.012*** (0.004)
Channelage [in years]	−0.003*** (0.001)
Constant	0.241*** (0.031)
Observations	2,305
R^2	0.140
Adjusted R^2	0.129
Residual Std. Error	0.139 (df = 2273)
F Statistic	11.966*** (df = 31; 2273)
Note: controlled for age, channel topic, education	*p<0.1; **p<0.05; ***p<0.01

Sentiment differences depending on Race



controlled for age, channel age, channel topic, education, sex, subscriber

Sentiment differences depending on Race



controlled for age, channel age, channel topic, education, sex, subscriber

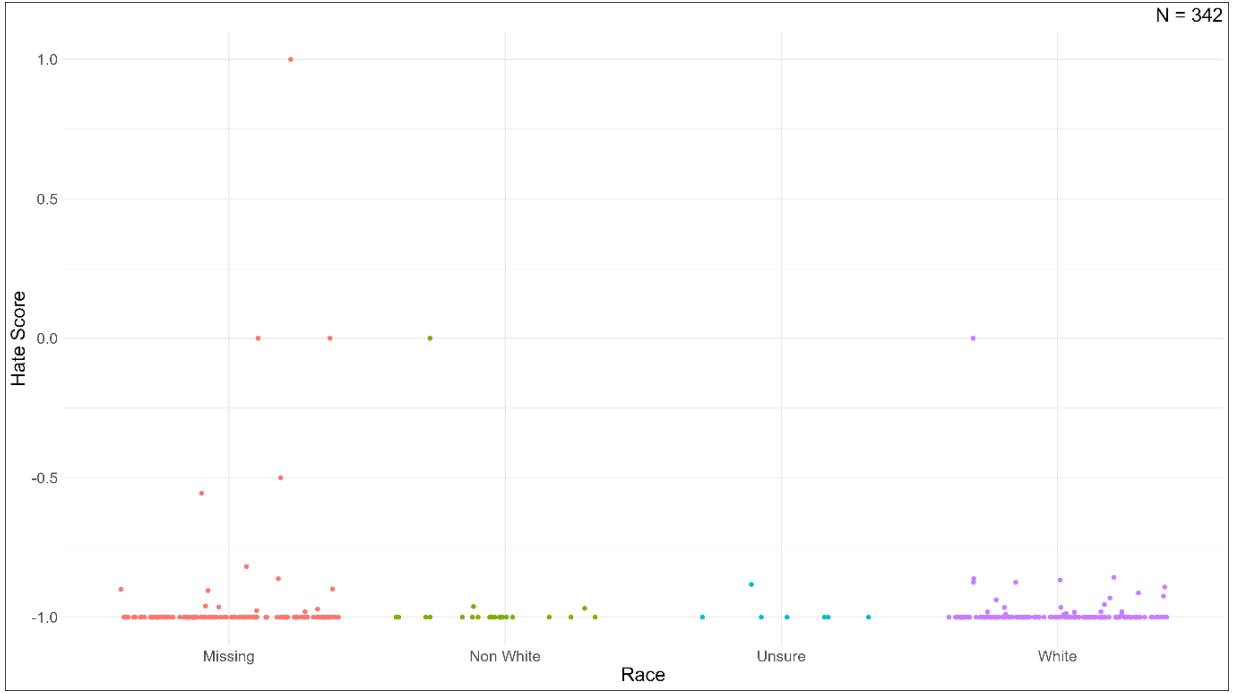
Channel Polarisation depending on Gender and Race

	Dependent Var.: SD Sentiment on Channel Level
Sex [Ref: Male]	
Missing	−0.003 (0.007)
Female	**−0.012* (0.006)**
Race [Ref.: White]	
Missing	0.003 (0.007)
Non White	**0.036** (0.015)**
Unsure	0.026 (0.019)
Visibility [Ref.: No]	
Yes	−0.002 (0.006)
Interaction Race x Visibility	
Missing x Yes	0.038 (0.025)
Non White x Yes	−0.010 (0.019)
Unsure x Yes	−0.031 (0.027)
Subscriber [in 100.000]	0.006** (0.003)
Channelage [in years]	−0.003*** (0.001)
Constant	0.288*** (0.016)
Observations	2,246
R^2	0.101
Adjusted R^2	0.089
Residual Std. Error	0.095 (df = 2214)
F Statistic	8.033*** (df = 31; 2214)
Note: controlled for age channel topic, education	*p<0.1; **p<0.05; ***p<0.01

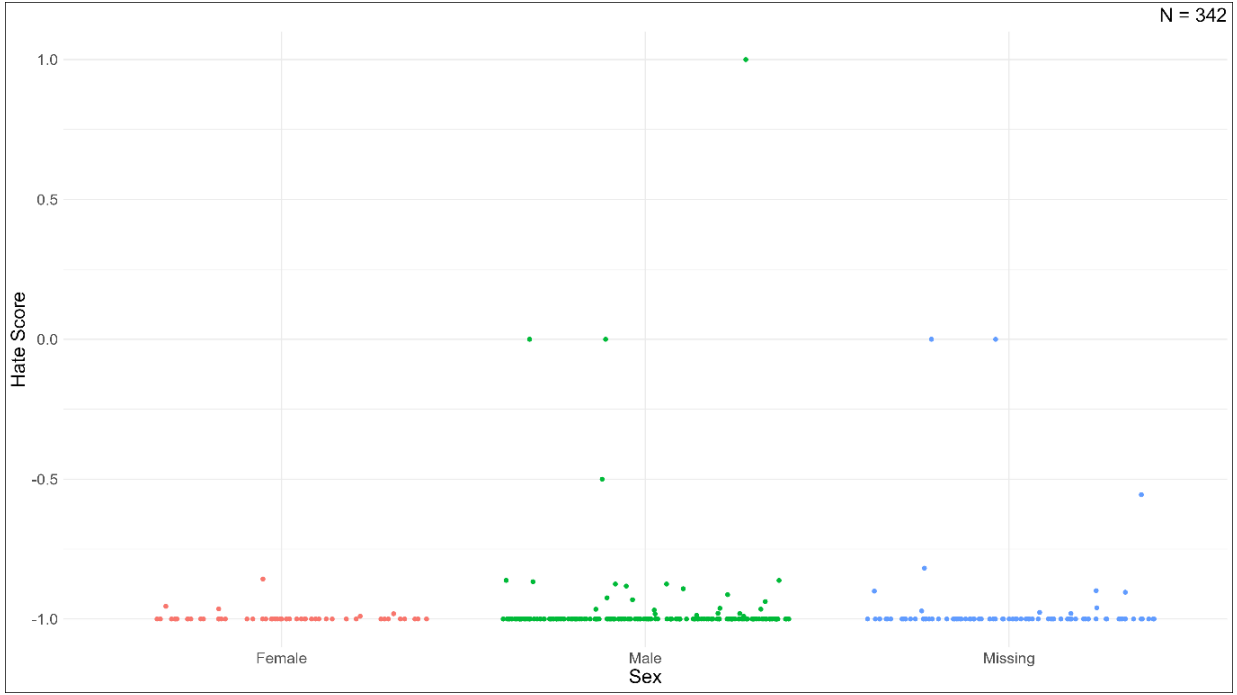
Hate-Speech Detection

- Analyses of text to see if they incite violence or hatred against a group of people or a member of such a group (e.g., race, gender, religion, sexual orientation, gender identity, disability, age)
  - Identification of discrimination (no hate/hate)
- Multilanguage mDeBERTa V3 Algorithm
  - Finetuning: German Hatespeech Dataset DeTox (Demus et al. 2022)
    - Accuracy: 0.97
    - F1 score: 0.62
    - Still currently calculating at our HPC-Cluster
    - For results of the preliminary and descriptive analysis based on N= 342 channels
- Resulting Variables:
  - Aggregated hate-speech score on channel level

Hate-Speech on Channel Level - Race



Hate-Speech on Channel Level - Sex



Our Contribution

- Solving methodological challenges of digital trace data
- Analyzing social media data regarding social-structural differences with significant effects in sex (rejection H1) and race (evidence for H2) for sentiment
  - No significant effect for age and education but for sex and race
  - Visibility as key explaining variable for differences sentiment by race
  - Platform structure/channel characteristics (channel topic, channel age, channel popularity) are related to sentiments patterns (as cluster analysis showed)
- A ‘low’ prevalence of hate in the YouTube Context (0.7 %)
- Potential biases:
  - Moderation of comments (deleting, reporting etc.) cannot be captured
  - Algorithmic structure of the platform is unknown (effects of polarization on views, moderation algorithms etc.)

Thank you for your attention!

References

- Amarasekara, I., & Grant, W. J. (2019). Exploring the YouTube science communication gender gap: A sentiment analysis. Public Understanding of Science, 28(1), 68–84. https://doi.org/10.1177/0963662518786654
- Barbieri, F., Anke, L. E., & Camacho-Collados, J. (2021). Xlm-t: Multilingual language models in twitter for sentiment analysis and beyond. arXiv preprint arXiv:2104.12250.
- Demus, C., Pitz, J., Schütz, M., Probol, N., Siegel, M., & Labudde, D. (2022). A comprehensive dataset for german offensive language and conversation analysis. In Proceedings of the Sixth Workshop on Online Abuse and Harms (WOAH), 143-153.
- Döring, N., & Mohseni, M. R. (2019). Fail videos and related video comments on YouTube: a case of sexualization of women and gendered hate-speech?. Communication Research Reports, 36(3), 254-264.
- Ghaffari, S. (2022). Discourses of celebrities on Instagram: digital femininity, self-representation and hate-speech. Critical Discourse Studies, 19(2), 161-178.
- He, P., Gao, J., & Chen, W. (2021). Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. arXiv preprint arXiv:2111.09543.
- Jahan, M. S., & Oussalah, M. (2023). A systematic review of hate-speech automatic detection using Natural Language Processing. Neurocomputing, 126232.

