Human Resources, Artificial Intelligence, & Bias

Nigel Guenole, PhD

What is do we mean by bias in AI

How have we addressed it before AI

How do we address it in the age of AI

The bias challenge for AI in HR

There is now a broad consensus that AI research is progressing steadily, and that its impact on society is likely to increase…. Because of the great potential of AI, it is important to research how to reap its benefits while avoiding potential pitfalls”

- Stephen Hawking

35%

of firms have the statistical capability to detect bias

Pulse check

At your organization, is the topic of bias in A.I for H.R:

a) unimportant?

b) important?

Something old, and something new.

- AI systems apply selection criteria at remarkable speed and at vast scale.
- Heightened need to systematically tackle bias.
- Approaches to addressing bias in talent management exist in assessment.
- I'll describe how we can integrate psychological methods with newer ML.

Lay perspectives on bias

- Familiar with idea from media, heuristics and biases, stereotypes, unconscious bias etc.
- Believe it's undesirable, struggle to say how it manifests in a *testable* way
- If pressed, say different selection rates e.g. more men than women receive offers
- Consistent with media, ML community, but *not* psychological science

The New York Times

- *X is the first hiring platform designed entirely around the psychology of decision-making that helps firms make recruitment decisions smart (more predictive of performance), fair (less biased) and easy.*

October 3, 2017

In I-O psychology fairness is a *social judgment*, bias is a *technical issue,* but in the ML community fairness and bias often used interchangeably as is the case here.

Because fairness is a social judgment, you will not ‘fix’ claims and beliefs about *fairness* with algorithms or statistics alone.

The New York Times

- *We test the algorithms to ensure that women and men (as well as people of different ethnic backgrounds) are getting similar scores, and if they aren't, we adjust the inputs until they are.*

October 3, 2017

HR has been concerned with whether or not the differences reflect *real* differences, if not there is bias. ML community implicitly treats observed differences as unreal differences and removes them. Consequences?

Impact validity trade-off

- Some of the most predictive selection methods show adverse impact.
- ML and Psychology both observe trade-off between diversity / optimal prediction.
- Be cautious if you hear that you can eliminate differences *and* improve prediction.

I-O Psychology on bias: Two fundamental issues

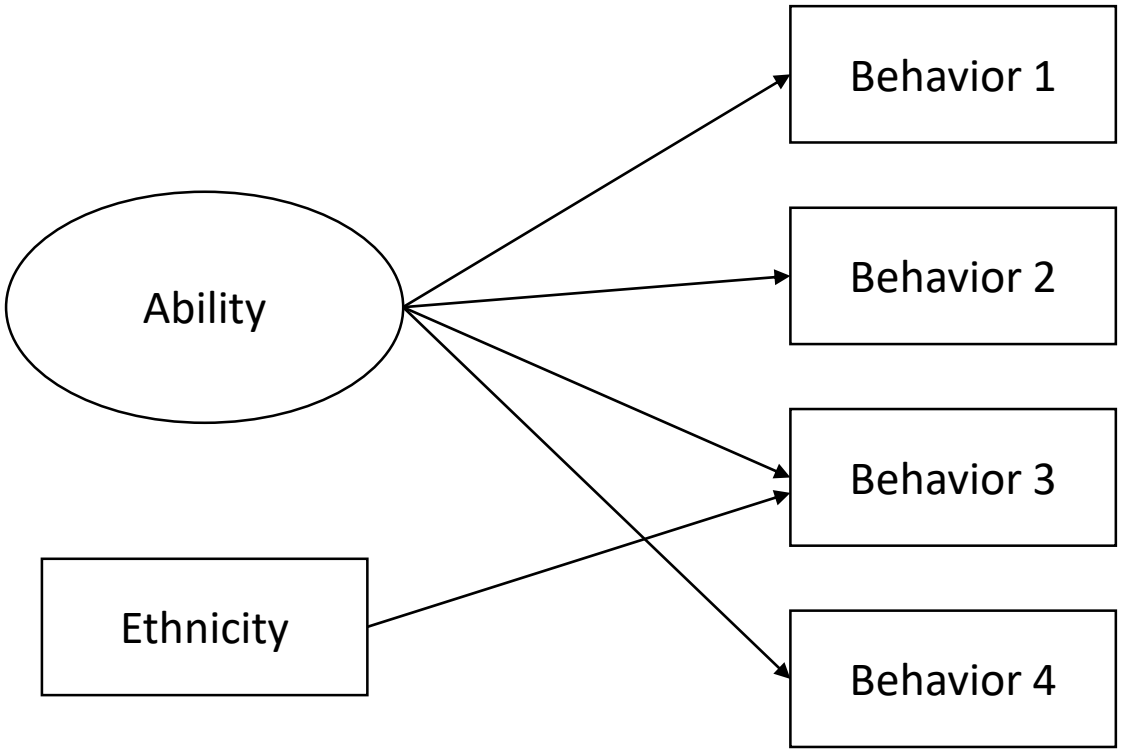
- Measurement bias

Can a randomly chosen and equally capable man / woman expect to get the same score?

