DR. DEANNA NEEDELL, UBC / UCLA

What we didn't choose: multiplicity of predictions, explanations, and power

We routinely attach a single number to a person, a predicted risk, an explanation for that prediction, a measure of a voter's power, and treat it as determined, when in fact it is partly an accident of something we never chose. I will trace this through three settings. In observational multiplicity (joint with Erin George and Berk Ustun), a classifier is trained on binary labels that are themselves a single random draw from a Bernoulli distribution; had the labels come out differently we would fit a different, equally valid model with different predictions. We define the resulting regret, the variance of an individual's prediction across label re-draws, give a closed form for logistic regression, estimate it at the scale of thirteen million loans, and show that its burden falls unevenly on those most at risk. In explainability multiplicity, we ask whether the explanations we hand people, such as SHAP values, are as arbitrary as the predictions, and quantify their regret under the same label randomness. Finally, in voting power, the Banzhaf index measures how often a voter is decisive; relaxing its classical assumptions reveals that power is equally contingent, since a player cannot raise their own power, only diminish that of others, and natural equilibria exist but are unstable. Across all three, the moral is the same: the quantities we compute about people are not unique. They depend on labels, models, and assumptions we did not get to pick, and that contingency is distributed unevenly, sometimes strategically.