References

- Keum, B. T., & Miller, M. J. (2017). Racism in digital era: Development and initial validation of the Perceived Online Racism Scale (PORS v1.0). *Journal of Counseling Psychology, 64*(3), 310–324.
- Kumar, A., & Jaiswal, A. (2020). Systematic literature review of sentiment analysis on Twitter using soft computing techniques. Concurrency and Computation: Practice and Experience, 32(1), e5107.
- Latorre, J. P., & Amores, J. J. (2021). Topic modelling of racist and xenophobic YouTube comments. Analyzing hate-speech against migrants and refugees spread through YouTube in Spanish. In Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21), 456-460.
- Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., & Meira Jr, W. (2020). Auditing radicalization pathways on YouTube. In Proceedings of the 2020 conference on fairness, accountability, and transparency, 131-141.
- Siegel, A. A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., Nagler, J., & Tucker, J. A. (2021). Trumping hate on Twitter? Online hate-speech in the 2016 US election campaign and its aftermath. Quarterly Journal of Political Science, 16(1), 71-104.
- Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: implications for clinical practice. American Psychologist, 62(4), 271–286.

References

- Thomas, K., Kelley, P. G., Consolvo, S., Samermit, P., & Bursztein, E. (2022). “It’s common and a part of being a content creator”: Understanding How Creators Experience and Cope with Hate and Harassment Online. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1-15.
- Tynes, B. M., Lozada, F. T., Smith, N. A., & Stewart, A. (2018). From racial microaggressions to hate crimes: A model of online racism based on the lived experiences of adolescents of color. In C. Capodilupo, K. Nadal, D. Rivera, D. W. Sue, & G. Torino (Eds.), Microaggression theory: Influence and implications. Hoboken, NJ: Wiley.
- Veletsianos, G., Kimmons, R., Larsen, R., Dousay, T. A., & Lowenthal, P. R. (2018). Public comment sentiment on educational videos: Understanding the effects of presenter gender, video format, threading, and moderation on YouTube TED talk comments. PLOS ONE, 13(6), e0197331. https://doi.org/10.1371/journal.pone.0197331

Appendix

Evaluation Bert Models

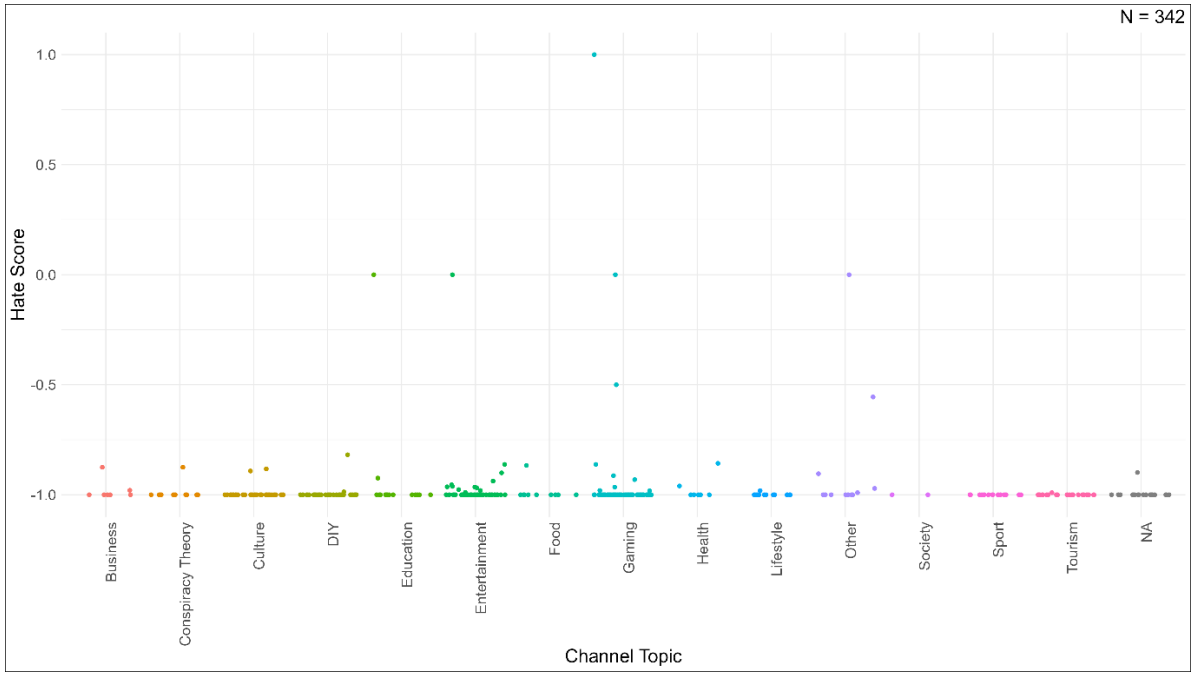
mDeberta Model V3

Accuracy	F1-Macro	Accuracy balanced	F1-Micro	Precision Macro	Recall Macro	Precision Micro	Recall Micro
0.9726	0.6181	0.5982	0.9726	0.6491	0.5983	0.9726	0.9726

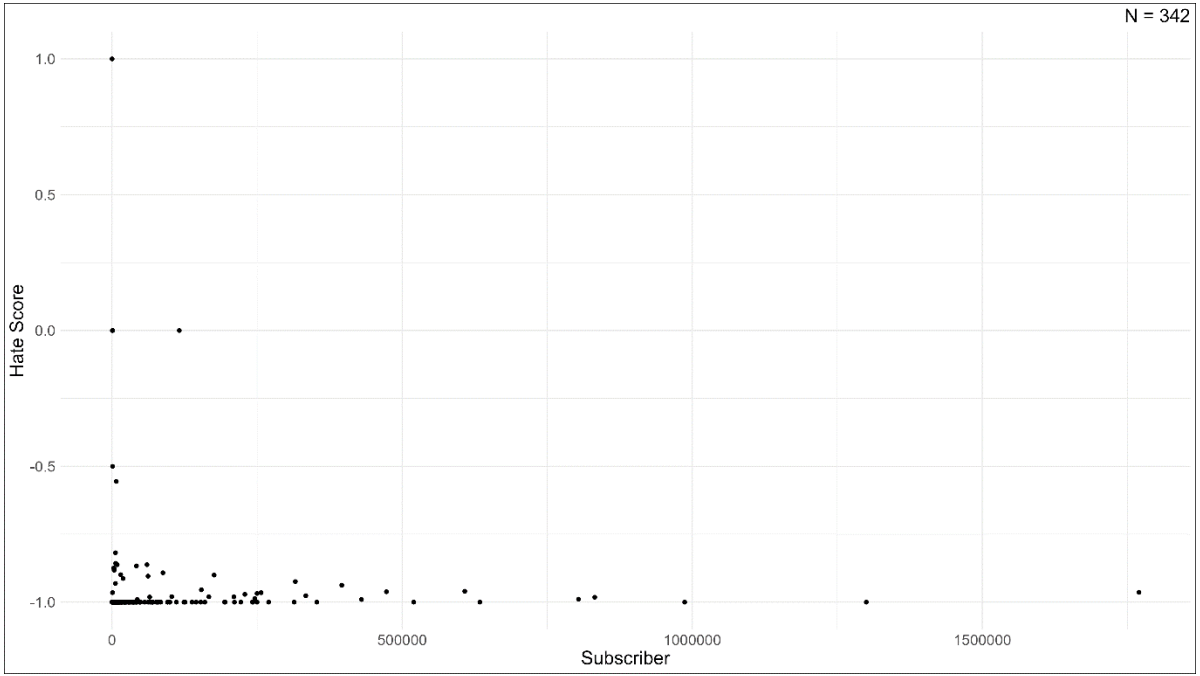
German Bert Model

Accuracy	F1-Macro	Accuracy balanced	F1-Micro	Precision Macro	Recall Macro	Precision Micro	Recall Micro
0.9678	0.5688	0.5489	0.9678	0.6302	0.5489	0.9672	0.9672

