Achieving Group Fairness with Social Welfare Optimization

John Hooker
Carnegie Mellon University

Joint work with

Violet (Xinying) Chen
Stevens Institute of Technology

Derek Leben
Carnegie Mellon University

INFORMS Optimization Society 2024

Group Parity Metrics

- Group parity metrics are widely used in AI
    - To assess whether demographic **groups** are treated **equally**
    - **Selection rates** are compared for:
        - Job interviews
        - University admissions
        - Mortgage loans, etc.
- A “**protected group**” is compared with the **rest** of the population
    - Groups defined by **race**, **gender**, **ethnicity**, **region**, etc.
    - Sometimes based on **legal** mandates
- We study parity metrics as an **assessment tool**
    - Rather than a selection criterion

Problems with Group Parity

- Group parity is intuitively appealing **at first**…
    - But is it really **fair**?
    - On closer examination, it raises many **problems**:

Problems with Group Parity

- Group parity is intuitively appealing **at first**…
    - But is it really **fair**?
    - On closer examination, it raises many **problems**:
- Failure to account for actual **welfare consequences**
    - Considers only **frequency** of selection
    - For example, rejection may be **more harmful** to a protected group

Problems with Group Parity

- Group parity is intuitively appealing **at first**…
    - But is it really **fair**?
    - On closer examination, it raises many **problems**:
- Failure to account for actual **welfare consequences**
    - Considers only **frequency** of selection
    - For example, rejection may be **more harmful** to a protected group
- Controversy over **which metric** is appropriate
    - **Many metrics** have been proposed
    - Some are mutually **incompatible**

Problems with Group Parity

- Group parity is intuitively appealing **at first**…
  - But is it really **fair**?
  - On closer examination, it raises many **problems**:
- Failure to account for actual **welfare consequences**
  - Considers only **frequency** of selection
  - For example, rejection may be **more harmful** to a protected group
- Controversy over **which metric** is appropriate
  - **Many metrics** have been proposed
  - Some are mutually **incompatible**
- Unclear how to **identify** protected groups
  - Groups often have **conflicting interests**
  - **No limit** to groups that may cry “unfair.”

Some Parity Metrics

- **Demographic parity**.
    - Same **fraction of group** is selected.

$$P(D|Z) = P(D|\neg Z)$$

Selected
Protected
Not protected

- **Equalized odds** (specifically, equality of opportunity)
    - Same fraction of **qualified** members of group is selected
    - Qualified = offered a job, repays mortgage, success in school.

$$P(D|Y, Z) = P(D|Y, \neg Z)$$

Qualified

- **Predictive rate parity**
    - Same fraction of **selected** members of a group are **qualified**

$$P(Y|D, Z) = P(Y|D, \neg Z)$$

Example: Parole Decisions

- **Objective: Select prisoners for parole**.
  - Based on AI-predicted recidivism rates.
  - Without discriminating against minority candidates
  - Northpointe (now Equivant) developed the COMPAS system for parole decisions.

Example: Parole Decisions

- **Objective: Select prisoners for parole**.
  - Based on AI-predicted recidivism rates.
  - Without discriminating against minority candidates
  - Northpointe (now Equivant) developed the COMPAS system for parole decisions.
- **Controversy**
  - ProPublica claimed that COMPAS is **unfair** because it fails to **equalize odds**.
    - Minority candidates must be **less likely to recidivate** to obtain parole.
  - Norrhpointe claimed that COMPAS is **fair** because it achieves **predictive rate parity**
    - Paroled minority and majority candidates have **equal recidivism rates**
  - **Which** parity metric is appropriate?

Fairness as Social Welfare

- Group fairness through population-wide social welfare
  - As measured by a **social welfare function**
  - Perhaps a **broader concept of distributive justice** can assess parity metrics and achieve fairness across multiple groups
    - while taking **welfare** into account.

Fairness as Social Welfare

- Group fairness through population-wide social welfare
  - As measured by a **social welfare function**
  - Perhaps a **broader concept of distributive justice** can assess parity metrics and achieve fairness across multiple groups
    - while taking **welfare** into account.
- Focus on **alpha fairness** as a social welfare function
  - Frequently used in engineering, etc.
  - Studied for over 70 years.
    - In particular, by 2 Nobel laureates (John Nash, J.C. Harsanyi).
  - Defended by axiomatic and bargaining arguments
    - *Axiomatic arguments:* Nash (1950), Lan, Kao & Chiang (2010,2011)
    - *Bargaining arguments:* Harsanyi (1977), Rubinstein (1982), Binmore, Rubinstein & Wolinksy (1986)

Alpha Fairness

- The **alpha fairness** social welfare function:

$$W_\alpha(\boldsymbol{u}) = \begin{cases} \frac{1}{1-\alpha} \sum_i u_i^{1-\alpha} & \text{for } \alpha \geq 0,\ \alpha \neq 1 \\ \sum_i \log(u_i) & \text{for } \alpha = 1 \end{cases}$$

where u_i is the utility allocated to individual i

  - **Utilitarian** when α = 0, **maximin** (Rawlsian) when $\alpha \rightarrow \infty$
  - **Proportional fairness** (Nash bargaining solution) when α = 1

- To achieve alpha fairness:

Maximize $W_\alpha(\boldsymbol{u})$ subject to resource constraints.

Alpha Fairness

- Alpha fair selection

Let $x_i = 1$ if individual i is selected, 0 otherwise.
Then $u_i = a_i x_i + b_i$, where $a_i =$ **selection benefit**
$b_i =$ base utility .

Now

$$W_\alpha(\boldsymbol{u}) = \begin{cases} \dfrac{1}{1-\alpha} \displaystyle\sum_i (a_i x_i + b_i)^{1-\alpha} & \text{for } \alpha \geq 0,\ \alpha \neq 1 \\ \displaystyle\sum_i \log(a_i x_i + b_i) & \text{for } \alpha = 1 \end{cases}$$

We want to maximize $W_\alpha(\boldsymbol{u})$ subject to $x_i \in \{0,1\}$ and

$$\sum_i x_i = m$$

Number of individuals selected

Alpha Fairness

- An algebraic trick leads to a solution algorithm

If $\alpha \neq 1$, we have

$$W_\alpha(\boldsymbol{u}) = \boxed{\frac{1}{1-\alpha}\sum_i b_i^{1-\alpha}} + \frac{1}{1-\alpha}\sum_i \Big((a_i x_i + b_i)^{1-\alpha} - b_i^{1-\alpha}\Big)$$

