Detecting extreme ideologies in shifting landscapes

2023-02-02

Also check out [Rohit](../authors/rohit-ram/)’s and [Andrei](../authors/ma-rizoiu/)’s article that just appeared in The Conversation: [Can ideology-detecting algorithms catch online extremism before it takes hold?](https://theconversation.com/can-ideology-detecting-algorithms-catch-online-extremism-before-it-takes-hold-200629)

In this [our latest working paper](https://arxiv.org/pdf/2208.04097.pdf), we propose a completely automatic end-to-end ideology detection pipeline for the detection and psychosocial profiling of left-right political ideology, as well as far-right ideological users. The pipeline fills a crucial gap by providing flexible methodology and tooling for understanding ideologies and building early warning systems for extreme ideology-motivated (and potentially violent) activity.

**Paper citation:**

Ram, R. and Rizoiu, M.A., 2022. You are what you browse: A robust framework for uncovering political ideology.  
arXiv preprint arXiv:2208.04097.

(*see full paper here:* [*https://arxiv.org/pdf/2208.04097.pdf*](https://arxiv.org/pdf/2208.04097.pdf))

## Ideology in an Online World

Ideology determines how we make sense of much of the world, our opinions, and our political actions. It is not a new concept; throughout history, it served as the context for unrest. However, ideological spread and radicalization have entered a new paradigm in our ever-connected world. The internet is a significant source of information and spreads opinions quickly through social platforms. In particular, the anonymity and lack of accountability often associated with online communication set up a supportive environment for spreading far-right ideologies and radicalizing individuals into extremist groups. Far-right extremism is a form of ideology that advocates for ultranationalism, racism, and opposition to immigration and multiculturalism. These ideologies strongly correlate with violence and terrorism and threaten individual and collective security.

The Australian Security Intelligence Organisation (ASIO) raised concerns about Australians being radicalized very young and the rise of extremist movements in Australia through online technologies ([“Director-General’s Annual Threat Assessment” 2021](#ref-asio)). ASIO claimed that during the COVID period, 30-40% of their caseload was devoted to far-right extremism, up from 10-15% in 2016 ([Karp 2020](#ref-guardian1)).

Unfortunately, Ideologically Motivated Violent Extremism (IMVE) continues to be an issue in Australia. On December 12th, 2022, two Queensland police officers were killed while performing routine duties ([Gillespie and McGowan 2022](#ref-guardian2)). Later investigations would uncover that the three people, who killed the officers, were active online in producing deep-state and religious conspiratorial content. Their content has since been removed from mainstream social platforms but continues to be shared on conspiratorial websites. Such extreme-leaning content often serves as a lead indicator of violent extremism (as was the case in this incident and the Christchurch Mosque Shootings three years prior). However, the tools to identify and understand the psychosocial characteristics of these extreme individuals and communities are lacking.

In this work, we build an end-to-end ideology detection pipeline and psychosocial profiles of ideological groups. We find that right-leaning individuals tend to use moral-vice language more than left-leaning and that far-right individuals’ grievance language (violence, hate, paranoia, etc.) significantly differs from the moderates.