Hate-Speech on Channel Level - Channel Topic



Hate-Speech on Channel Level - Subscriber Count



Hate-Speech on Channel Level

	Dependent Var.: Mean Hate on Channel Level			
	(1)	(2)	(3)	(4)
NA sex [Ref.: male]	0.008 (0.021)	−0.001 (0.024)	−0.013 (0.027)	−0.014 (0.027)
female [Ref.: male]	−0.008 (0.017)	−0.007 (0.017)	−0.006 (0.017)	−0.005 (0.017)
NA race [Ref.: white]		0.013 (0.017)	0.003 (0.022)	0.004 (0.023)
non white [Ref.: white]		0.015 (0.020)	0.016 (0.021)	0.016 (0.021)
NA age [Ref.: under 30 years]			0.027 (0.025)	0.025 (0.026)
31-50 years [Ref.: under 30 years]			0.007 (0.015)	0.006 (0.015)
51+ years [Ref.: under 30 years]			0.010 (0.023)	0.009 (0.023)
NA education [Ref.: high education]				0.006 (0.019)
low education [Ref.: high education]				−0.008 (0.040)
subscriber [in 100.000]	−0.0001 (0.005)	−0.001 (0.005)	0.00003 (0.005)	−0.0002 (0.005)
education [Ref.: DIY]	0.058* (0.034)	0.055 (0.034)	0.052 (0.035)	0.052 (0.035)
finances [Ref.: DIY]	0.007 (0.038)	0.009 (0.038)	0.013 (0.039)	0.012 (0.039)
Gaming [Ref.: DIY]	0.023 (0.021)	0.021 (0.021)	0.025 (0.022)	0.024 (0.022)
cooking [Ref.: DIY]	0.011 (0.036)	0.010 (0.036)	0.012 (0.037)	0.011 (0.037)
music [Ref.: DIY]	−0.0001 (0.026)	−0.002 (0.026)	0.001 (0.026)	0.001 (0.027)
lifestyle [Ref.: DIY]	0.004 (0.029)	0.003 (0.029)	0.005 (0.030)	0.004 (0.030)
mobility [Ref.: DIY]	−0.002 (0.029)	−0.003 (0.029)	−0.001 (0.031)	−0.001 (0.031)
health [Ref.: DIY]	0.011 (0.040)	0.009 (0.040)	0.010 (0.040)	0.010 (0.041)
other [Ref.: DIY]	0.001 (0.034)	0.001 (0.034)	0.001 (0.034)	−0.0002 (0.035)
sport [Ref.: DIY]	−0.002 (0.036)	−0.002 (0.036)	0.002 (0.036)	0.001 (0.037)
entertainment [Ref.: DIY]	0.013 (0.023)	0.013 (0.023)	0.014 (0.023)	0.013 (0.023)
esotericism [Ref.: DIY]	0.008 (0.036)	0.008 (0.036)	0.003 (0.037)	0.003 (0.037)
Constant	0.003 (0.018)	−0.0005 (0.018)	−0.008 (0.022)	−0.012 (0.027)
Observations	235	235	235	235
R^2	0.028	0.032	0.038	0.039
Adjusted R^2	−0.039	−0.043	−0.052	−0.060
Residual Std. Error	0.088 (df = 219)	0.088 (df = 217)	0.088 (df = 214)	0.089 (df = 212)
F Statistic	0.419 (df = 15; 219)	0.428 (df = 17; 217)	0.426 (df = 20; 214)	0.394 (df = 22; 212)

Note: *p<0.1; **p<0.05; ***p<0.01

Methods of Analysis

- Hierarchical cluster analysis to study patterns in our sample
  - Sociodemographic variables: Age, Education, Race, Sex
  - Platform variables: Channel Age, Channel Topic, Subscriber count

- Regression analysis to investigate differences in sentiment and hate-speech
  - Dep. var.:
    - Mean Sentiment on Channel Level
    - SD Sentiment on Channel Level
    - (Hate-Speech on Channel Level & Hate-Speech on Video Level)
  - Independ. var.:
    - Age, Education, Race, Sex, Visibility
    - Channel Age, Channel Topic, number of Subscribers

Language Detection Performance for German Comments

	accuracy	precision	recall	f1
langdetect	0.881	0.679	0.984	0.803
textcat	0.829	0.561	0.976	0.713
cld3	0.873	0.641	0.995	0.780

Language Detection Performance for English Comments

	accuracy	precision	recall	f1
langdetect	0.938	0.960	0.974	0.967
textcat	0.939	0.947	0.989	0.967
cld3	0.961	0.967	0.993	0.980

Sentiment calculation in dictionary method

$$Sentiment^{1} = \frac{\#positive\ terms\ - \#negative\ terms}{\#all\ terms}$$

- Sentiment ϵ [-1;1]
- Sentiment > 0: positiv; Sentiment < 0: negativ
- Term-level Sentiment
- Dictionary: SentiWS (Universität Leipzig)[2] & GermanPolarityClues[3], AFINN & Emoji Sentiment Randing v1.0[4]

[1] Young, L., & Soroka, S. (2012). Affective News: The Automated Coding of Sentiment in Political Texts. *Political Communication*, *29*(2), 205–231. https://doi.org/10.1080/10584609.2012.671234

[2] Remus, R., Quasthoff, U., & Heyer, G. (2010). *SentiWS - A Publicly Available German-language Resource for Sentiment Analysis*.

[3] http://www.ulliwaltinger.de/german-polarity-clues/ (letzter Zugriff: 07.03.2023 14:20)

[4] https://kt.ijs.si/data/Emoji_sentiment_ranking/ (letzter Zugriff: 29.08.2023 10:20)

Comparison of different dictionarys

Performance of different Dictionarys		
	F1-Macro	Accuracy
English		
AFINN	0.426	0.478
VADER	0.409	0.451
AFINN, VADER	0.390	0.428
German		
German Polarity Clues	0.459	0.470
SentiWS	0.435	0.494
SentiWS, German Polarity Clues	0.467	0.491
SentiWS, German Polarity Clues, Emoji	0.463	0.504
All	0.460	0.501
All without VADER	0.467	0.508
N = 1217		
Treshold = +-0.05		

Performance of the dictionary method

Sentiment Classification with the Dictionary Method

	Truth		
Prediction	negativ[1]	neutral[2]	positiv[3]
negativ	111	87	18
neutral	84	86	36
positiv	99	186	329

[1] F1-Score: 0.435

[2] F1-Score: 0.304

[3] F1-Score: 0.66

F1-Macro = 0.467

Accuracy: 0.508

N = 1217

Performance of Lasso Regression

Sentiment Classification with Lasso Regression

	Truth		
Prediction	negativ[1]	neutral[2]	positiv[3]
negativ	245	30	15
neutral	5	253	19
positiv	66	200	384

[1] F1-Score: 0.809

[2] F1-Score: 0.666

[3] F1-Score: 0.719

F1-Macro = 0.731

Accuracy: 0.725

N = 1217

Trained with the DeTox Dataset

Sentiment Classification with Lasso Regression

	Truth		
Prediction	negativ[1]	neutral[2]	positiv[3]
negativ	38	21	7
neutral	38	87	62
positiv	3	13	36