Constant term

Alpha Fairness

- An algebraic trick leads to a solution algorithm

If $\alpha \neq 1$, we have

$$W_\alpha(\boldsymbol{u}) = \frac{1}{1-\alpha}\sum_i b_i^{1-\alpha} + \boxed{\frac{1}{1-\alpha}\sum_i \left((a_i x_i + b_i)^{1-\alpha} - b_i^{1-\alpha}\right)}$$

So we can maximize

$$\sum_{i|x_i=1} \boxed{\frac{1}{1-\alpha}\left((a_i + b_i)^{1-\alpha} - b_i^{1-\alpha}\right)}$$

Alpha Fairness

- An algebraic trick leads to a solution algorithm

If $\alpha \neq 1$, we have

$$W_\alpha(\boldsymbol{u}) = \frac{1}{1-\alpha}\sum_i b_i^{1-\alpha} + \boxed{\frac{1}{1-\alpha}\sum_i \left((a_i x_i + b_i)^{1-\alpha} - b_i^{1-\alpha} \right)}$$

So we can maximize

$$\sum_{i|x_i=1} \boxed{\frac{1}{1-\alpha}\left((a_i + b_i)^{1-\alpha} - b_i^{1-\alpha} \right)} = \sum_{i|x_i=1} \boxed{\Delta_i(\alpha)}$$

Welfare differential of individual *I*
= net increase in social welfare that
results from selecting individual *i*

Alpha Fairness

- An algebraic trick leads to a solution algorithm

If $\alpha \neq 1$, we have

$$W_\alpha(\boldsymbol{u}) = \frac{1}{1-\alpha}\sum_i b_i^{1-\alpha} + \boxed{\frac{1}{1-\alpha}\sum_i \left((a_i x_i + b_i)^{1-\alpha} - b_i^{1-\alpha}\right)}$$

So we can maximize

$$\sum_{i|x_i=1} \boxed{\frac{1}{1-\alpha}\left((a_i + b_i)^{1-\alpha} - b_i^{1-\alpha}\right)} = \sum_{i|x_i=1} \boxed{\Delta_i(\alpha)}$$

Welfare differential of individual *I*
= net increase in social welfare that
results from selecting individual *i*

…by selecting the m individuals with the largest welfare differentials $\Delta_i(\alpha)$. Similarly if $\alpha = 1$.

Alpha Fairness Example

$\alpha = 0.7$, Select 9 individuals

Majority group

a_i	$\Delta_I(0.7)$
1.5	**0.750**
1.4	**0.708**
1.3	**0.665**
1.2	**0.621**
1.1	**0.577**
1.0	**0.531**
0.9	0.484
0.8	0.436
0.7	0.387
0.6	0.336

Protected group

a_i	$\Delta_I(0.7)$
0.2	0.187
0.4	0.354
0.6	**0.505**
0.8	**0.643**
1.0	**0.770**

Alpha Fairness Example

α = 0.7, Select 9 individuals

Majority group

a_i	$\Delta_i(0.7)$
1.5	**0.750**
1.4	**0.708**
1.3	**0.665**
1.2	**0.621**
1.1	**0.577**
1.0	**0.531**
0.9	0.484
0.8	0.436
0.7	0.387
0.6	0.336

Protected group

a_i	$\Delta_i(0.7)$
0.2	0.187
0.4	0.354
0.6	**0.505**
0.8	**0.643**
1.0	**0.770**

9 individuals with highest welfare differentials

a_i	$\Delta_i(0.7)$
1.0	**0.770**
1.5	**0.750**
1.4	**0.708**
1.3	**0.665**
0.8	**0.643**
1.2	**0.621**
1.1	**0.577**
1.0	**0.531**
0.6	**0.505**

Alpha Fairness Example

$\alpha = 0.7$, Select 9 individuals

- Alpha fairness ($\alpha = 0.7$) corresponds to demographic parity.
  - 6 of 10 majority individuals selected
  - 3 of 5 protected individuals selected
  - 60% of both groups

Welfare differential of individual I
= net increase in social welfare that
results from selecting individual i

9 individuals with highest welfare differentials

a_i	$\Delta_I(0.7)$
1.0	0.770
1.5	0.750
1.4	0.708
1.3	0.665
0.8	0.643
1.2	0.621
1.1	0.577
1.0	0.531
0.6	0.505

Alpha Fairness Example

α = 0.7, Select 9 individuals

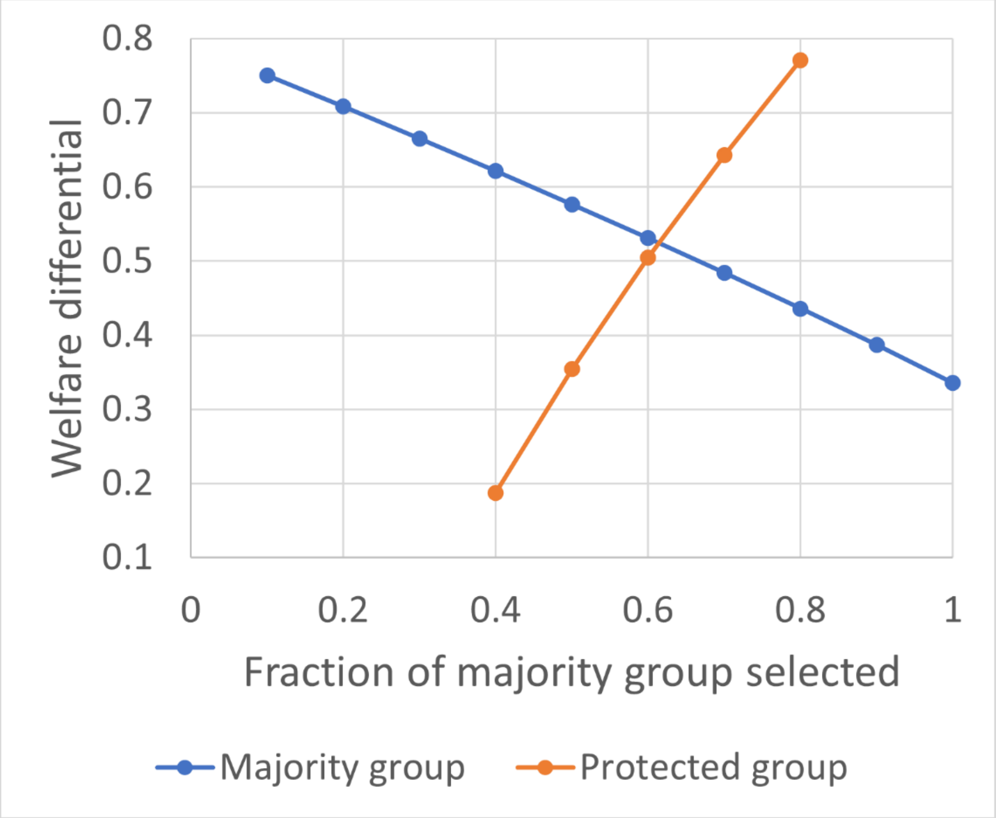
Majority group

a_i	$\Delta_i(0.7)$
1.5	**0.750**
1.4	**0.708**
1.3	**0.665**
1.2	**0.621**
1.1	**0.577**
1.0	**0.531**
0.9	0.484
0.8	0.436
0.7	0.387
0.6	0.336

