Individual Fairness in Pipelines

Cynthia Dwork, Christina Ilvento, and **Meena Jagadeesan**

https://drops.dagstuhl.de/opus/volltexte/2020/12023/pdf/LIPIcs-FORC-2020-7.pdf

Presented at Dwork Reading Group (7/7/20)

(Some of these slides are from Christina Ilvento's FORC presentation.)

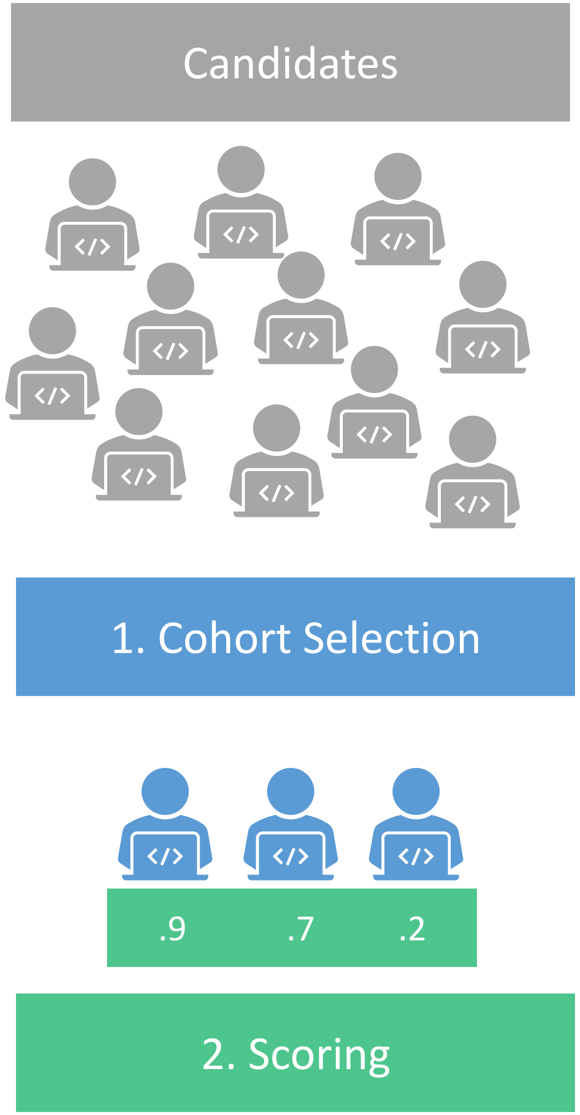
Cohort pipelines

Two-stage cohort pipeline: a *cohort selection* step followed by a *scoring* step within the chosen cohort

Examples:

- Hire team & promote *top performer*
- Screen a batch of resumes & interview *top candidates*

(Can also consider pipelines with many cohort selection/scoring steps in sequence.)



A motivating example-- employment

Majority group S; Minority group T

Cohort selection*:* hire every individual with the same probability.

- "Pack" high-potential $t \in T$ into same teams.
- Place all other hires on mixed skilled teams.

Scoring step*: s*core and promote according to *relative* performance.

$\Rightarrow$ Fewer high-potential $t \in T$ promoted than high-potential $s \in S$.

Fairness can degrade arbitrarily even in a 2-stage cohort pipeline.

The setup

- Universe U of individuals with similarity metric D
- Collection of "permissible" cohorts $\mathcal{C} \subseteq 2^U$
- Cohort selection mechanism A that chooses a cohort in $\mathcal{C}$
- Score function $f\colon \mathcal{C} \times U \to [0,1]$ for individuals within cohort context
- Pipeline $f \circ A$:
    1. Run A to select a cohort C $\in \mathcal{C}$.
    2. Score all individuals $u \in C$ according to $f(C, u)$.

Our goal: Ensure that the pipeline $f \circ A$ treats similar individuals similarly.

Fairness of each step in isolation

Starting point: *individual fairness* (Dwork et al. ‘12)

“Similar people should be treated similarly (w.r.t similarity metric D)”

- A is an *individually fair cohort selection mechanism* if:

  For all $u, v \in U, |\Pr[u \in C] - \Pr[v \in C]| \leq D(u, v)$

  (Dwork & Ilvento 2019).

