

Probabilities of predicting gendered nouns given attributes

probability

gender
female
male

1.0
0.8
0.6
0.4
0.2
0.0

Anger

Dangerousness

Fear

Coercion

Segregation

Avoidance

Help

Pity

Blame

stigma_category