Protected group

a_i	$\Delta_i(0.7)$
0.2	0.187
0.4	0.354
0.6	**0.505**
0.8	**0.643**
1.0	**0.770**

Graphical interpretation



Utility Model for 2 Groups

- We want a model that relates alpha fairness to the utility characteristics of the majority and projected groups.
  - …while reducing the number of utility parameters
  - Selection benefits uniformly distributed in each group
  - Base utility is constant in each group

Majority group

Selection benefits

$A_{\max}$ $A_{\min}$

Base utility = *B*

Protected group

Selection benefits

$a_{\max}$ $a_{\min}$

Base utility = *b*

Utility Model for 2 Groups

- We want a model that relates alpha fairness to the utility characteristics of the majority and projected groups.
  - …while reducing the number of utility parameters

Let S = fraction of majority group selected
s = fraction of protected group selected

Then the welfare differential of the last individual selected in the majority group is

$$\Delta_S(\alpha) = \begin{cases} \frac{1}{1-\alpha}\Big(\big((1-S)A_{\max} + SA_{\min} + B\big)^{1-\alpha} - B^{1-\alpha}\Big) & \text{if } \alpha \neq 1 \\ \log\big((1-S)A_{\max} + SA_{\min} + B\big) - \log(B) & \text{if } \alpha = 1 \end{cases}$$

and in the protected group is $\Delta'_s(\alpha)$, similarly defined.

Utility Model for 2 Groups

If β = fraction of population that is in the protected group
σ = fraction of population selected, then

$$(1-\beta)S + \beta s = \sigma,$$

which implies

$$s = s(S) = \frac{\sigma - (1-\beta)S}{\beta}$$

and...

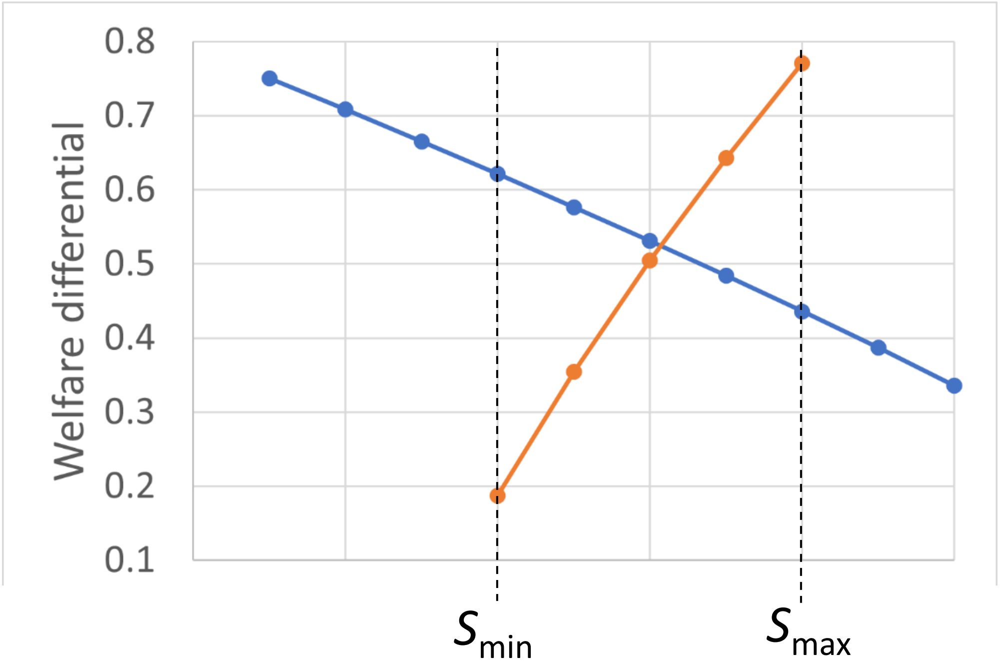
Utility Model for 2 Groups

If β = fraction of population that is in the protected group
σ = fraction of population selected, then

the min and max values of S are

$$S_{\min} = \max\left\{0,\ \frac{\sigma-\beta}{1-\beta}\right\},\quad S_{\max} = \min\left\{1,\ \frac{\sigma}{1-\beta}\right\}$$

$\sigma = 0.6$

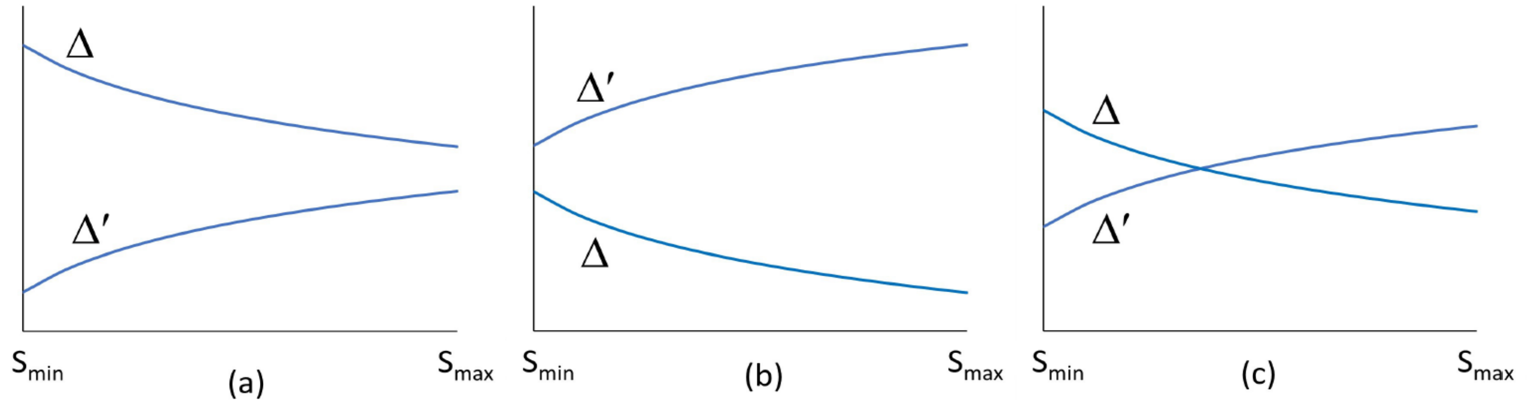


Utility Model for 2 Groups

Theorem. Selection rates (S, s) achieve alpha fairness for a given α if and only if $s = s(S)$ and