- f: $\mathcal{C} \times U \rightarrow [0,1]$ is *intra-cohort individually fair* if:

  For all $C \in \mathcal{C}$ and $u, v \in$ C, $|f(C, u) - f(C, v)| \leq D(u, v)$.

Fair components not enough

Hiring (A) followed by **promotion** (f).

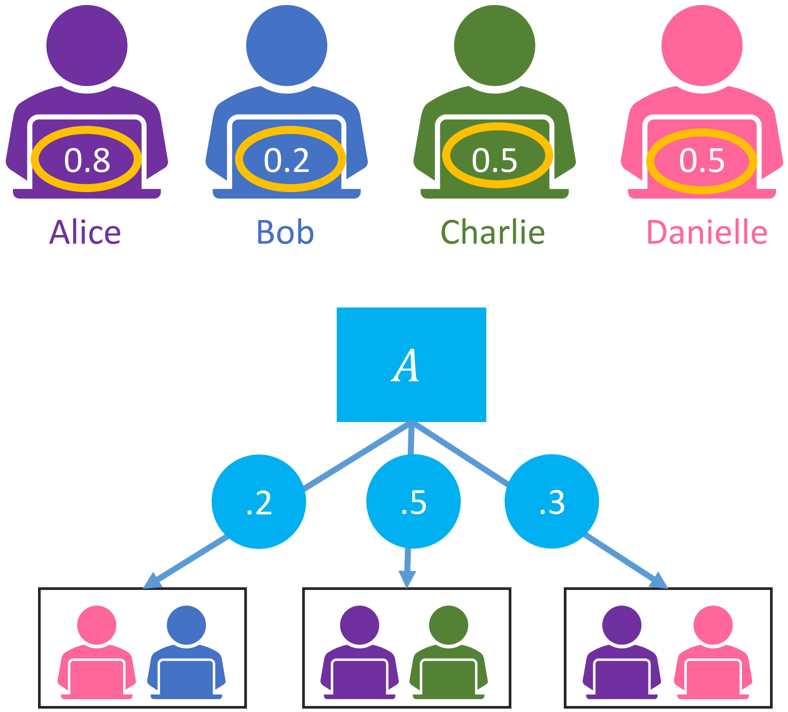
- Each candidate has a quality $q_i \in [0,1]$
- Minority group T; Majority group S
- Similarity metric given by $D(i,j) \coloneqq \left|q_i - q_j\right|$

A "packs" $\{\, t \in T \mid q_t \geq 0.8\}$ in same cohorts; balances other cohorts w.r.t. quality score.

f assigns weight proportional to quality so that $\sum_{u \in C} f(C,u) = 1$.

Intuition: High-quality $t \in T$ receive lower scores than high-quality $s \in S$.