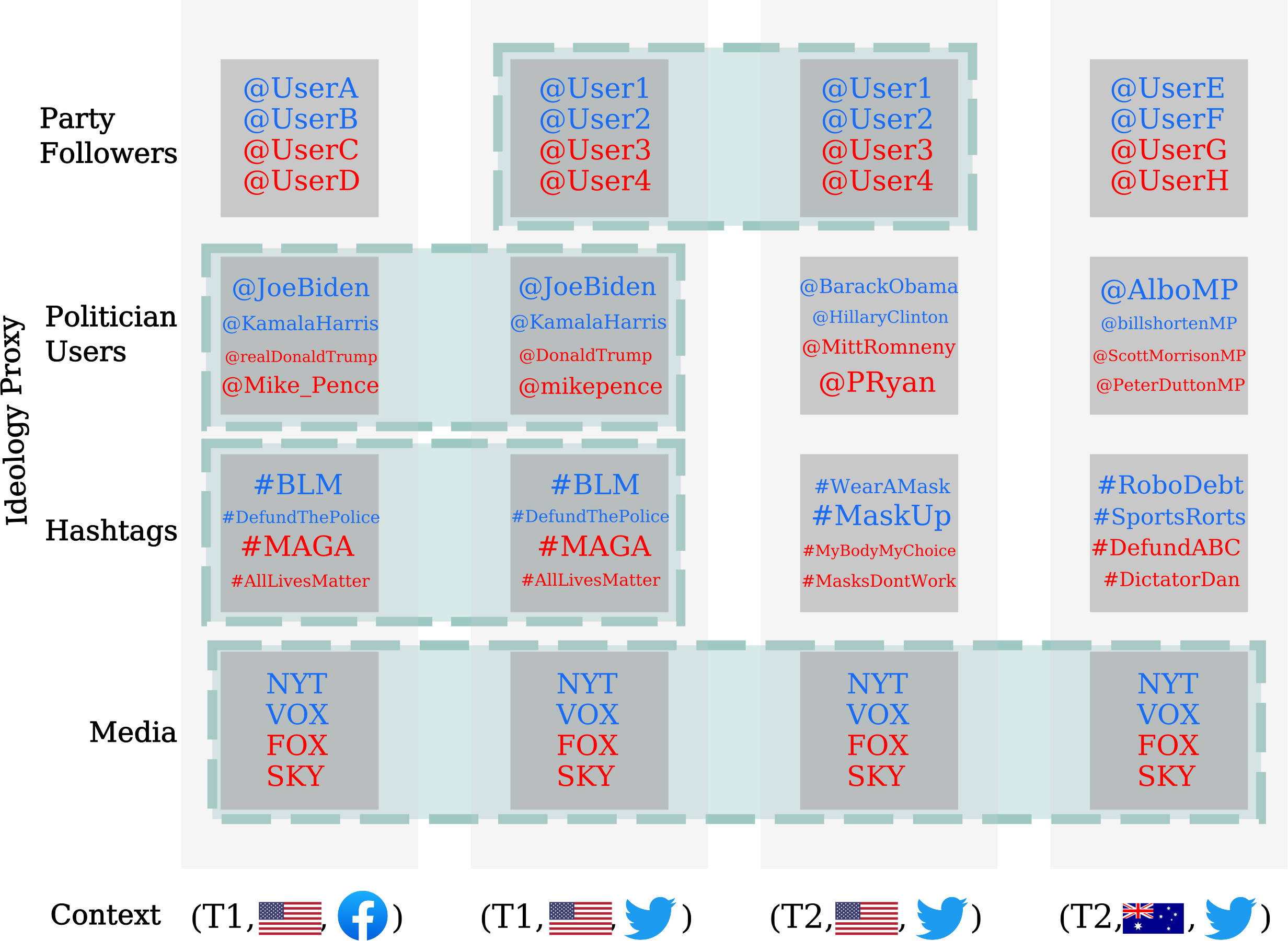
## Signals of Ideology

In online social settings, researchers face numerous barriers that prevent using traditional methods. Directly asking users for their ideologies has dubious success, infringes on platform T&Cs, and does not scale to online populations. Inferring users’ ideologies from their activity also does not scale as the data requires is prohibitively expensive and tedious to compile.

Instead, to reduce expert labor to feasible levels, researchers infer ideologies from signals in user behavior – such as whether they use political hashtags, retweet politicians, or follow political parties. We dub these signals *ideological proxies*.

Importantly, these *ideological proxies* for online users can still require laborious labeling by context-specific experts. For example, the hashtag *#ScottyFromMarketing* requires an up-to-date expert in Australian politics to uncover that it expresses an anti-right-wing ideology. For many researchers: - access to contextual experts is difficult, - labeling of signals is still laborious and expensive, - and context switches require relabelling (exasperating the above problems).