$$
\begin{cases}
(S,s) = \left(\min\left\{1, \dfrac{1}{1-\beta}\right\}, \dfrac{\sigma}{\beta}\left[1 - \min\left\{1, \dfrac{1-\beta}{\sigma}\right\}\right]\right) & \text{in case (a)} \\
(S,s) = \left(\dfrac{\sigma}{1-\beta}\left[1 - \min\left\{1, \dfrac{\beta}{\sigma}\right\}\right], \min\left\{1, \dfrac{\sigma}{\beta}\right\}\right) & \text{in case (b)} \\
\Delta_S(\alpha) = \Delta'_s(\alpha) & \text{in case (c)}
\end{cases}
$$

where the cases are



Alpha-fair Selection Rates

- Consider 3 qualitatively different utility scenarios…

	Scenario 1	Scenario 2	Scenario 3
Majority group	$A_{\min}$ 0.5 — 1.5 $A_{\max}$	$A_{\min}$ 0.5 — 0.8 $A_{\max}$	$A_{\min}$ 0.5 — 1.0 $A_{\max}$
Protected group	$a_{\min}$ 0.2 — 1.0 $a_{\max}$	$a_{\min}$ 0.2 — 1.0 $a_{\max}$	$a_{\min}$ –0.5 — 1.0 $A_{\max}$ $a_{\max}$

Protected group **benefits somewhat less** from selection

*For example, granting **job interviews***

Some protected individuals **benefit most**

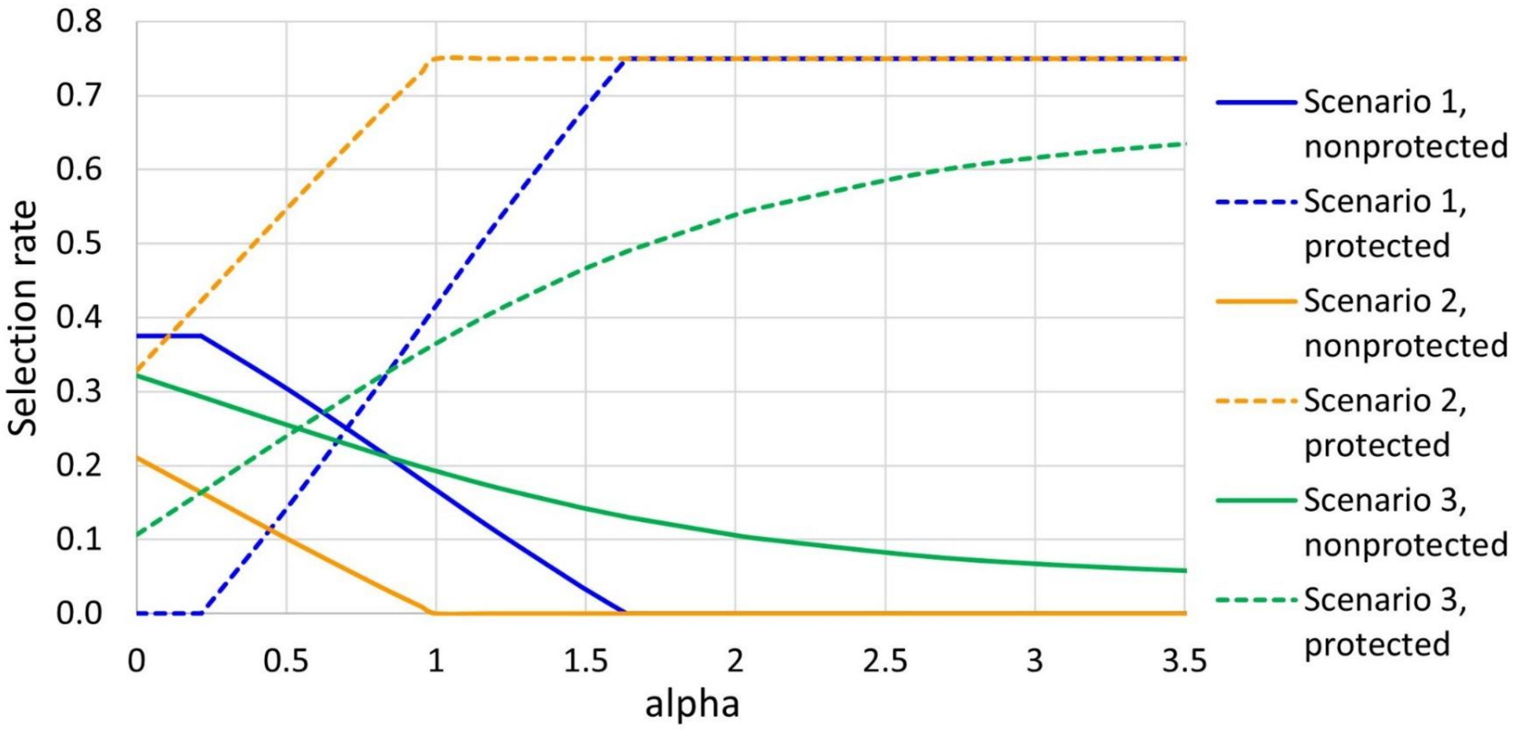
*For example, **admission** of talented individuals **to university***