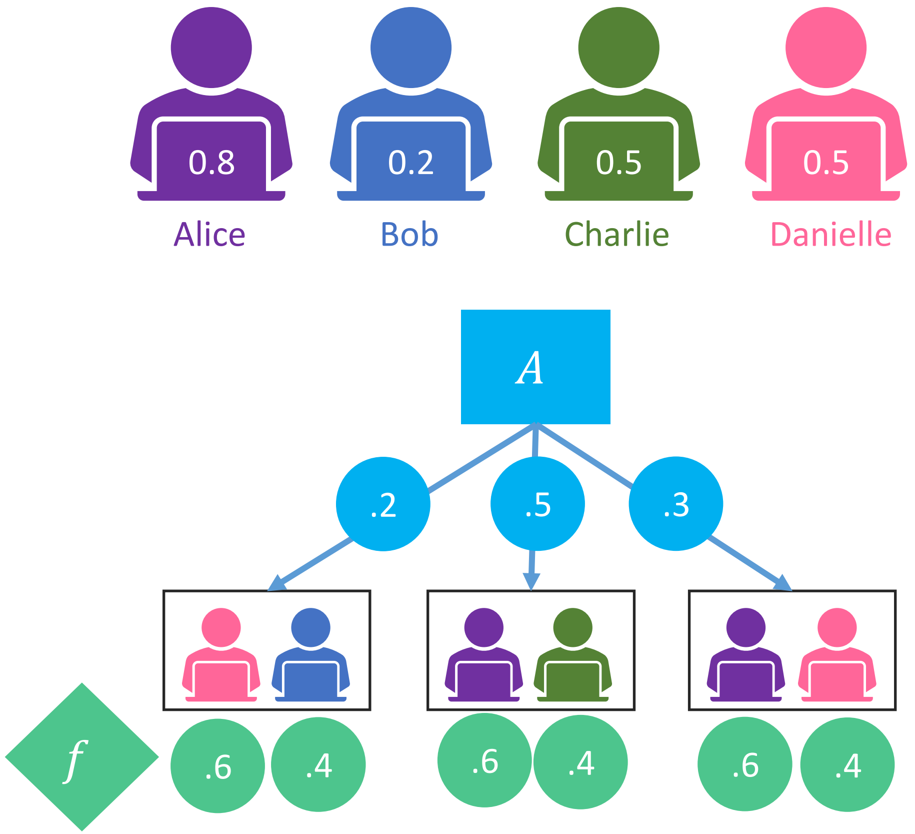
Example



	Alice	Bob	Charlie	Danielle
Hire	.8	.2	.5	.5

Example

- A is individually fair
- f is intra-cohort individually fair
- But the pipeline results in different promotion outcomes for equal individuals Charlie and Danielle



	Alice	Bob	Charlie	Danielle
Hire	.8	.2	.5	.5

Our contributions

1. Formalize definitions of *pipeline fairness* and extensions to a family of scoring functions.
2. Provide sufficient conditions for achieving pipeline fairness. These conditions allow for *flexible design* of the cohort selection mechanism and scoring functions by different bodies.
3. Construct explicit cohort selection mechanisms for two families of scoring functions. These mechanisms achieve pipeline fairness and are expressive.

DEFINITIONS

Pipeline fairness and robustness (Informal)

Notation: A cohort selection mechanism, f scoring function, D similarity metric

Definition: α-individually fairness for pipelines (Informal)

$f \circ A$ is α-individually fair if for all $u, v \in U$, $d\big([f \circ A](u), [f \circ A](v)\big) \leq \alpha D(u, v)$.

But A and f might be designed by separate bodies!

$\Rightarrow$ Not ideal to "lock" into a single scoring function f.

Instead, we require that f lives in some pre-specified family $\mathcal{F}$:

Definition: α-robustness for pipelines (Informal)

A is α-robust with respect to $\mathcal{F}$ if $f \circ A$ is α-individually fair for every $f \in \mathcal{F}$.

Defining pipeline individual fairness formally

To formally define pipeline individual fairness, we need to specify $[f \circ A](u)$ and d.

Outcome is either not selected or a score.

- Outcome space $O_{pipeline} = [0,1] \cup \perp$.
- $\Delta(O_{pipeline})$ is space of distributions over outcomes
  (so $[f \circ A](u) \in \Delta(O_{pipeline})$)

What metric d over $\Delta(O_{pipeline})$ captures fairness desiderata?

We design metrics d over $\Delta(O_{pipeline})$ in two steps:

1. Interpret $\Delta(O_{pipeline})$ as a distribution over $[0,1]$.
2. Select a metric over $\Delta([0,1])$.

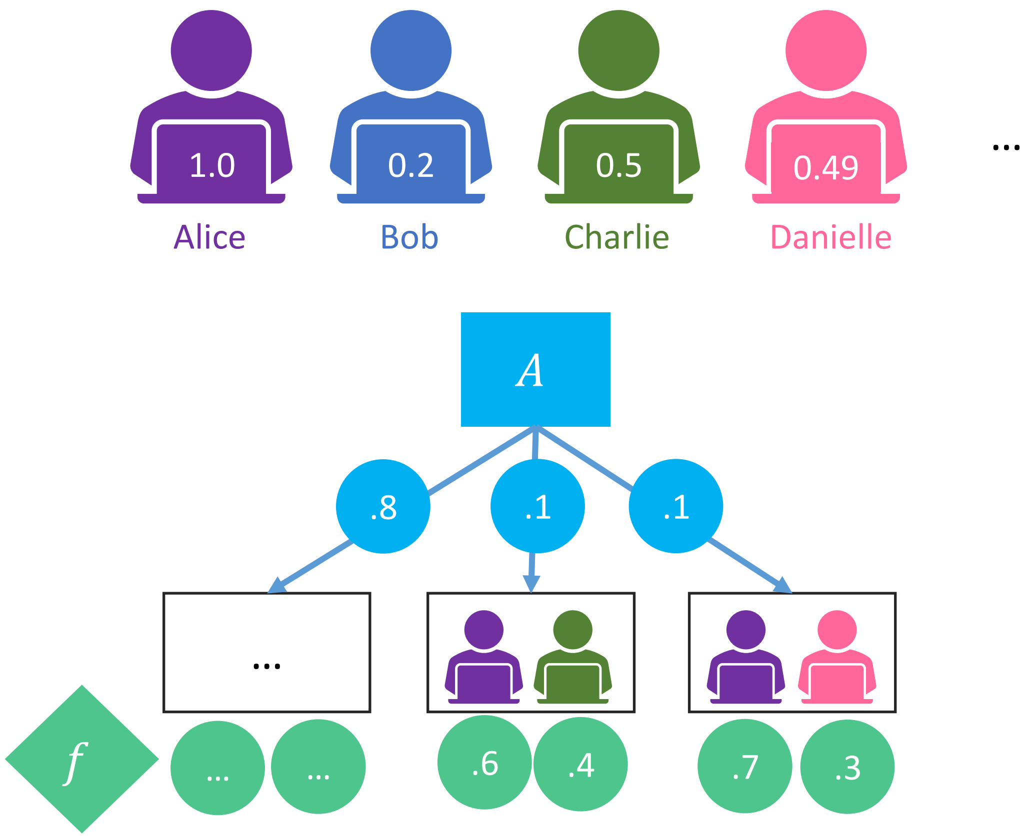
Step 1: Interpret the distribution

Two approaches to map $\Delta([0,1] \cup \perp)$ to $\Delta([0,1])$:

1. View not selected as a score of 0.
2. Consider distribution conditioned on being selected.

Example

- Equal hiring rate
- Difference in promotion respects metric
- Conditional probability of promotion 10x the metric distance!



	Alice	Bob	Charlie	Danielle
Hire	…	…	.1	.1
Promote	…	…	.04	.03

Conditional vs. Unconditional Interpretations

$\Pr_A[C]$ represents the probability over the randomness of A that A outputs the cohort C.

Unconditional Distribution

Treats "not selected" as score of 0. Places probability mass

- $1 - \sum_{C \in \mathcal{C}} \Pr_A[C] \Pr[f(C, u) \neq 0]$ on score 0.
- $\sum_{C \in \mathcal{C}} \Pr_A[C] \Pr[f(C, u) = s]$ on score s.

Conditional Distribution

Conditions on selection in the cohort. Places probability mass

- $\frac{\sum_{C \in \mathcal{C}} \Pr_A[C] \Pr[f(C,u)=s]}{\sum_{C \in \mathcal{C}, u \in C} \Pr_A[C]}$ on each score s.