Unfortunately, such context switches are commonplace, as the context changes with time, country, or social platform. Figure @ref(fig:teaser) showcases the problem: most commonly used *ideological proxies* can only be transferred in narrow circumstances (represented by the green dotted regions). For example, following political parties is country-dependent, politicians come and go with time, and hashtags are platform-dependent. As such, we desire an *ideological proxy* that is robust to changes in context, requires no expert labeling and is true to the gold standard.



Furthermore, the *ideological proxies* are often sparse among users; however, we would ideally like to detect influence amongst the entire population of users (as taking only active users could bias our inferences). We further desire a method for inferring the ideology of (potentially inactive) users without direct *ideological proxy* information.

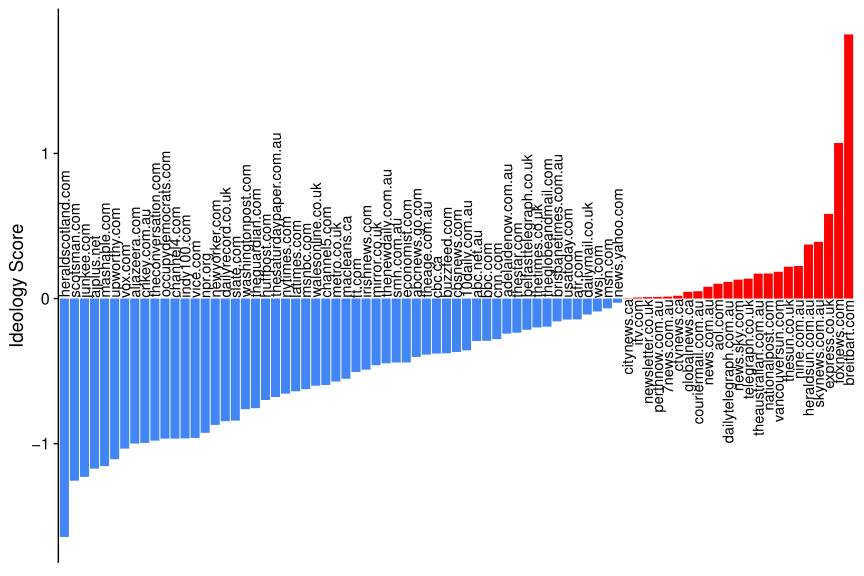
## Our Solution: You are what you browse

Our solution is a large-scale end-to-end ideology detection pipeline that can be used to profile entire populations of users. The solution has two main components; the media proxy and the inference architecture. The media proxy allows for labeling a subset of users, and the inference architecture allows for propagating these labels to the remaining users via socially-informed homophilic lenses.

### The Media Proxy

For the first part of our work, we generate a proxy based on media-sharing behavior, which satisfies the desiderata.

We generate the media proxy via two media slant datasets (although many are widely available). The first is an extensive survey of media consumption behaviors conducted by Reuters ([Newman et al. 2019](#ref-newman2019reuters)) in several countries in 2020 and 2021. Participants reported the media publications they consume and their own political ideology. We estimate the slant of a media source for each country and year as the average ideology of the participants who consume it. The second dataset is the Allsides Media Bias Dataset ([Sides 2018](#ref-sides2018media)), which contains an expert-curated set of media publications. The Allsides dataset contains mostly American-based media; conversely, Reuters covers the major media outlets in each country. Given that each country and period will have a different conception of ideologies, we calibrate Reuter’s media slants to approximate the Allsides (minimizing the mean-squared error). Figure @ref(fig:slants) shows the slants for each media website within the Reuters dataset.



Finally, we quantify a user’s ideology as the average ideology of their shared media.

The media proxy resolves the issue of context switching; since it is applicable across many contexts and can be used widely in a fully automated fashion. This allows us to create an end-to-end ideology detection pipeline.

