

DATA STATEMENT FOR DRAX (Direct and Reported Aporophobia on X)

A. Curation Rationale

Aporophobia has been defined as the "rejection, aversion, fear and contempt for the poor" (Cortina, 2022). It increases the burden of poverty, impacting the well-being of this vulnerable group, and constitutes an obstacle to poverty mitigation. When the poor are blamed for their situation and considered undeserving of help, it is harder for policy makers to approve and implement poverty-mitigation strategies (Arneson, 1997; Everatt, 2009; Nunn and Biressi, 2009). In this study, we identify and characterize aporophobia by analyzing how it is expressed through language. We consider both personal views and discussions on the views of others. For this, we collect and annotate English texts from various regions of the world that refer to poor people. This provides us with a better understanding of the diversity of commonly expressed beliefs and behaviors regarding the poor.

The DRAX dataset contains 1,816 tweets referring to poor people. Every tweet is manually annotated for one of the three categories: (1) '**Direct Aporophobia**', defined as text expressing the speaker's own aporophobic views, (2) '**Reporting Aporophobia**', defined as text stating or criticizing the aporophobic views and behaviors of others, or (3) '**None**' (none of the above). There are 520 (29%) instances labeled as 'Direct Aporophobia', 723 (40%) instances labeled as 'Reporting Aporophobia', and 573 (32%) instances labeled as 'None'.

The tweets were collected through Twitter API between 25 August 2022 and 23 November 2022 using the following query terms: *the poor* (used as a noun as opposed to an adjective, as in 'the poor performance'), *poor people*, *poor ppl*, *poor folks*, *poor families*, *homeless*, *on welfare*, *welfare recipients*, *low-income*, *underprivileged*, *disadvantaged*, *lower class*. (Further details on query term selection and tweet pre-processing are available in the paper.) By using tweet location, where available, or user location field, we grouped tweets into the following six regions: North America, Europe, Africa, South Asia, Oceania, and Other. We applied unsupervised topic modeling, using BERTopic, on the collected tweets and selected 15 topics highly relevant to the concept of aporophobia. Then, tweets from each of the 15 topics were randomly sampled to satisfy the following two conditions: (1) uniform distribution by region (equal amounts of tweets are sampled from each of the six geographical regions), and (2) uniform temporal distribution (equal amounts of tweets are sampled from each of the three months of the collected data). Finally, the selected tweets were manually annotated by three experts (authors of the paper). Each tweet was independently annotated by two annotators and the disagreements were resolved through discussions.

B. Language Variety

The data was collected via Twitter API with the language option set for English; therefore, any variety of English recognized by the Twitter language identification tool as English can be present. We specifically selected tweets with the user location in North America, Europe, Africa, South Asia, Oceania, as well as tweets with unknown location.

C. Speaker Demographic

No direct speakers' demographic information is available. Location information is available for some of the tweets; these tweets originate from North America, Europe, Africa, South Asia, and Oceania.

According to [Statista](#), Twitter users worldwide tend to be [male](#), [between the ages of 18 and 49](#). The United States of America has the most users. [According to Pew Research Center](#), in the US, Twitter users are younger, more highly educated and have higher income than the general public. However, the DRAX corpus was collected using specific query terms and restricted to English-language tweets. Thus, its user demographics might differ from the general Twitter demographics.

D. Annotator Demographic

The three annotators of the DRAX dataset are authors of this paper. Two of them identify as females, and one as male. The ages vary from 20s to 40s. All three have received higher education at Western institutions, but have different cultural backgrounds. They have extensive knowledge on social biases, aporophobia, and NLP.

E. Speech Situation

The DRAX dataset was collected between 25 August 2022 and 23 November 2022. The tweets mostly represent informal, spontaneous, asynchronous written language. The intended audience is friends and followers of the user or the general Twitter audience. Each tweet is limited to 280 characters.

F. Text Characteristics

The tweets are sampled from 15 topics listed in the paper. Topics with the highest proportion of 'Direct' aporophobic statements are those referring to drug addiction and mental health issues, immigrants and refugees. Other topics with a high proportion of tweets in the 'Direct' category refer to crime, homeless encampments, smell, alcohol addiction, and fear. Such messages often stereotype poor people, and especially the homeless group, as substance addicts and criminals, or express the general attitudes of fear and contempt toward the group. Aporophobic texts in topic 10 communicate the views of rejecting immigrants and refugees since they do not bring any resources and depend on the state's support. Topics with a high rate of 'Reporting Aporophobia' refer to racism, crime, hatred, the military, laws and courts, laws and regulations, and blaming the poor. Messages in these topics often criticize the governments and people in power for taking advantage and discriminating against poor people through unfair enforcement of laws and regulations, and blaming all social and economic issues on the lower socio-economic class. Black communities are seen as the most targeted since race-based discrimination results in both social and economic disadvantages.

Based on the qualitative analysis of the DRAX dataset, the conceptual framework for the nature of prejudices, bias and discrimination (Allport, 1954; Kahneman, 2011; Taylor, 1931; Honneth,

1996; Blodgett et al., 2020; Fuchs, 2018) and the concept of aporophobia (Cortina, 2022; Comim et al., 2020), we devise a taxonomy of aporophobic actions expressed through speech:



G. Recording Quality N/A

H. Other N/A

I. Provenance Appendix N/A