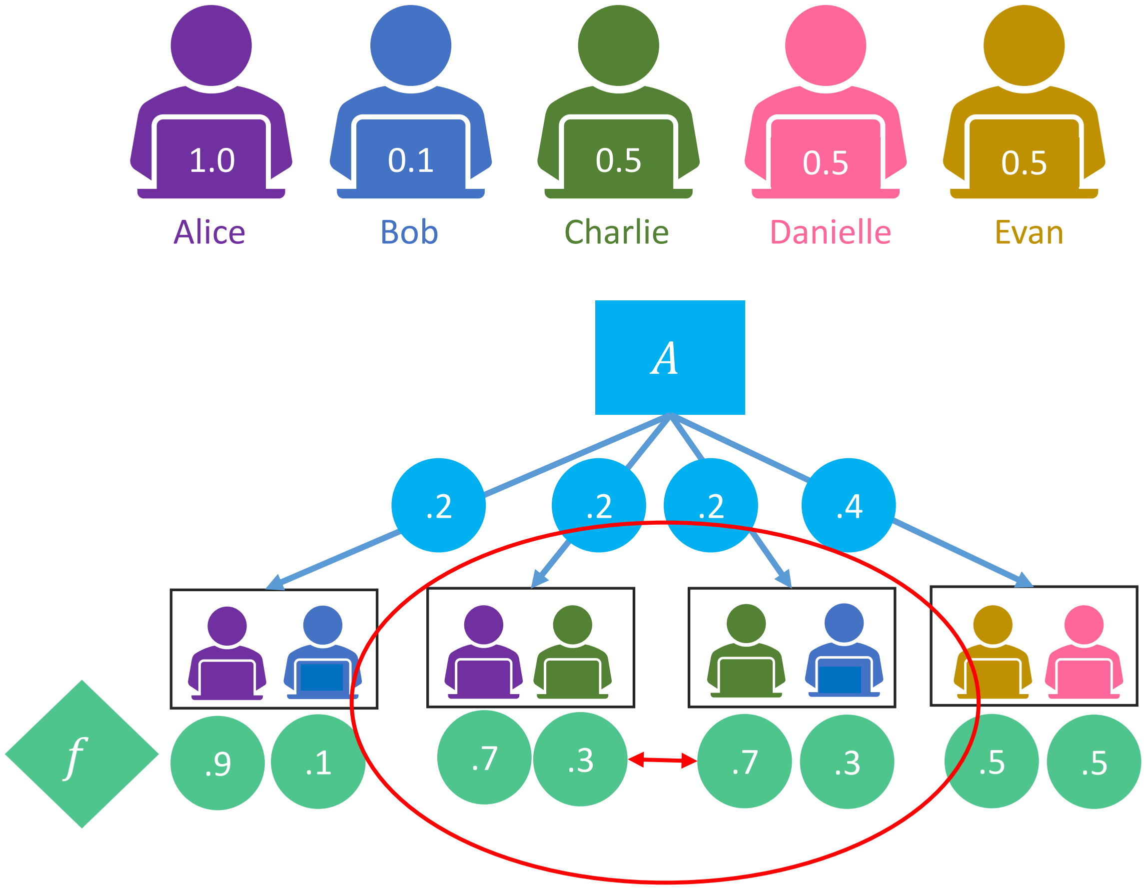
Step 2: Select distance metric over $\Delta([0,1])$

Two approaches to select distance metric over $\Delta([0,1])$:

1. Consider differences in *expected score.*
2. Account for uncertainty through *mass-moving distance.*

Example 3.

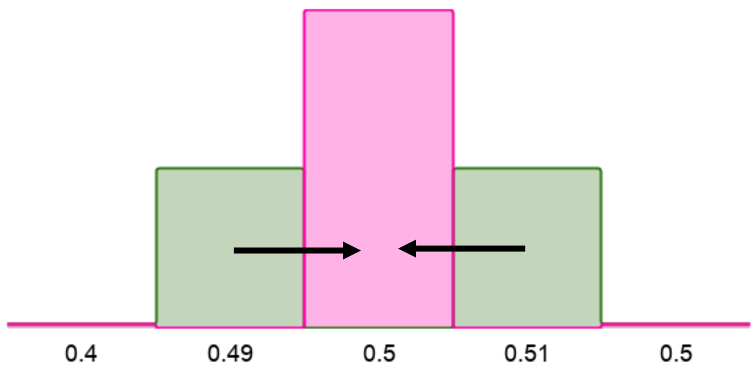
- Equal hiring rate
- Equal promotion rate
- But, compared with Danielle and Evan, Charlie has much higher **certainty** of promotion (or not)



	Alice	Bob	Charlie	Danielle	Evan
Hire	.4	.4	.4	.4	.4
Promote	.32	.08	.2	.2	.2

Choice for distance metric over $\Delta([0,1])$

- Expectation is often suitable
  - Simple, captures difference in binary outcomes or scores well
  - But hides **certainty**
- Total variation distance is a natural choice, but too strict:
  - e.g., Charlie has probability $0.5 + \varepsilon$ or $0.5 - \varepsilon$
- Mass-moving distance
  - Combines total variation distance with earth-mover's distance.
  - Similar individuals should receive similar distributions over close (but not necessarily identical) scores



Robustly fair pipelines

Define robustness w.r.t. different metrics over $\Delta(O_{pipeline})$:

	Expectation	Mass-moving distance
Conditional	$d^{cond,\mathbb{E}}$	$d^{cond,MMD}$
Unconditional	$d^{uncond,\mathbb{E}}$	$d^{uncond,MMD}$

Definition: Robust pipeline

Let $(d, D, A, \alpha, \mathcal{C}, \mathcal{F})$ be a pipeline consisting of a distance metric $d \in \{d^{cond,\mathbb{E}}, d^{cond,MMD}, d^{uncond,\mathbb{E}}, d^{uncond,MMD}\}$ over $\Delta(O_{pipeline})$, a set of permissible cohorts $\mathcal{C}$, a cohort selection mechanism A, and a set of scoring functions $\mathcal{F}$.

