WORKSHOP

COLLECTIVE DECISION-MAKING AND DEMOCRATIC INSTITUTIONS

7-8 JULY 2023

Algorithmic Fairness and Human Discretion

Arna Wömmel
(Universität Hamburg)

Abstract

Machine-learning algorithms are increasingly used to assist humans in high-stakes decision-making. For example, loan officers apply algorithmic credit scores to inform lending decisions, HR managers use data-driven predictions in selecting applicants, and judges turn to recidivism risk tools when setting bail. Despite their pervasiveness, there are growing concerns that such predictive tools may discriminate against certain groups, which has led to numerous efforts to exclude information about protected group membership (e.g., race, gender) from input data. While, technically, such interventions can increase overall fairness levels, there is little evidence on how human decision-makers, who take these predictions as input, ultimately react to them. Do they consider the elimination of protected characteristics in algorithmic predictions when making decisions about others? To address this question, I conduct a lab experiment in which subjects predict the performance of others in a quantitative task. They receive (i) an algorithmic performance prediction ('suggestion'), and (ii) information about the other participants' social identity ('profile'). The treatments vary between subjects in the level of algorithmic fairness, i.e. whether the prediction includes protected social identity variable(s) (e.g. gender) or not, which is communicated to the subjects. I explore how potential reactions to various fairness properties might be influenced by subjects' biased beliefs about differences in performance levels across protected groups.