[1] F1-Score: 0.524

[2] F1-Score: 0.565

[3] F1-Score: 0.459

F1-Macro = 0.516

Accuracy: 0.528

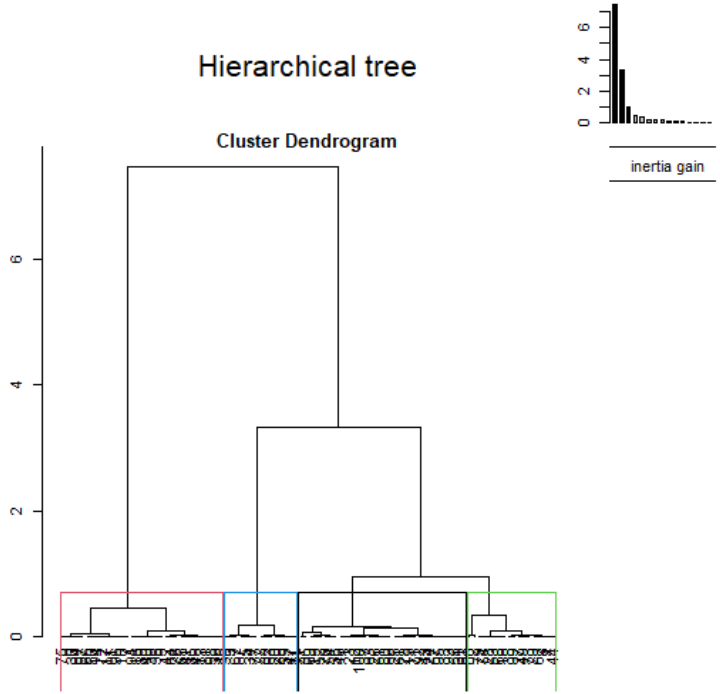
N = 305

Trained with 912 hand-coded comments

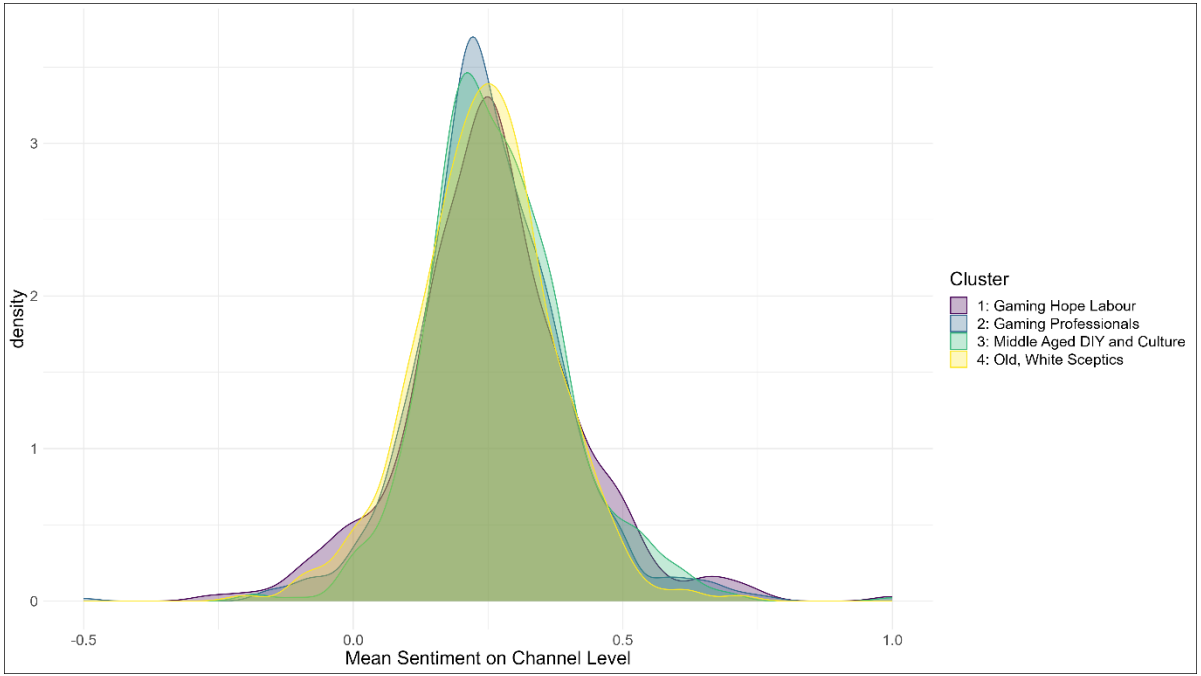
v.Test values for the Cluster analysis

Variable	1: Gaming Hope Labour	2: Gaming Professionals	3: Middle Aged Culture and DIY	4: Old White Sceptics
Conspiracy Theory	-1,86	0,0756	-1,49	3,11
Culture	-1,91	-2,92	1,43	3,54
DIY	-2,03	-2,25	2,68	1,84
Gaming	4,82	3,18	-2,06	-6,86
under 20 years	10,3	2,58	-5,2	-11,5
21 - 30 years	-1,53	3,44	0,164	-2,03
31 - 40 years	-2,49	-2,35	1,64	3,46
40+ years	-6,05	-4,27	3,76	6,83
High Education	-2,71	-2,09	1,97	3,07
Low Education	6,39	0,574	-3,24	-5,9
White	-4,65	-1,34	2,29	4,38
Channel Age	-35	-6,85	8,85	37,4
Polarisation (SD Sentiment)	4,23	3,37	-2	-6,23
Mean Sentiment	-1,53	0,502	3,38	-1,93
Channel Size (Subscriber)	-4,12	3,54	-2,19	2,97