The pipeline is **robust** if $f \circ A$ is α-individually fair for all $f \in \mathcal{F}$.

Which metric is most appropriate is context-dependent.

CONDITIONS FOR SUCCESS

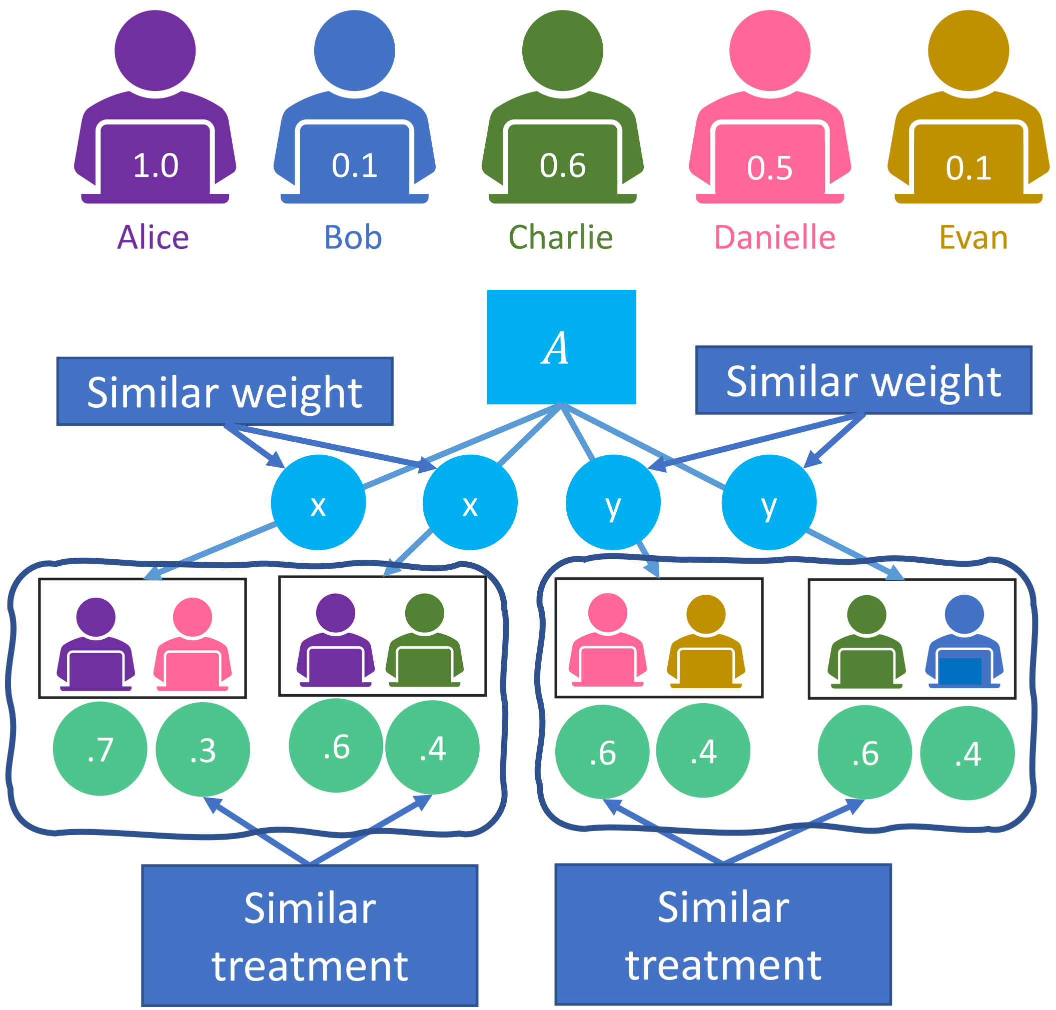
Constructing robustly fair pipelines

Our goal: Simple conditions on A that guarantee pipeline robustness with respect to $\mathcal{F}$**.**

The strength of the conditions on A heavily depends on $\mathcal{F}$.