Some protected individuals **harmed** by selection

*For example, **mortgage loans** with possible **foreclosure***

Alpha-fair Selection Rates

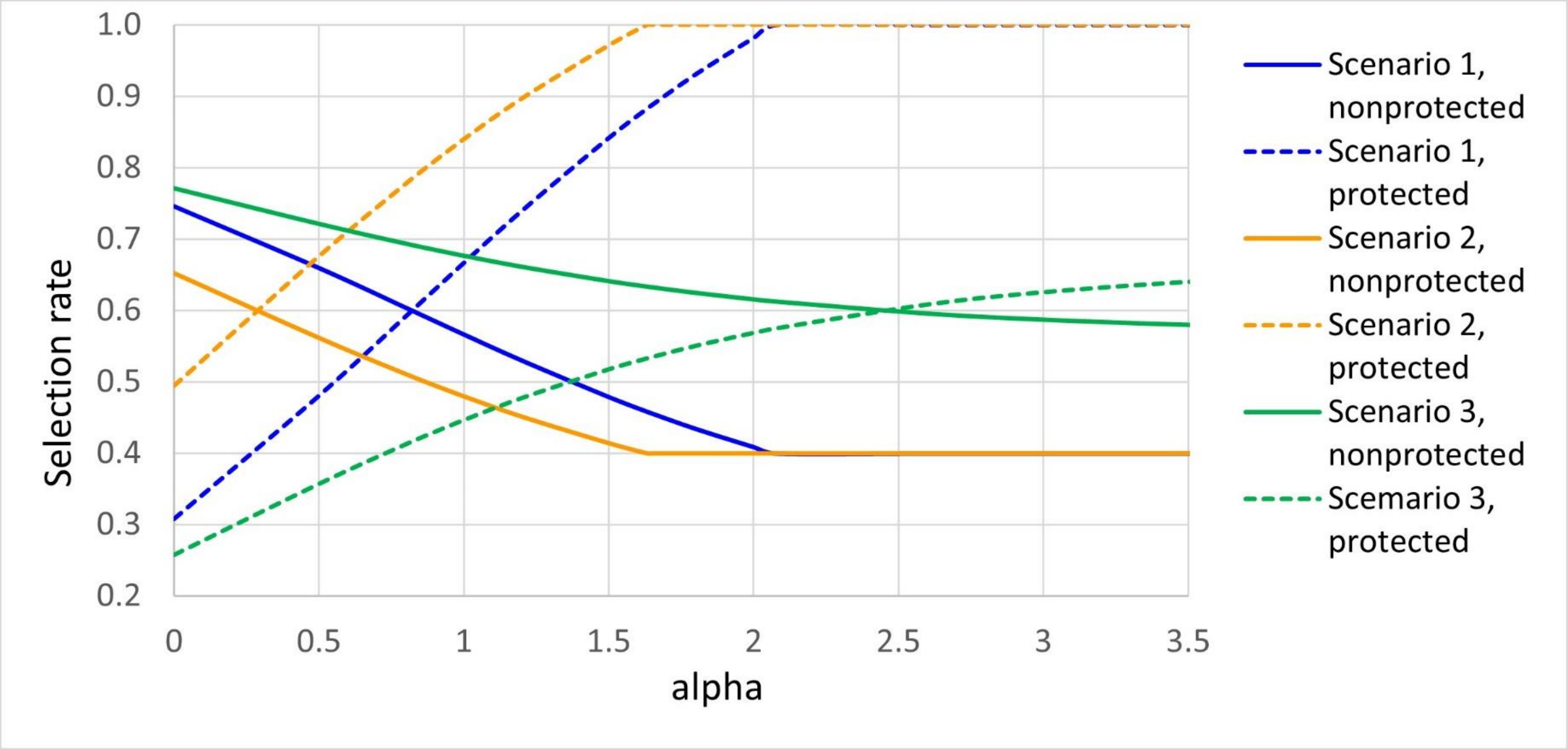
- Overall selection rate = 0.25



- Protected group has **lower** selection rates in Scenario 1 than in Scenario 2 due to **higher utility cost** of fairness in scenario 1.
- Protected group selection rate approaches 2/3 asymptotically because 1/3 of group is **harmed** by selection.

Alpha-fair Selection Rates

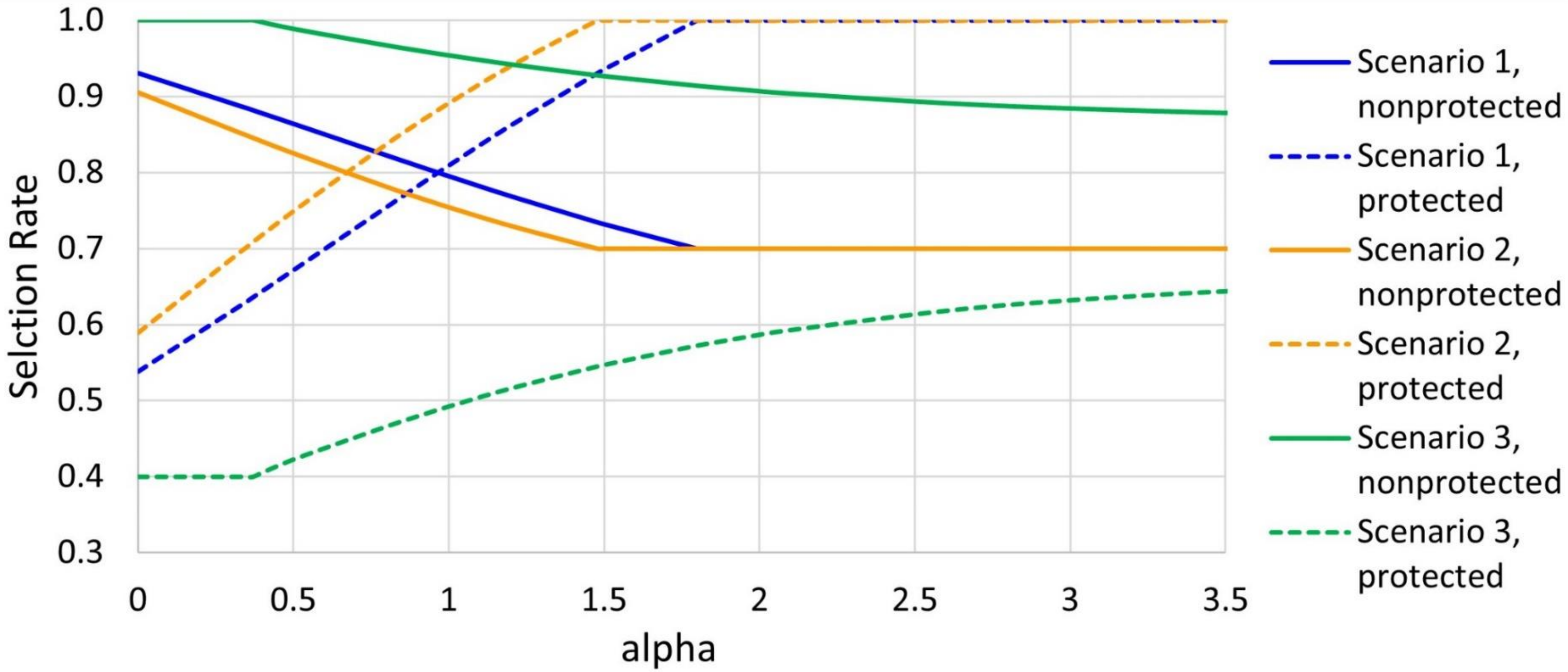
- Overall selection rate = 0.6



- Similar pattern, higher rates.