We further define methods to classify far-right users from their media-sharing behaviors, which we fully describe in the paper.

### The Inference Architecture

In the second part of our work, we define an inference architecture that allows inferring the ideological labels of the remaining users – e.g., users who do not share any URLs. Our inference architecture relies on the sociological principle of homophily, where we hypothesize that similar users will share a similar ideology. We measure homophily through three distinct lenses;

1. *Lexical*: Users with similar language will have similar ideology
2. *Hashtag*: Users who participate in similar topics of discussion share a similar ideology
3. *Resharing*: Users who consume similar content (signaled via resharing of other users) will share a similar ideology

Through these lenses, we utilize an AutoML model, FLAML ([Wang et al. 2021](#ref-wang2021flaml)) (with the LightGBM architecture), trained on users identified via an ideological proxy to propagate the labels to the remaining users and generate a complete ideological profile for a dataset.

## The Data

We utilize several large-scale datasets from various platforms to showcase the relative ease of applying our end-to-end pipeline. The datasets’ characteristics are described in Table @ref(tab:datasets).

The datasets used through-out the analysis, with the number of users, posts, and affliated country.

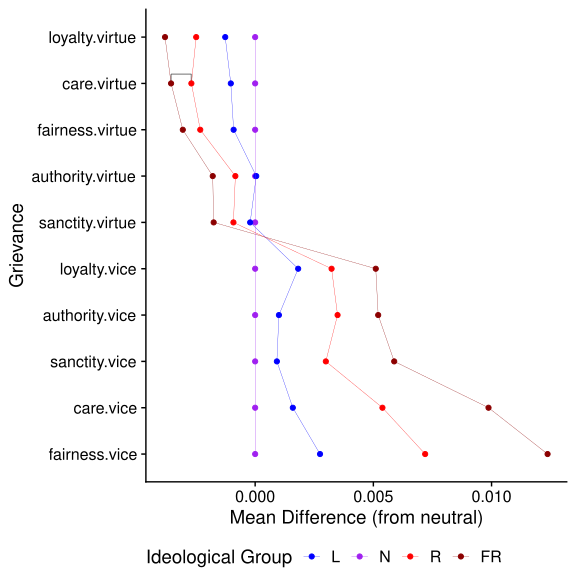
| Dataset | Users | Posts | Country |
| --- | --- | --- | --- |
| #Qanda | 103,074 | 768,808 | AUS |
| #Ausvotes | 273,874 | 5,033,982 | AUS |
| #SocialSense | 49,442 | 358,292 | AUS |
| Riot | 574,281 | 1,067,794 | US |
| Parler | 120,048 | 603,820 | US |

### Psychosocial profiles of the Ideological Groups

Large-scale profiling of entire online populations gives us significant insights into the characteristics of online populations. We apply our inferred ideological labels of online users in two critical ways:

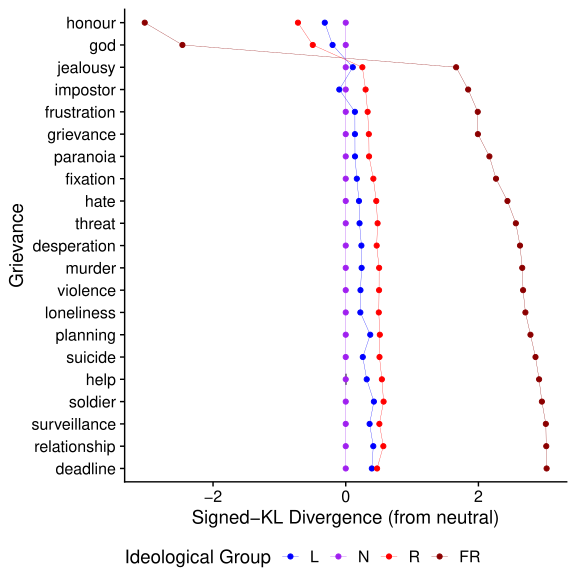
1. We derive a method for distinguishing the left and the right in terms of their moral language.
2. We derive a method for distinguishing moderate and extreme ideologies in terms of their grievance language.