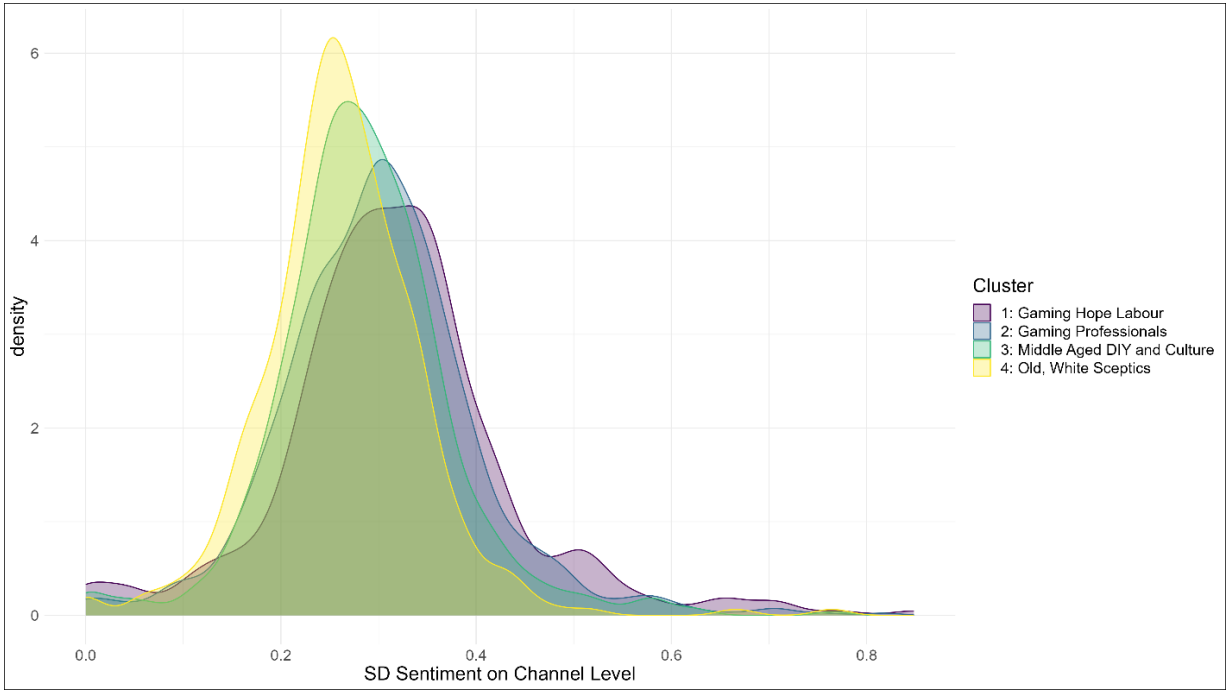
Dendogramm Clustering



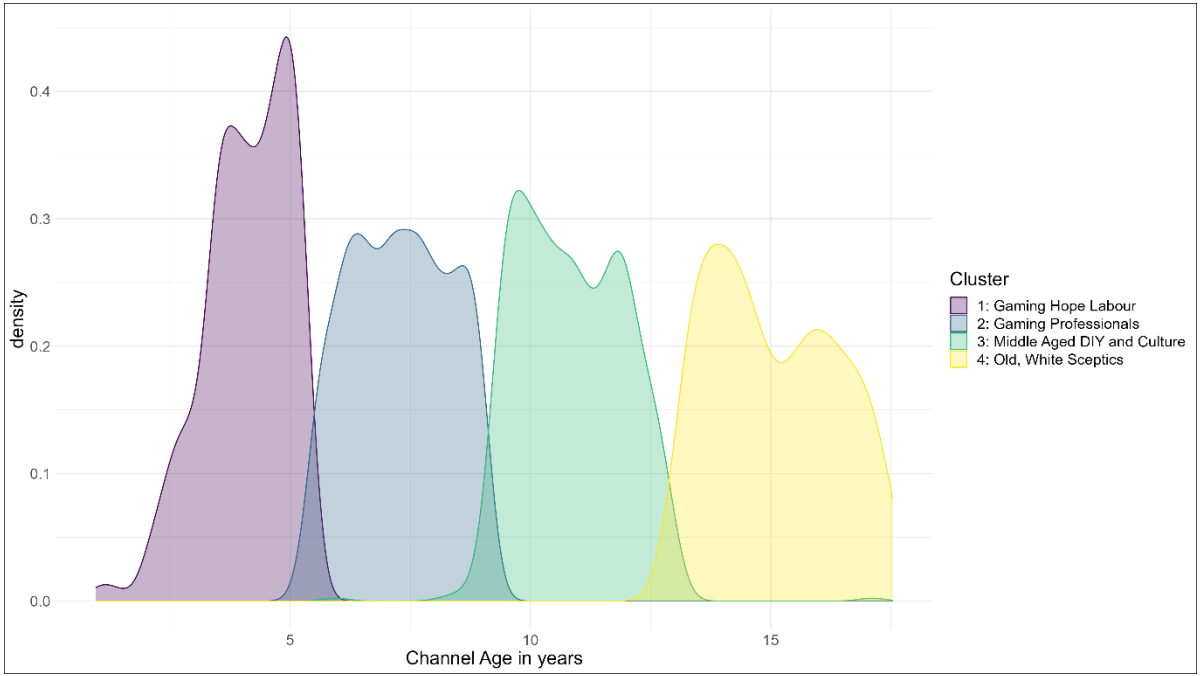
Cluster characteristics



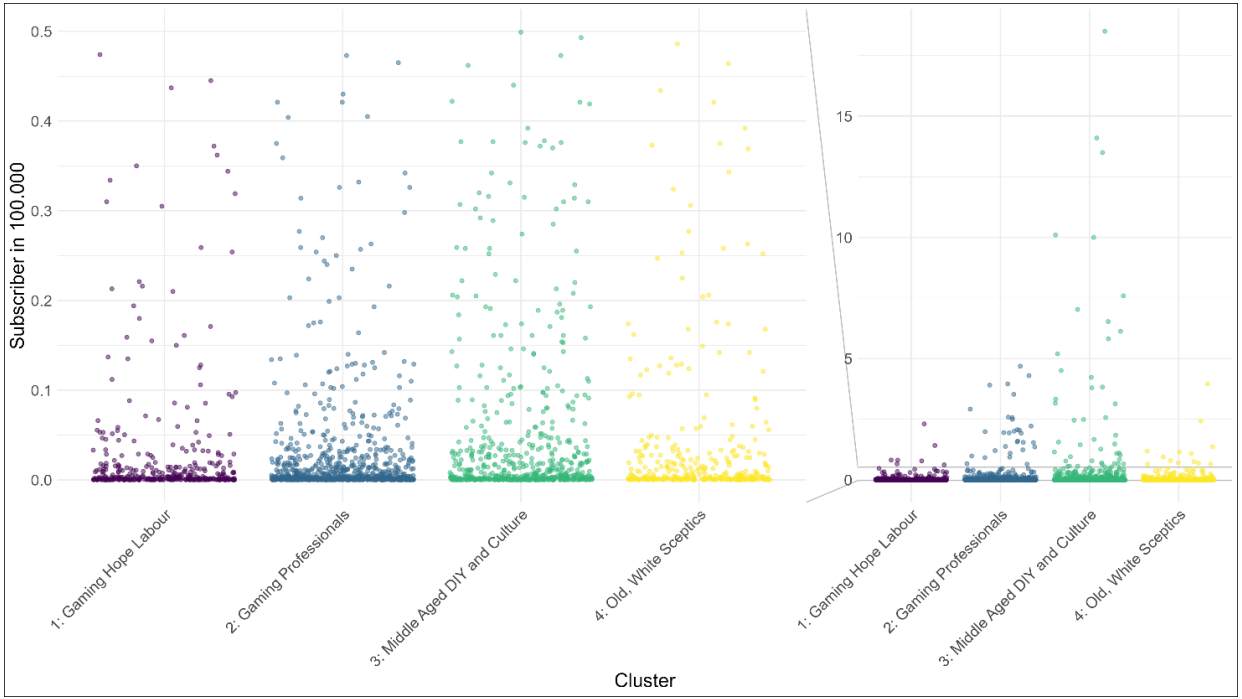
Cluster characteristics



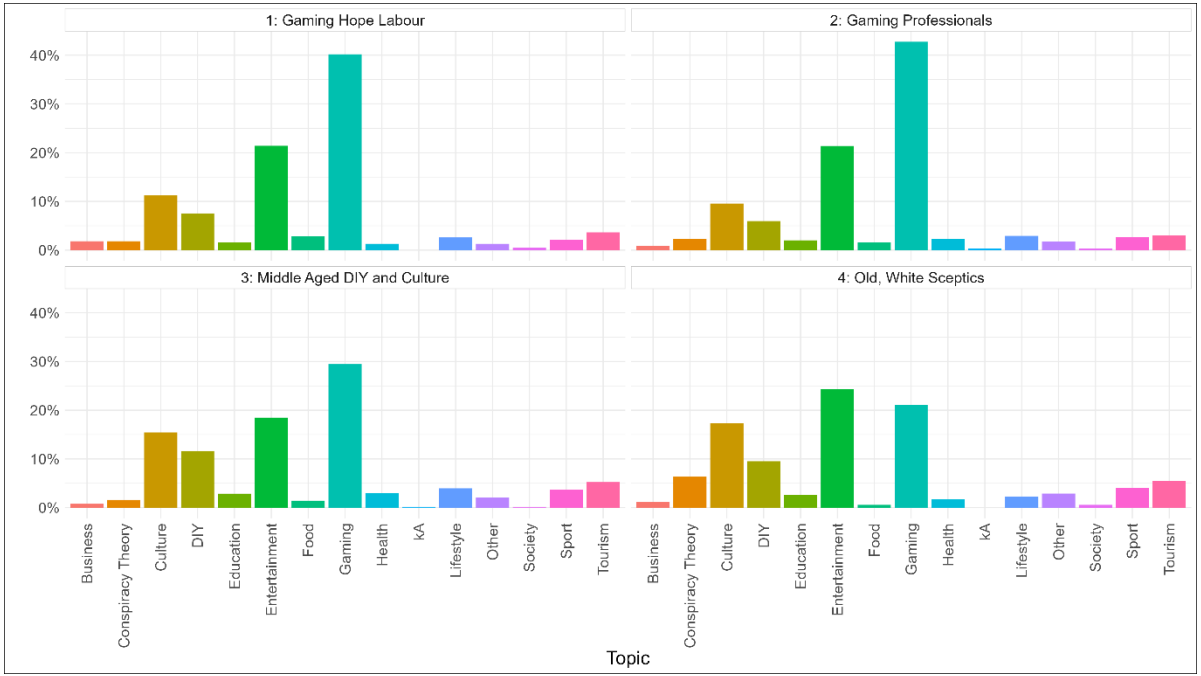
Cluster characteristics



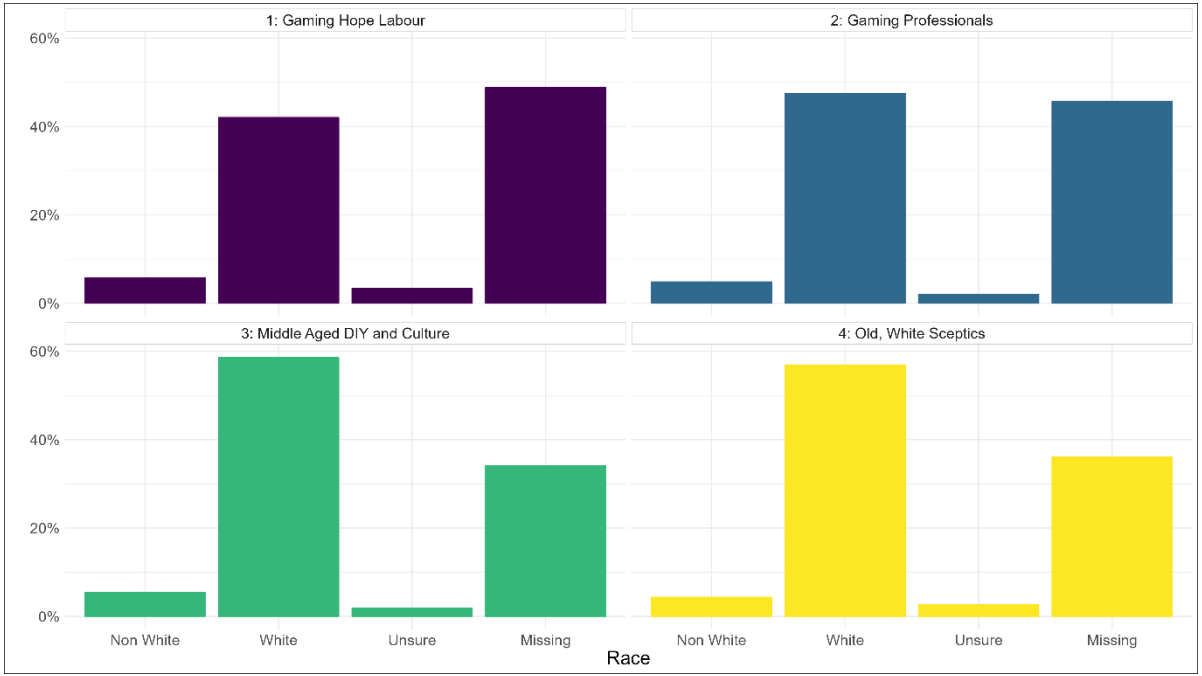
Cluster characteristics



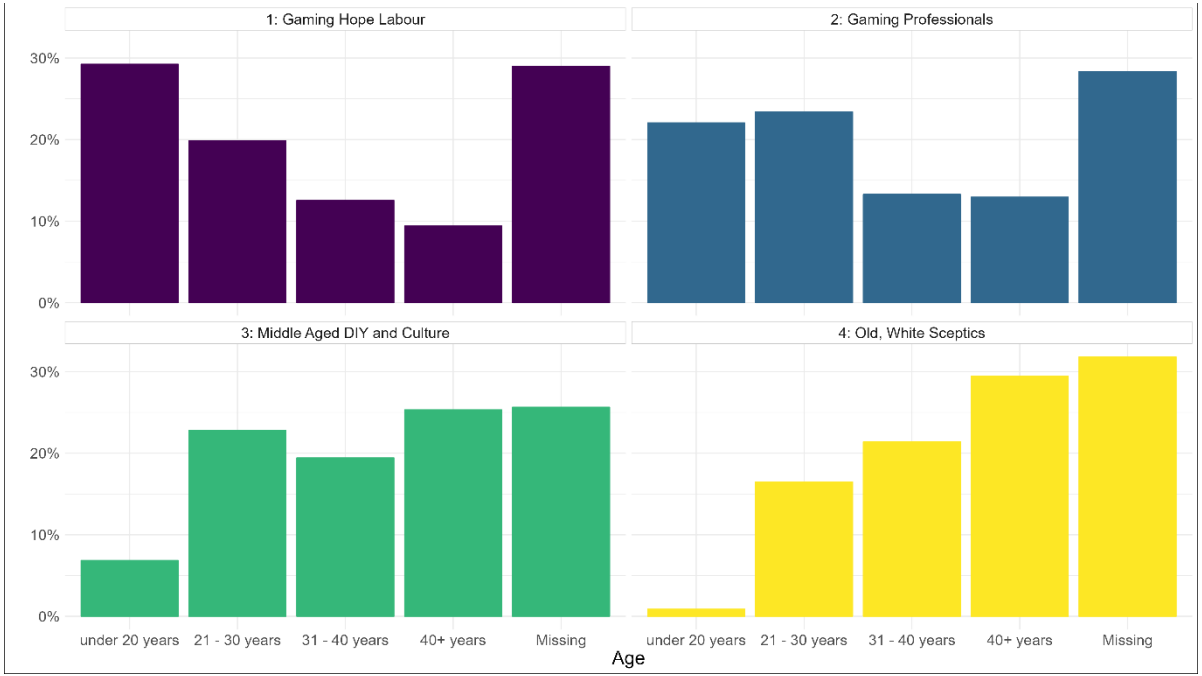
Cluster characteristics



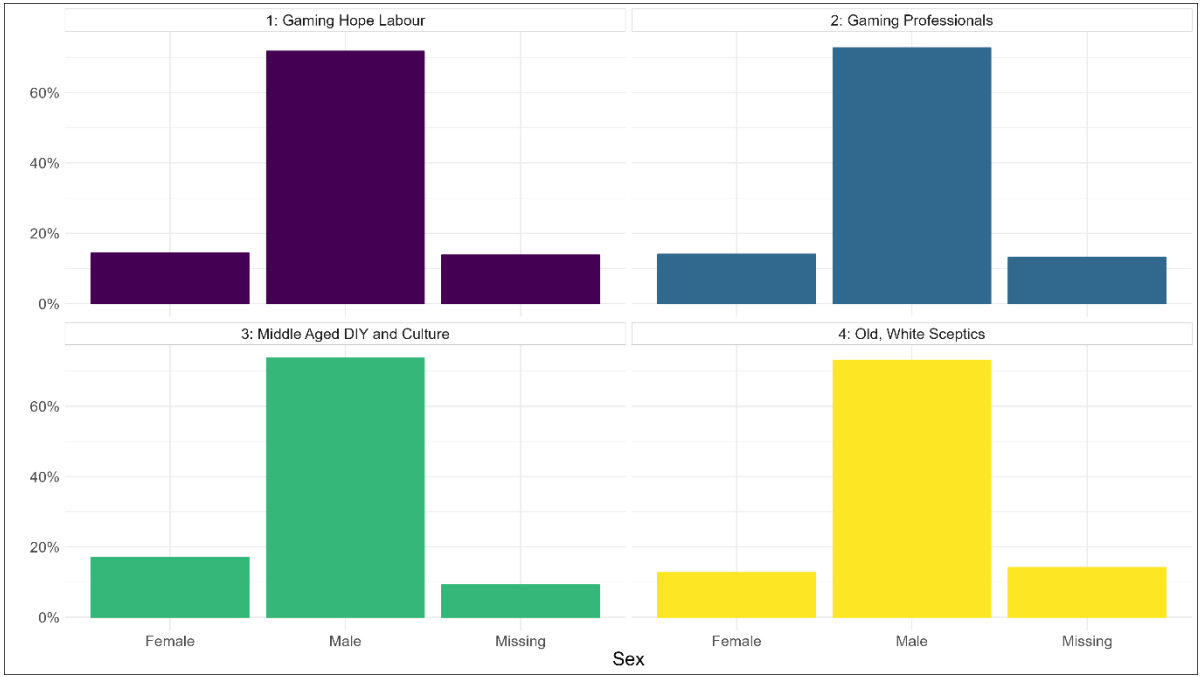
Cluster characteristics



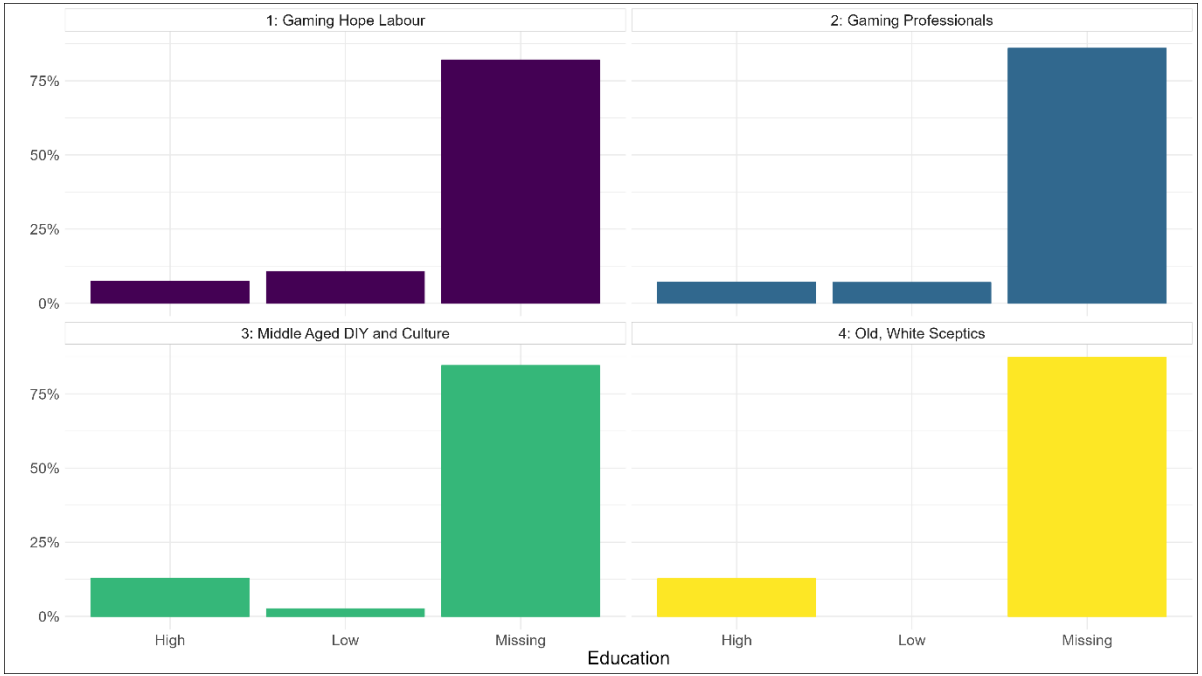
Cluster characteristics



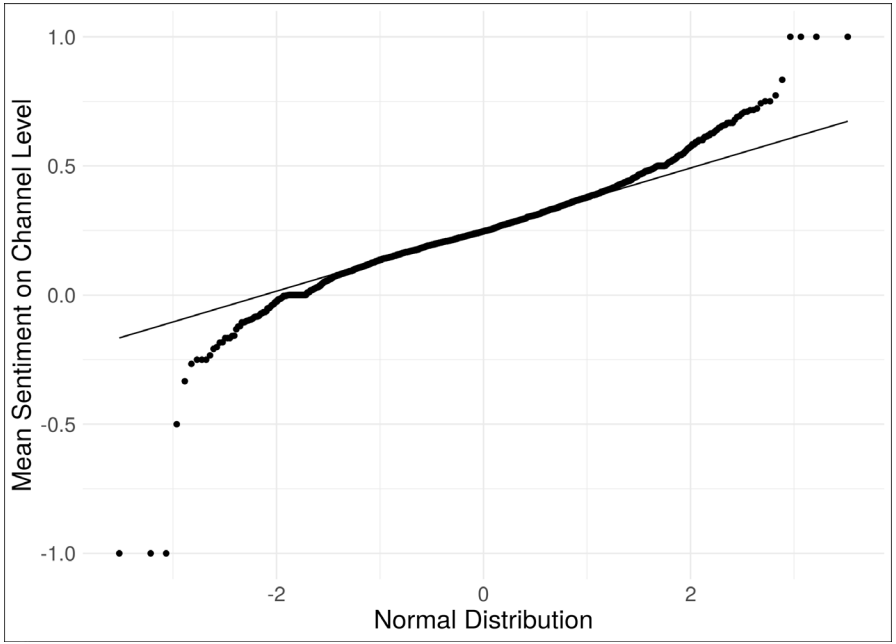
Cluster characteristics

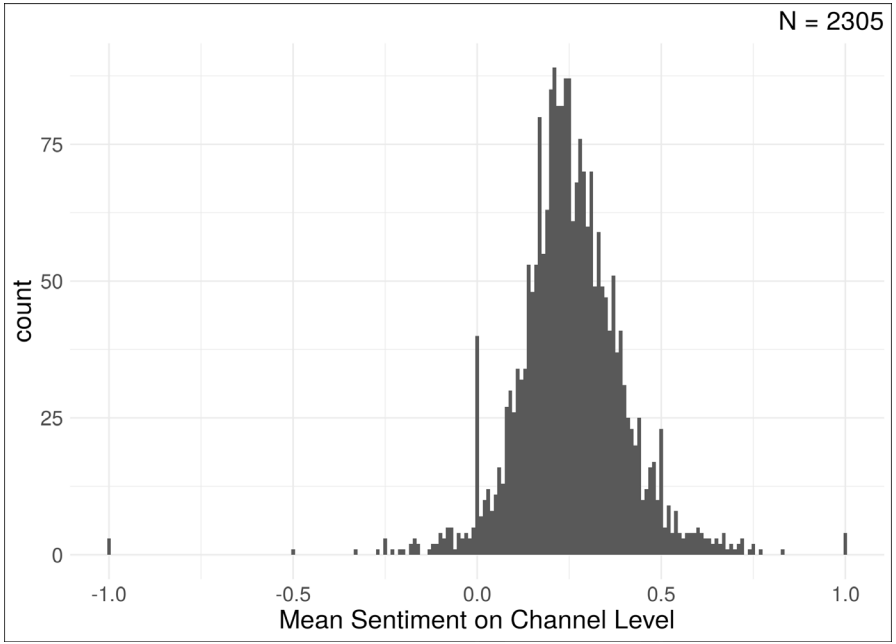


Cluster characteristics

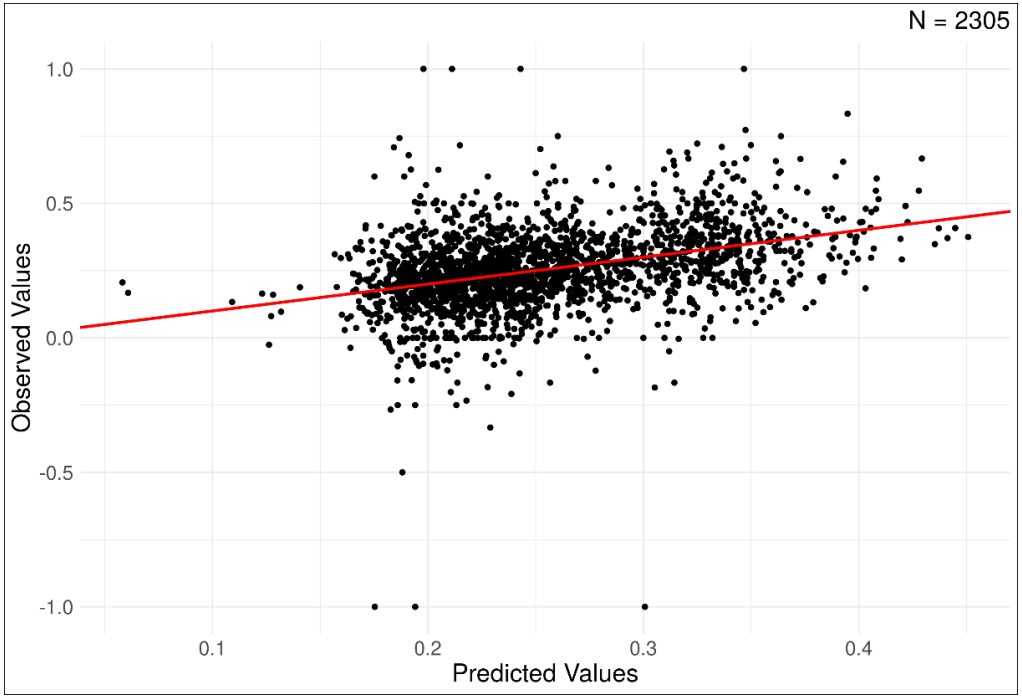


Distribution of mean sentiment on channel level





Predicted vs. observed plot of mean sentiment on channel level



Channel sentiment with single predictor

	Dependent Var.: Mean Sentiment on Channel Level		
	(1)	(2)	(3)
Sex [Ref: Male]			
Missing	−0.021** (0.009)		
Female	**0.097*** (0.009)**		
Race [Ref.: White]			
Missing		−0.035*** (0.008)	
Non White		**0.064*** (0.022)**	
Unsure		0.019 (0.029)	
Visibility [Ref.: No]			