- Relational bias

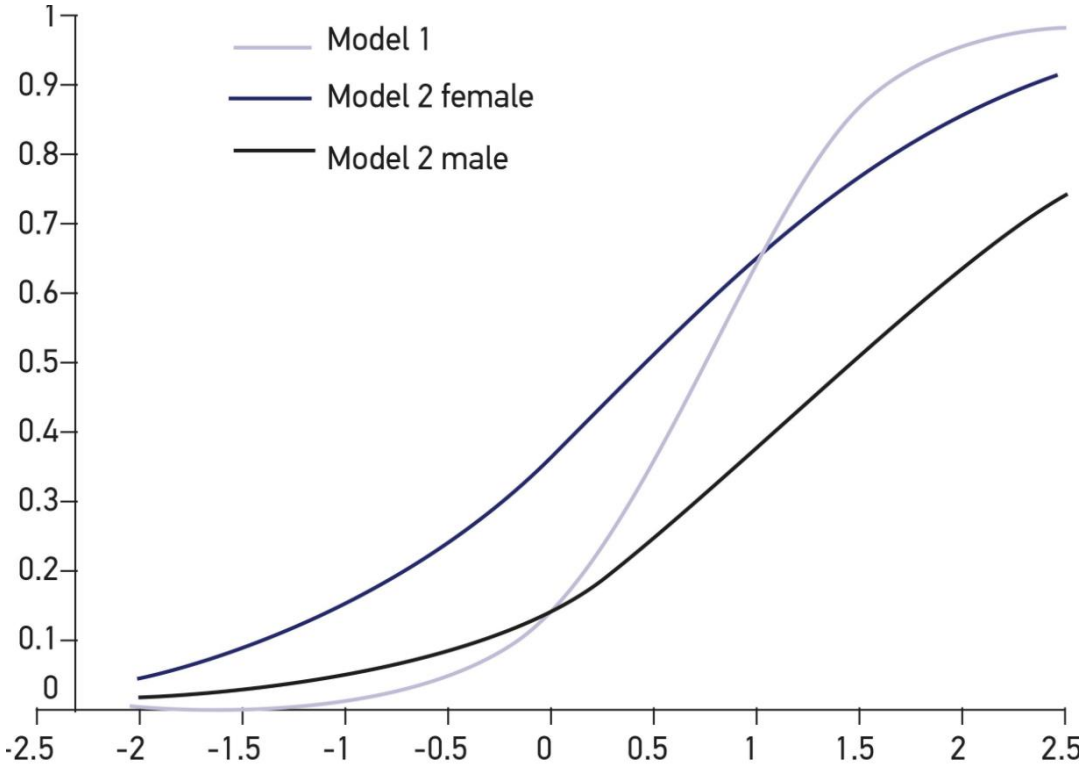
Do male scores predict male performance as well as females scores predict female performance?