- When $\mathcal{F}$ consists of functions that ignore the cohort, then A just needs to be individually fair.
- When $\mathcal{F}$ accounts for *relative performance,* conditions are stronger.

Key idea*: Similar individuals need to be assigned to similar distributions over cohort*s. Similarity of distributions is dependent on $\mathcal{F}$.



1.0
0.1
0.6
0.5
0.1
Alice
Bob
Charlie
Danielle
Evan
A
Similar weight
Similar weight
x
x
y
y
.7
.3
.6
.4
.6
.4
.6
.4
Similar treatment
Similar treatment

The policy: $\delta^{\mathcal{F}}$

Can summarize $\mathcal{F}$ as a distance function:

$$\delta^{\mathcal{F}}((C,u),(C',v)) \coloneqq \sup_{f\in\mathcal{F}} |f(C,u) - f(C',v)|$$

$\delta^{\mathcal{F}}$ is a simple form of **communication** between A and $\mathcal{F}$.

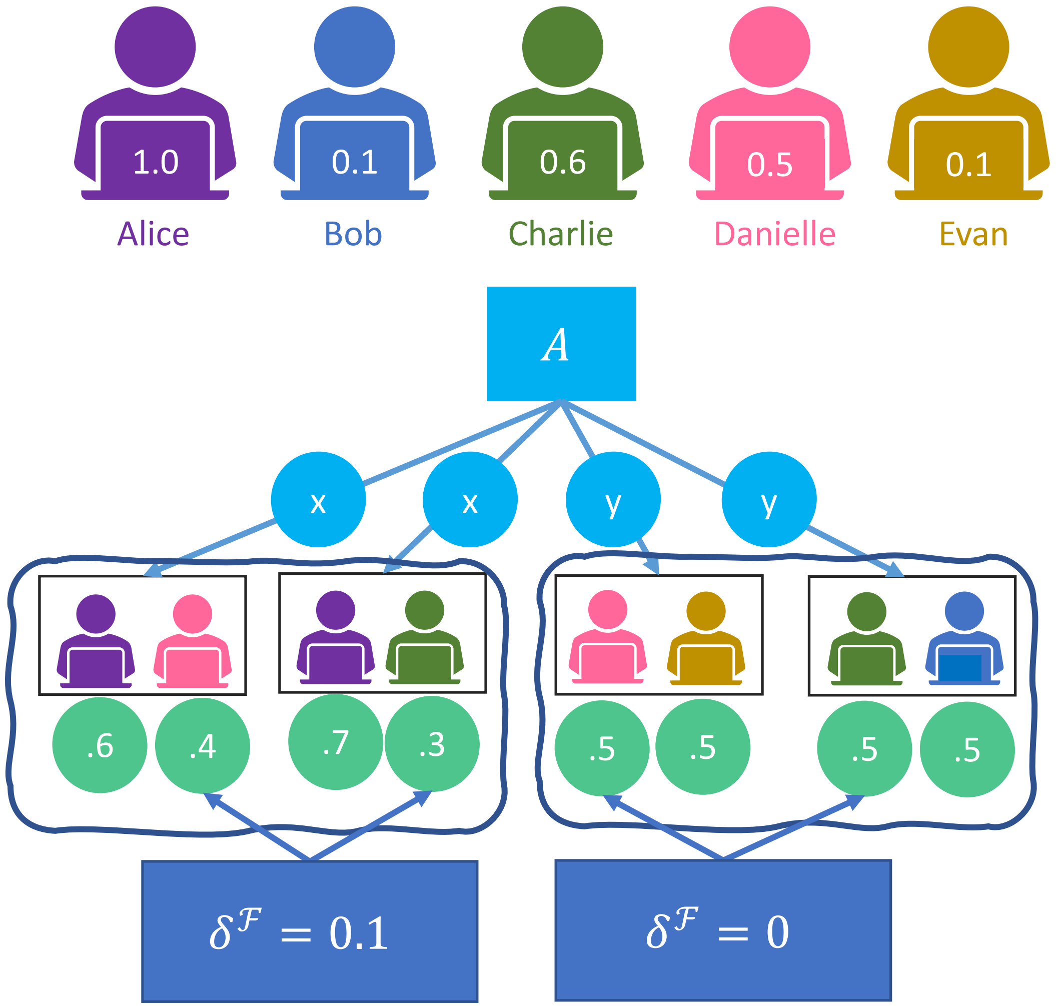
$\delta^{\mathcal{F}}$ can be thought of a **“policy”** agreed upon by both parties.