Yes		0.017* (0.009)	
Interaction Race x Visibility			
Missing x Yes		0.009 (0.038)	
Non White x Yes		**−0.085*** (0.029)**	
Unsure x Yes		−0.021 (0.042)	
Age [Ref.: 40+ years]			
Missing			−0.041*** (0.009)
≤ 20 years			−0.059*** (0.011)
21-30 years			−0.009 (0.010)
31-40 years			0.003 (0.011)
Constant	0.242*** (0.004)	0.262*** (0.006)	0.276*** (0.007)
Observations	2,305	2,305	2,305
R^2	0.059	0.026	0.024
Adjusted R^2	0.058	0.023	0.022
Residual Std. Error	0.145 (df = 2302)	0.148 (df = 2297)	0.148 (df = 2300)
F Statistic	71.731*** (df = 2; 2302)	8.860*** (df = 7; 2297)	14.144*** (df = 4; 2300)
Note:			*p<0.1; **p<0.05; ***p<0.01

Channel sentiment with single predictor

	Dependent Var.: Mean Sentiment on Channel Level	
	(4)	(5)
Education [Ref.: High]		
Missing	−0.017 (0.011)	
Low	−0.059*** (0.017)	
Subscriber [in 100.000]		−0.011*** (0.004)
Channelage [in years]		−0.003*** (0.001)
Channel Topic [Ref.: Business]		
Missing		0.109 (0.077)
Conspiracy Theory		0.033 (0.034)
Culture		0.108*** (0.030)
DIY		0.033 (0.031)
Education		0.025 (0.035)
Entertainment		0.023 (0.030)
Food		0.087** (0.038)
Gaming		−0.021 (0.030)
Health		0.082** (0.035)