Measurement bias / invariance



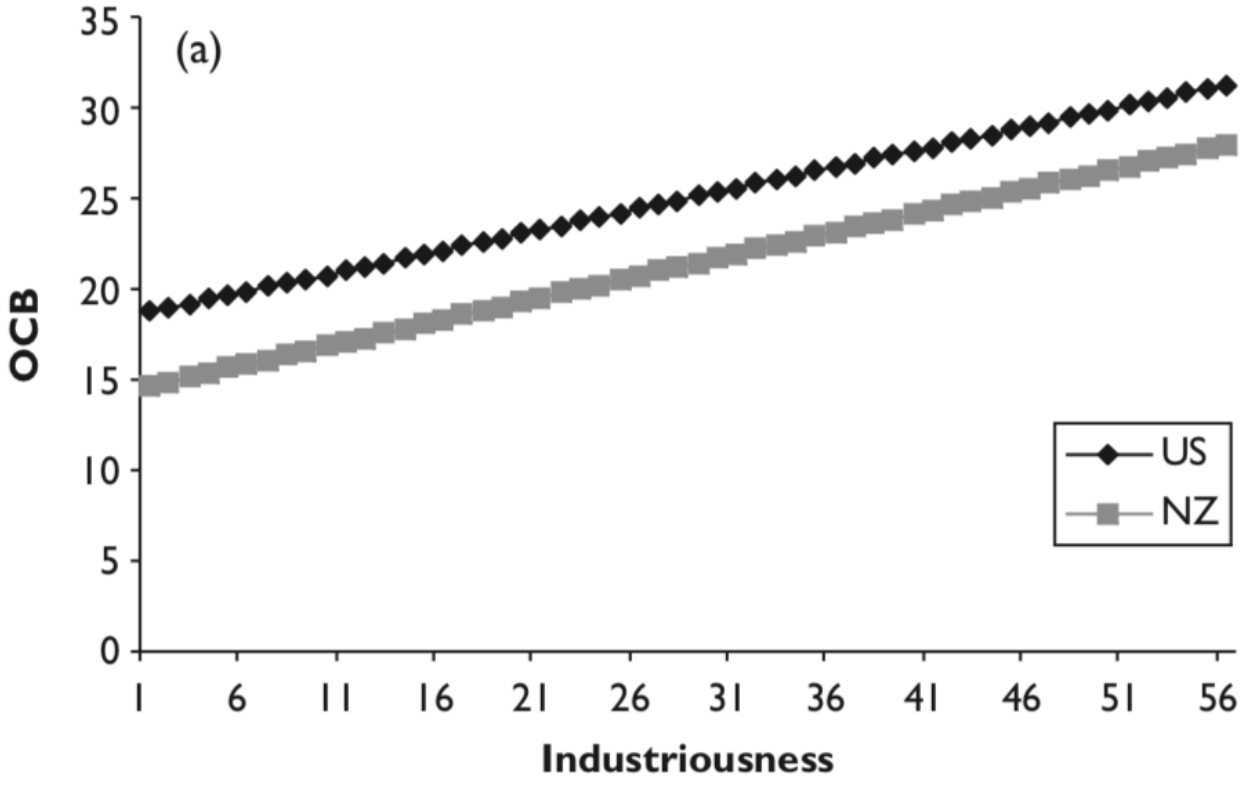
Show the model fits and ethnicity, and ability/ethnicity interaction (not shown) does not predict item/task scores

Measurement bias / invariance



Alternatively, compare item characteristic curve parameters, or area between the curves.

Relational bias



Stark, Chernyshenko, Guenole

I-O Psychology on Impact (‘bias’, ‘fairness’ to ML guys)

- Different selection rates may occur in the presence or absence of bias
- Tested with methods opposite, these *do not* test bias in IO terms

- *Impact* is checked with 4/5 rule, flip flop rule, Z-test (2SD rule), Fisher’s exact, Yates continuity correction, etc, *many free* packages in R

Different disciplines, different strategies

- Psychologists want to accurately measure people's skills, build tests to minimize *bias*, influence selection rates separately

- Machine learning community focuses on eliminating *impact* (which they call bias) directly in their algorithms

I-O psychology on minimizing impact (well understood)

- Change or weight predictors e.g. use narrow ability measures, use personality
- Broaden the criterion to include contextual performance or weight criterion
- Use score banding (fixed, sliding, criterion referenced)
- Pareto optimal selection methods (De Corte and colleagues)
- Affirmative action and equal employment opportunity

AI / ML on minimizing impact (be careful!)

- Alternative definitions e.g. Evaluate the 4/5 rule separately in selected and non-selected populations
- Pre-processing, e.g. remove protected attribute, label switching, forcing independence, re-weighting
- Training time constraints, e.g. optimize an alternative model that guarantees ‘fairness’
- Post processing, e.g. techniques like thresholding (categorizing scores)

AI / ML on minimizing impact (be careful!)

- Evaluate the 4/5 rule separately within selected versus non-selected populations
- Pre-processing, e.g. remove protected attribute, label switching, forcing independence
- Training time constraints, e.g. optimize an alternative model that guarantees ‘fairness’
- Post processing, e.g. techniques like thresholding (categorizing scores)

Reasons for the differences

Measurement error is often ignored in the machine learning literature because it is largely unimportant in traditional applications. The image algorithms that underlie the self-driving car, for example, are trained on data where it is easy to objectively label the presence of road boundaries, pedestrians, trees, and other obstacles. In the large image datasets on which vision algorithms are trained, the judgment of human observers who label images defines the truth that algorithm designers seek to predict.

- Mullanathan and Obermeyer, 2017, p479.

Measurement error has been the *exception* in ML, but its presence is the *rule* in HR, HR experts, psychologists and data scientists must work collaboratively

Bias is a tough problem. We don’t have all the answers. But we are better prepared than media reports lead us to believe.