Alpha-fair Selection Rates

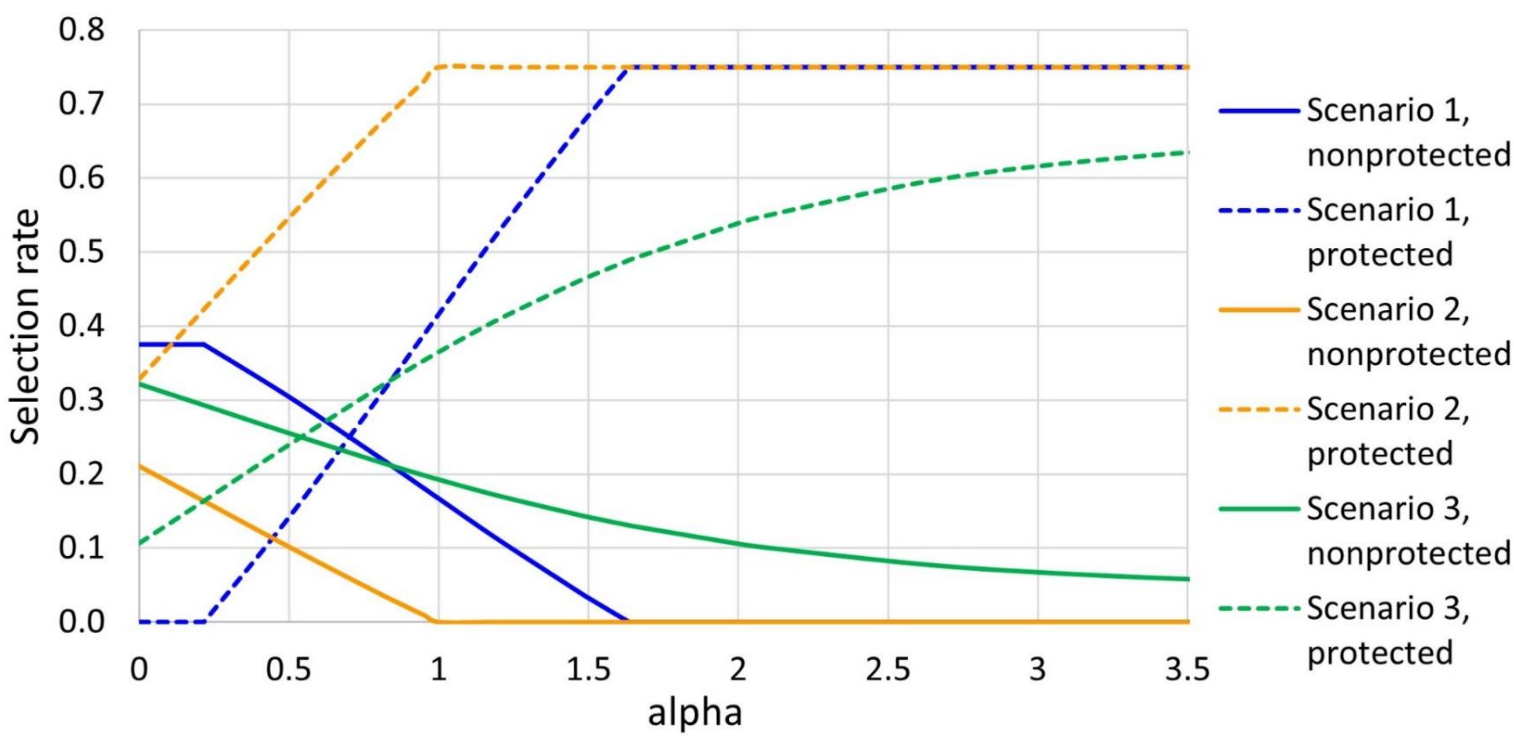
- Overall selection rate = 0.8



- Similar pattern, still higher rates.