Lifestyle		0.093*** (0.034)
Society		−0.068 (0.058)
Sport		0.054 (0.034)
Tourism		0.054 (0.033)
Other		0.063* (0.036)
Constant	0.271*** (0.010)	0.251*** (0.030)
Observations	2,305	2,305
R^2	0.005	0.091
Adjusted R^2	0.005	0.085
Residual Std. Error	0.149 (df = 2302)	0.143 (df = 2288)
F Statistic	6.240*** (df = 2; 2302)	14.356*** (df = 16; 2288)
Note:		*p<0.1; **p<0.05; ***p<0.01

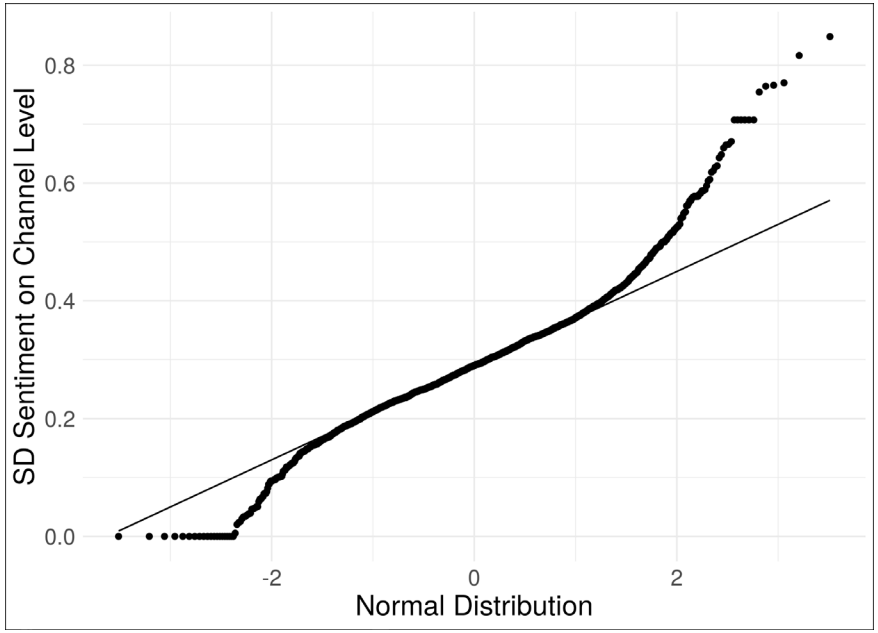
Channel sentiment (full model)

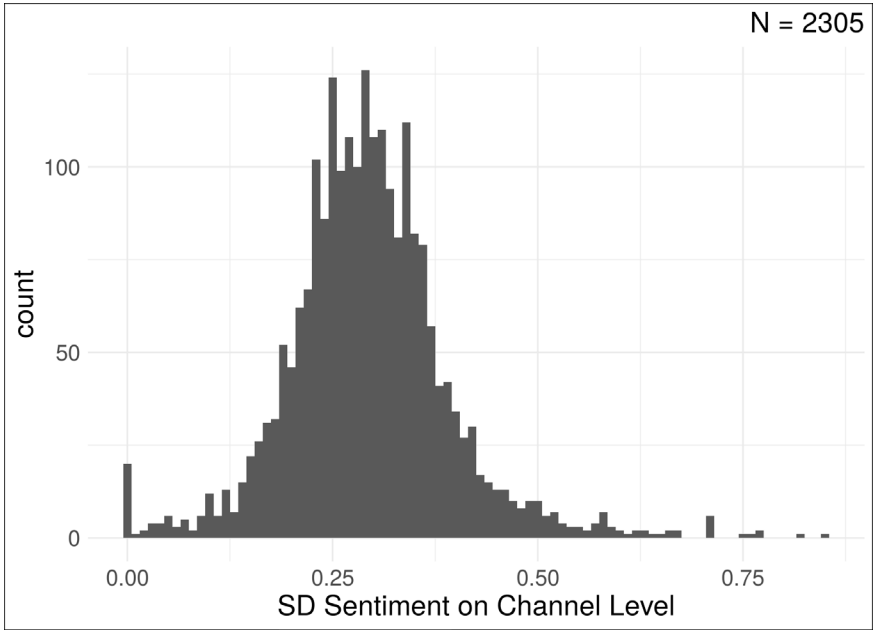
	Dependent Var.: Mean Sentiment on Channel Level
Sex [Ref: Male]	
Missing	−0.019* (0.011)
Female	0.080*** (0.009)
Race [Ref.: White]	
Missing	−0.019** (0.010)
Non White	0.042** (0.021)
Unsure	0.003 (0.028)
Visibility [Ref.: No]	
Yes	0.001 (0.009)
Interaction Race x Visibility	
Missing x Yes	−0.009 (0.037)
Non White x Yes	−0.067** (0.027)
Unsure x Yes	−0.007 (0.040)
Age [Ref.: 40+ years]	
Missing	0.004 (0.012)
$\leq$ 20 years	−0.019 (0.013)
21-30 years [Ref.: 40+ years]	0.005 (0.010)
31-40 years	0.011 (0.010)
Education [Ref.: High]	
Missing	0.015 (0.011)
Low	−0.001 (0.018)
Note:	*p<0.1; **p<0.05; ***p<0.01

Channel sentiment (full model)

