

Lab 1 Follow-up Discussion

The paper "Deep neural networks are more accurate than humans at detecting sexual orientation from facial images" ([paywalled peer-reviewed article](#), and [freely available preprint](#)) makes some big claims about machine learning. Are they justified?

As background, this article was published online in a preprint archive (a public online repository for research to publish findings prior to the article being submitted for peer review) in the summer of 2017. It was published in the Journal of Personality and Social Psychology in February 2018. The final article is available only via subscription through the journal website, although the preprint provides an opportunity to review the paper in its entirety.

In early 2018, Simpson et al. published a [critique of the paper](#). We will use this critique and the preprint as jumping off points for a discussion about the paper.

Exercise 1 The authors state: "We obtained facial images from public profiles posted on a U.S. dating website. We recorded 130,741 images of 36,630 men and 170,360 images of 38,593 women between the ages of 18 and 40, who reported their location as the U.S."

In [Simpson et al's critique](#), they write "[these results] may well be telling us more about the samples than about the general population they are presumed to represent" (p. 4).

Generate some hypotheses about why or why not the samples of photos used in this analysis provide results that are generalizable to contexts outside of this study. Personally, how big of a concern do you think generalizability is for this study?

Exercise 2 The first sentence of the abstract of [the original article](#) reads "We show that faces contain much more information about sexual orientation than can be perceived and interpreted by the human brain." What data do they present to back up this claim, and do you find it convincing? If not, what additional data would you like to see?

Exercise 3 Using the equations given in the critique on page 2 (section 2), calculate the probability that an individual is gay given they have been classified as gay if the population prevalence is 4% ($p=0.04$), the probability that a gay person is correctly classified as gay is 100% ($\alpha = 1$), and the probability that a straight person is correctly classified as straight is 90% ($\beta = 0.9$).

In the critique, they show that the probability of being gay given an individual was classified as gay is fairly low (only 27%). Implying that almost 3/4 of the time someone was classified as gay they were not actually gay. How does the result you calculate above compare with the scenario presented in the critique? What is the misclassification rate, when $\alpha = 1$? Why does improving our classification of gay people not improve our misclassification rate more? What do we need to improve to change our misclassification rate more meaningfully?

Exercise 4 Overall, which arguments for or against the conclusions of this paper do you personally find most compelling or interesting?

Are there lessons, particularly from the criticisms on generalizability, that are relevant to the data collection you did for your Lab 1 project?