Demographic Parity

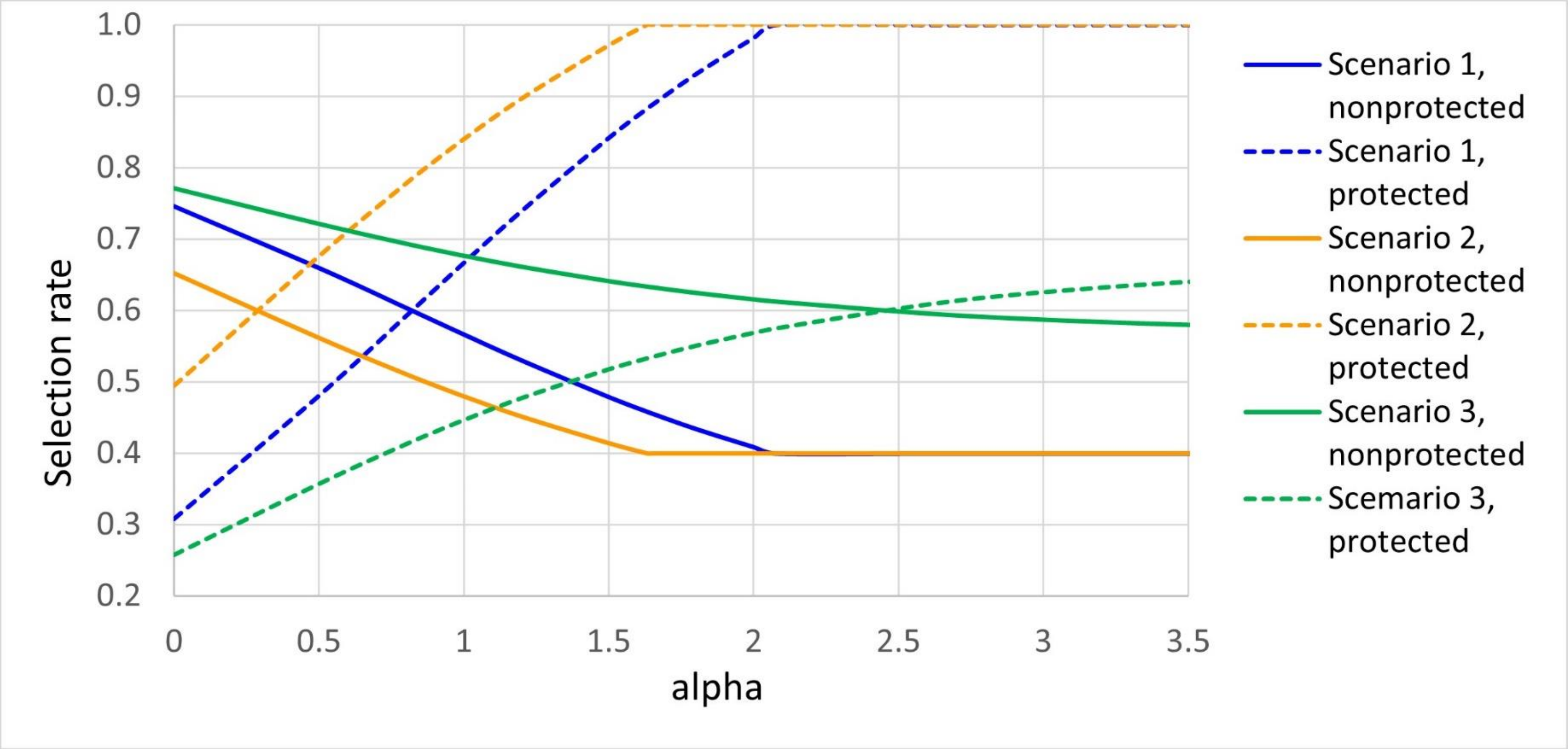
- Overall selection rate = 0.25



- Parity achieved when majority & protected curves **intersect**.
- Parity corresponds to relatively **low** degree of fairness.
- Protected group in Scenario 2 has higher rate even with **α = 0**.

Demographic Parity

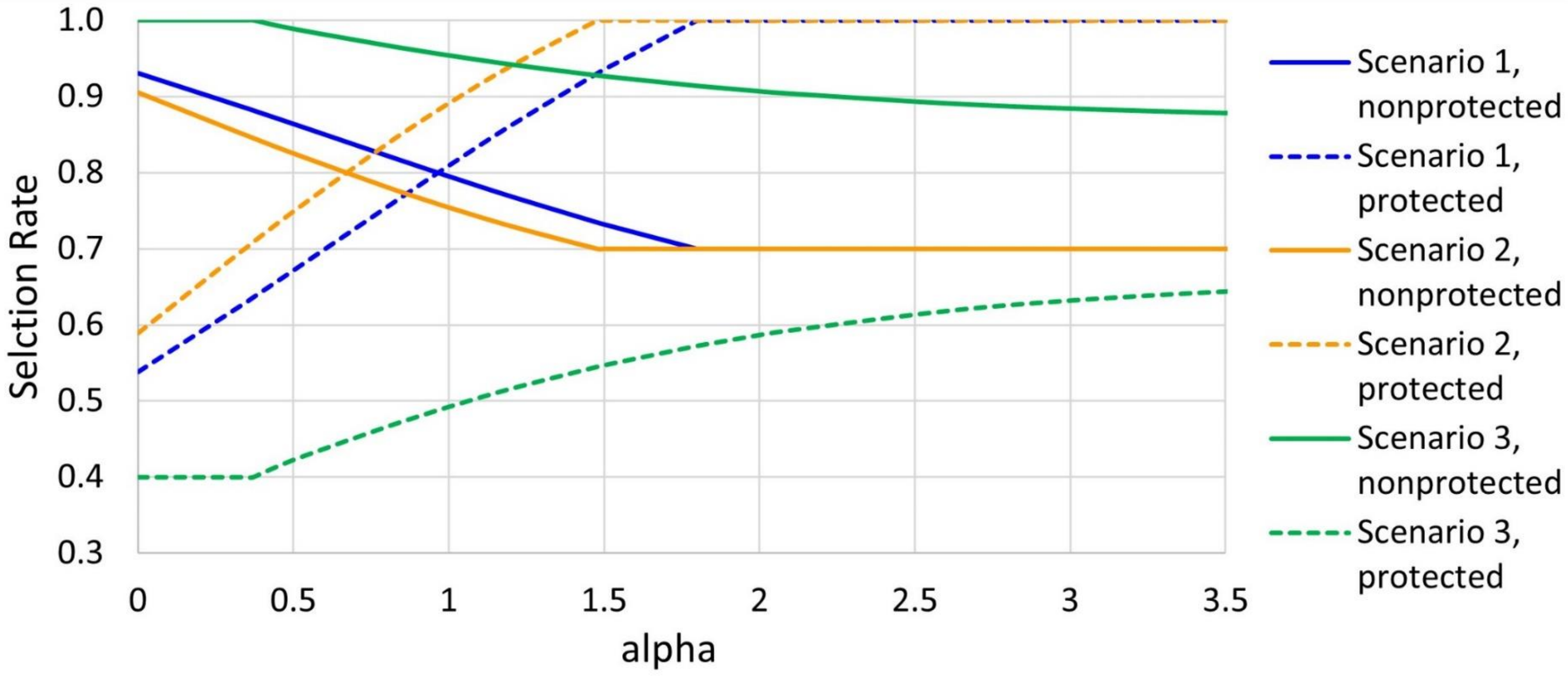
- Overall selection rate = 0.6



- Parity in Scenario 2 now requires a **slight** degree of fairness.
- Scenario 3 parity requires **large** α due to high cost of fairness.