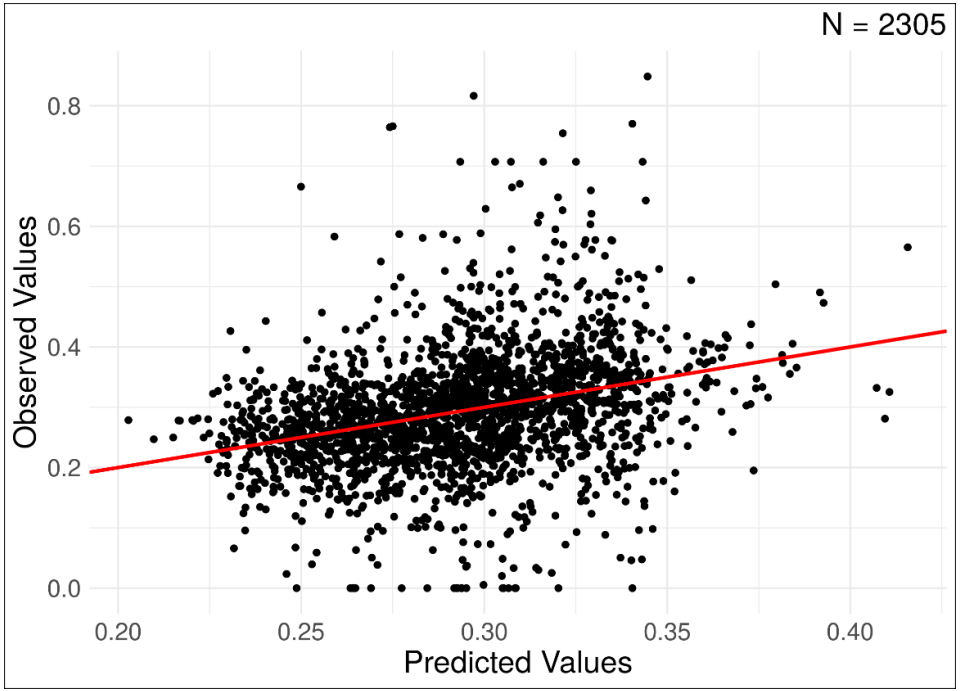
	Dependent Var.: Mean Sentiment on Channel Level
Subscriber [in 100.000]	−0.012*** (0.004)
Channelage [in years]	−0.003*** (0.001)
Channel Topic [Ref.: Business]	
Missing	0.149* (0.076)
Conspiracy Theory	0.015 (0.034)
Culture	0.105*** (0.030)
DIY	0.024 (0.030)
Education	0.023 (0.034)
Entertainment	0.025 (0.030)
Food	0.038 (0.037)
Gaming	−0.008 (0.029)
Health	0.055 (0.035)
Lifestyle	0.040 (0.034)
Society	−0.075 (0.057)
Sport	0.042 (0.033)
Tourism	0.057* (0.032)
Other	0.050 (0.035)
Constant	0.241*** (0.031)
Observations	2,305
R^2	0.140
Adjusted R^2	0.129
Residual Std. Error	0.139 (df = 2273)
F Statistic	11.966*** (df = 31; 2273)
Note:	*p<0.1; **p<0.05; ***p<0.01

Distribution of SD sentiment on channel level





Predicted vs. observed plot of SD sentiment on channel level



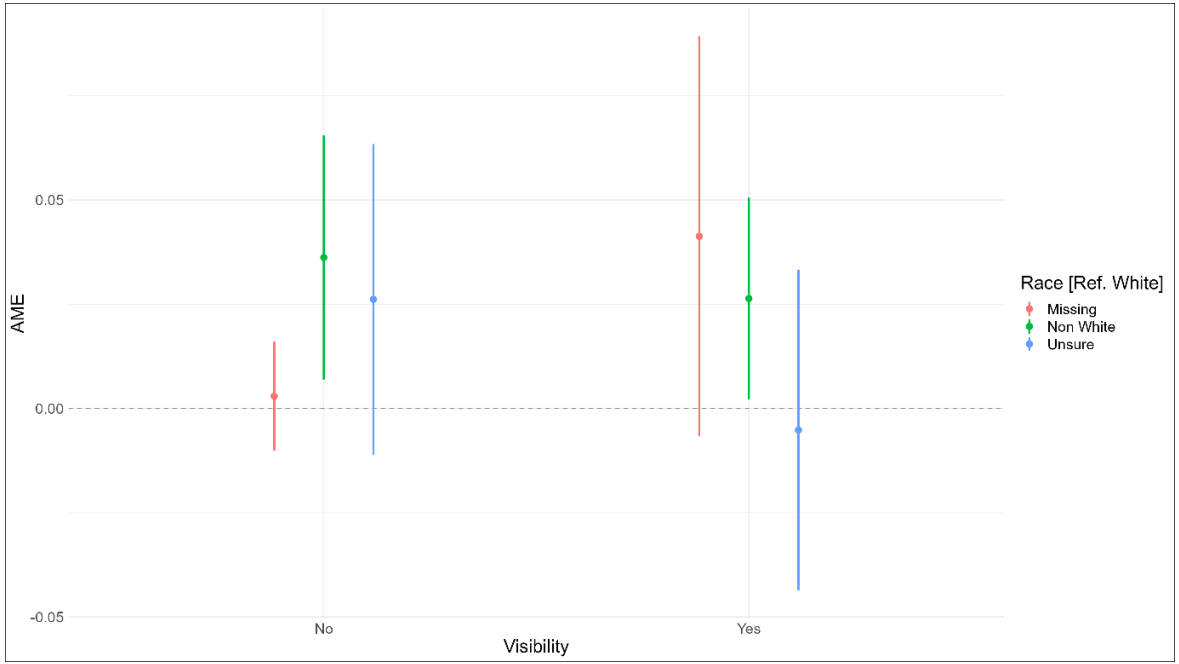
Channel Polarisation (full model)

	Dependent Var.: SD Sentiment on Channel Level
Sex [Ref: Male]	
Missing	−0.003 (0.007)
Female	−0.012* (0.006)
Race [Ref.: White]	
Missing	0.003 (0.007)
Non White	0.036** (0.015)
Unsure	0.026 (0.019)
Visibility [Ref.: No]	
Yes	−0.002 (0.006)
Interaction Race x Visibility	
Missing x Yes	0.038 (0.025)
Non White x Yes	−0.010 (0.019)
Unsure x Yes	−0.031 (0.027)
Age [Ref.: 40+ years]	
Missing	0.033*** (0.008)
≤ 20 years	0.059*** (0.009)
21-30 years	0.035*** (0.007)
31-40 years	0.009 (0.007)
Education [Ref.: High]	
Missing	−0.008 (0.007)
Low	−0.001 (0.013)
Note:	*p<0.1; **p<0.05; ***p<0.01

Channel Polarisation (full model)

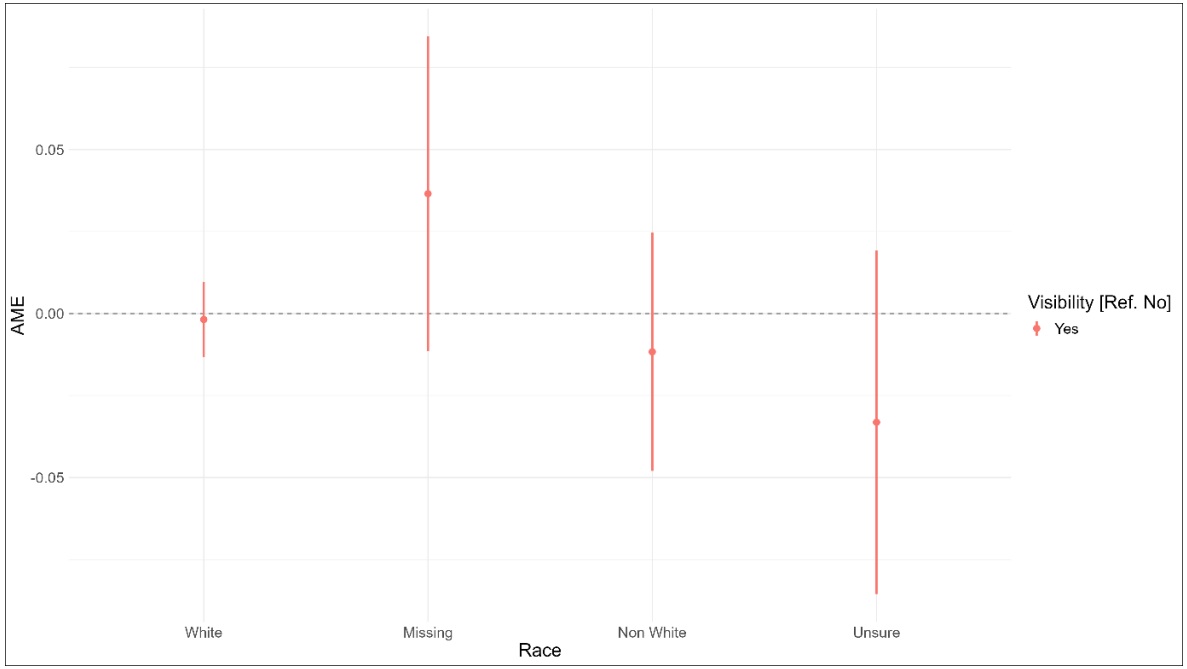
	Dependent Var.: SD Sentiment on Channel Level
Subscriber [in 100.000]	0.006** (0.003)
Channelage [in years]	−0.003*** (0.001)
Channel Topic [Ref.: Business]	
Missing	−0.007 (0.017)
Conspiracy Theory	0.007 (0.021)
Culture	−0.005 (0.015)
DIY	−0.018 (0.024)
Education	0.006 (0.014)
Entertainment	0.085* (0.050)
Food	0.042*** (0.015)
Gaming	0.011 (0.018)
Health	−0.030 (0.036)
Lifestyle	−0.004 (0.019)
Society	0.003 (0.018)
Sport	0.016 (0.014)
Tourism	−0.008 (0.018)
Other	0.005 (0.020)
Constant	0.288*** (0.016)
Observations	2,246
R^2	0.101
Adjusted R^2	0.089
Residual Std. Error	0.095 (df = 2214)
F Statistic	8.033*** (df = 31; 2214)
Note:	*p<0.1; **p<0.05; ***p<0.01

Polarisation differences in dependence of race



controlled for age, channel age, channel topic, education, sex, subscriber

Polarisation differences in dependence of race



controlled for age, channel age, channel topic, education, sex, subscriber

Concepts // Definitions

- **Sentiment analysis** or opinion mining is the computational study of people's opinions, sentiments, emotions, appraisals, and attitudes towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. (Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.)

- **Hate-Speech:** All conduct publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, color, religion, descent or national or ethnic. (Wigand, C. & Voin, M., (2017). Speech by commissioner jourová–10 years of the eu fundamental rights agency: A call to action in defence of fundamental rights, democracy and the rule of law.)