**Distinguishing Left and Right**: We utilize the FrameAxis ([Mokhberian et al. 2020](#ref-mokhberian2020moral)) methodology to metricize each user’s association with each of the five Moral Foundations ([Graham et al. 2013](#ref-graham2013moral)) in terms of their vice and virtue axes. Given the measures of the Moral Foundations in the user language, we can start to detect in what way the left and the right differ. To do this, we find each ideological group’s mean vice and virtue scores and compare these to the neutral group. Figure @ref(fig:mft) shows the outcome of this analysis on the SocialSense dataset.



We see that the left prefers the language of virtue, while the right prefers the language of vice. This trend is largely consistent across all datasets.

**Distinguishing Moderates and Extremes**: We similarly generate measures via the Grievance Dictionary ([Van der Vegt et al. 2021](#ref-van2021grievance)), a threat assessment tool designed to highlight potential threats through their language. Similar to the previous plot, we investigate the distribution of grievance scores for the ideological groups. However, here we measure the difference between distributions with the Signed KL-divergence (a measure of the difference in the location and shape of distributions). Figure @ref(fig:grievance) shows the results for the Ausvotes dataset.



We observe that the far-right’s usage of grievance language is significantly different from the moderate ideological groups. This adds evidence to the growing concern that members of the far-right may vent their frustration and participate in violent behavior.

## Conclusion

In this work, we build a fully automatic end-to-end ideology detection pipeline for left-right and far-right detection. Importantly, with the pipeline, we can show the differences between the left and right, and moderates and extremes in terms of psychosocial language, across a range of diverse datasets.

### References

“Director-General’s Annual Threat Assessment.” 2021. *ASIO*. <https://www.asio.gov.au/resources/speeches-and-statements/director-generals-annual-threat-assessment-2021>.

Gillespie, Eden, and Michael McGowan. 2022. “Queensland Shooting: Gareth and Stacey Train Published YouTube Video After Killing Police Officers.” <https://www.theguardian.com/australia-news/2022/dec/16/queensland-shooting-gareth-and-stacey-train-youtube-video-published-after-killing-police>; The Guardian.

Graham, Jesse, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P Wojcik, and Peter H Ditto. 2013. “Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism.” In *Advances in Experimental Social Psychology*, 47:55–130. Elsevier.

Karp, Paul. 2020. “Asio Reveals up to 40% of Its Counter-Terrorism Cases Involve Far-Right Violent Extremism.” <https://www.theguardian.com/australia-news/2020/sep/22/asio-reveals-up-to-40-of-its-counter-terrorism-cases-involve-far-right-violent-extremism>; The Guardian.

Mokhberian, Negar, Andrés Abeliuk, Patrick Cummings, and Kristina Lerman. 2020. “Moral Framing and Ideological Bias of News.” In *International Conference on Social Informatics*, 206–19. Springer.

Newman, Nic, Richard Fletcher, Antonis Kalogeropoulos, DAL Levy, and Rasmus Kleis Nielsen. 2019. “Reuters Institute Digital News Report 2018. Reuters Institute for the Study of Journalism.” Oxford.

Sides, All. 2018. “Media Bias Ratings.” *Allsides. Com*. <https://www.allsides.com/media-bias/ratings>.

Van der Vegt, Isabelle, Maximilian Mozes, Bennett Kleinberg, and Paul Gill. 2021. “The Grievance Dictionary: Understanding Threatening Language Use.” *Behavior Research Methods* 53 (5): 2105–19.

Wang, Chi, Qingyun Wu, Markus Weimer, and Erkang Zhu. 2021. “FLAML: A Fast and Lightweight Automl Library.” *Proceedings of Machine Learning and Systems* 3: 434–47.