Conditions on A based on $\delta^{\mathcal{F}}$

Similarity of distributions over cohorts is dictated by $\delta^{\mathcal{F}}$.

For each pair $u, v \in U$:

1. Consider cohort contexts $\{(C, u) \mid u \in C, C \in \mathcal{C}\} \cup \{(C, v) \mid v \in C, C \in \mathcal{C}\}$
2. Group cohort contexts into clusters so that $\delta^{\mathcal{F}}\big((C, x), (C', y)\big) \leq D(u, v)$ within cluster.
3. Obtain distributions p_u and p_v over clusters.
4. **Requirement**: $TV(p_u, p_v)$ small



1.0
0.1
0.6
0.5
0.1
Alice
Bob
Charlie
Danielle
Evan
A
x
x
y
y
.6
.4
.7
.3
.5
.5
.5
.5
$\delta^{\mathcal{F}} = 0.1$
$\delta^{\mathcal{F}} = 0$

TWO SAMPLE CONSTRUCTIONS

Two policies $\delta^{\mathcal{F}}$

Individual Interchangeability

- Scoring function “stable” if a single individual is swapped in the cohort.

Quality-based Scoring

- Cohort contexts with similar “quality profiles” are treated similarly.

Individual interchangeability

- Scoring function "stable" if a single individual is swapped.

$$\delta^{\mathsf{int}}((C,u),(C',v)) = \begin{cases} \mathscr{D}(u,v) & \text{if } C = C' \\ \mathscr{D}(u,v) & \text{if } C' = (C \setminus \{u\}) \cup \{v\}\,. \\ 1 & \text{otherwise.} \end{cases}$$

- With $d^{uncond,MMD}$, any *monotonic* mechanism works.
  (If $\Pr[u \in C] \leq \Pr[v \in C]$, then $A(C \cup \{u\}) \leq A(C \cup \{v\})$.

- With $d^{cond,MMD}$, stronger requirements are necessary.
  We design a mechanism (Conditioning Mechanism) that works.

Conditioning Mechanism

Mechanism 4.7 (Conditioning Mechanism)**.** Given a target cohort size k, a universe U and a distance metric $\mathscr{D}$, initialize an empty set S. For each individual $u \in U$:

(1) Assign a weight $w(u)$ such that $|w(u) - w(v)| \leq \mathscr{D}(u,v)$, *i.e.*, the weights are individually fair.

(2) Draw from $\mathbb{1}_u \sim \text{Bern}(w(u))$, (*i.e.* flip a biased coin with weight $w(u)$). If $\mathbb{1}_u$, add u to S.

If $|S| \geq k$, return a uniformly random subset of S of size k.[19] Otherwise, repeat the mechanism.

- Mechanism is *expressive* (dissimilar people can be treated dissimilarly; people can have very different probabilities of being selected).
- But mechanism yields "unstructured cohorts".

Quality-based scoring

- Universe can be partitioned into “quality groups” where metric is closer within each quality group than between quality groups.
- Scores $f(C, u)$ determined by
  1. *Quality group membership* of u
  2. *Quality profile:* number of people from each quality group in C.

- **High-level idea**: Mechanism can select cohorts with “structure” based on quality profile. Flexibility in choosing individuals within each quality group.

Conclusion & Future Work

- Fairness degrades ungracefully in cohort pipelines.
- We proposed *pipeline individual fairness* where similar individuals have similar distributions over outcomes w.r.t. careful selections of a metric over distributions over outcomes. We proposed *pipeline robustness* that requires pipeline individual fairness for every scoring function in a family.
- We provided conditions under which pipeline fairness is achieved. We proposed a "*policy"* as a means of communication, and we proved a sufficient condition for success in terms of distributions over the appropriate clusters.
- We constructed explicit cohort selection mechanisms for two policies.

Future work: different metrics; formalize tradeoffs between $\delta^{\mathcal{F}}$ policy complexity and the expressiveness of cohort selection; ranking instead of scoring