Demographic Parity

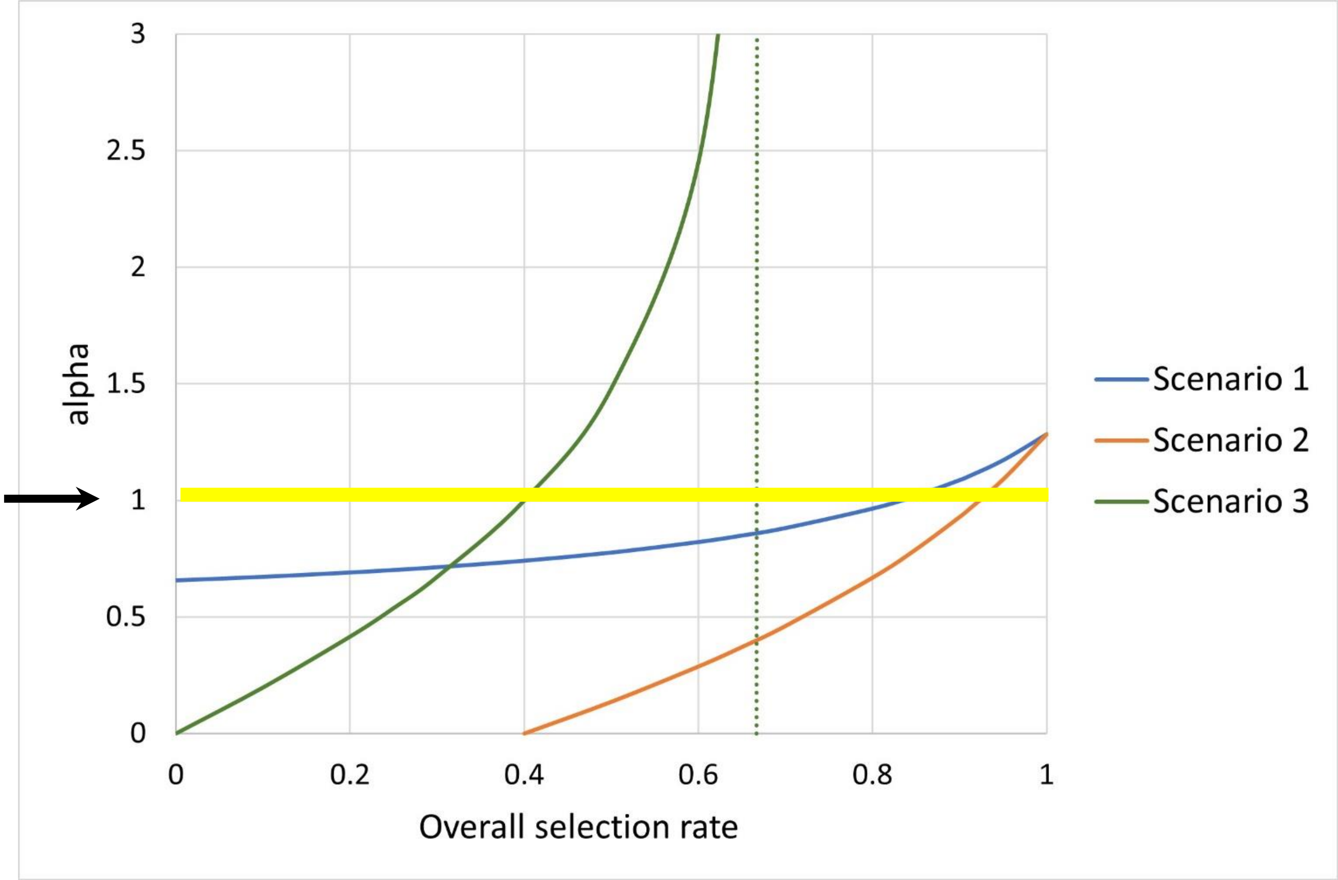
- Overall selection rate = 0.8



- Parity **impossible** in Scenario 3 because alpha fairness never calls for harmful selections.

Demographic Parity

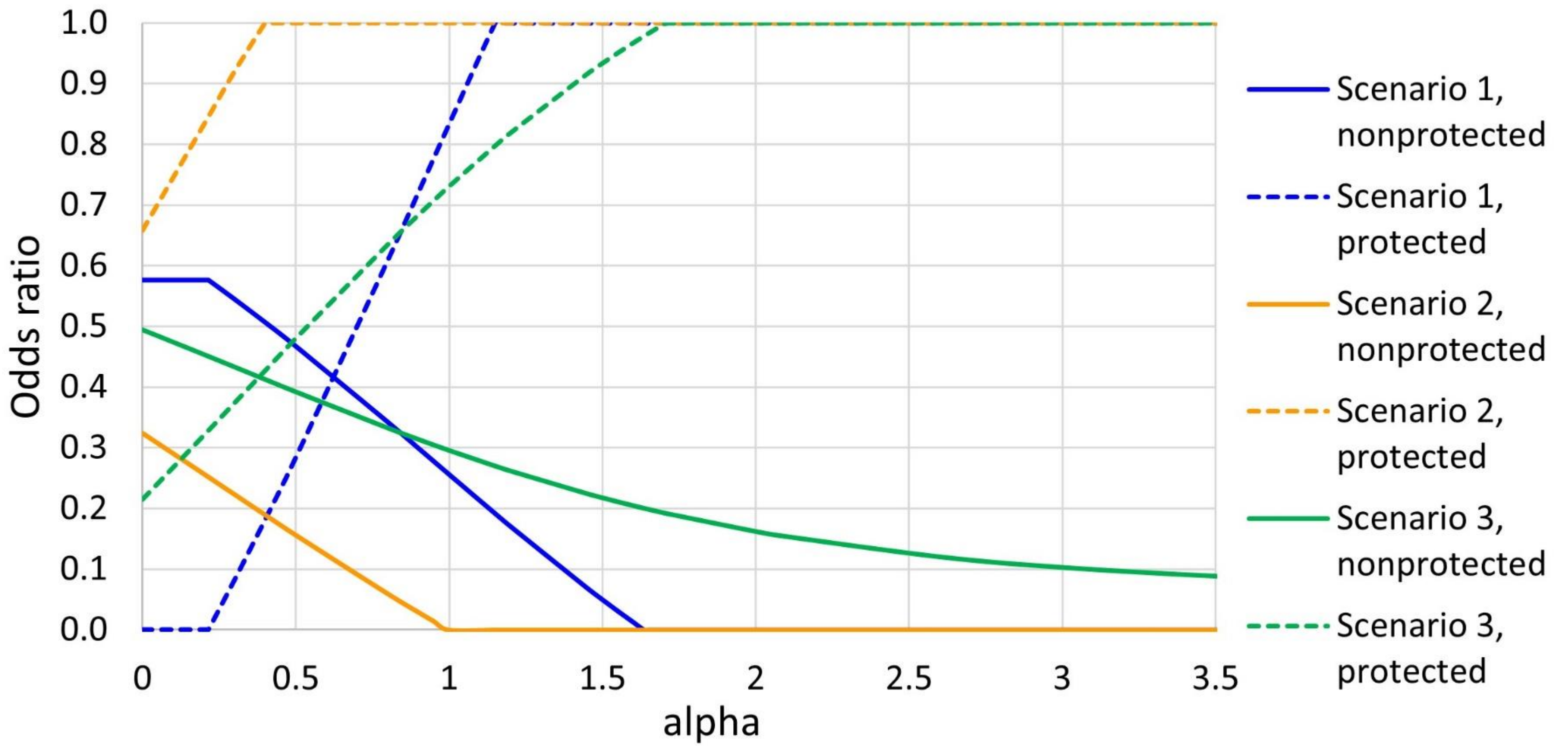
Proportional fairness



- Parity generally corresponds to **less than proportional fairness**.

Equalized Odds

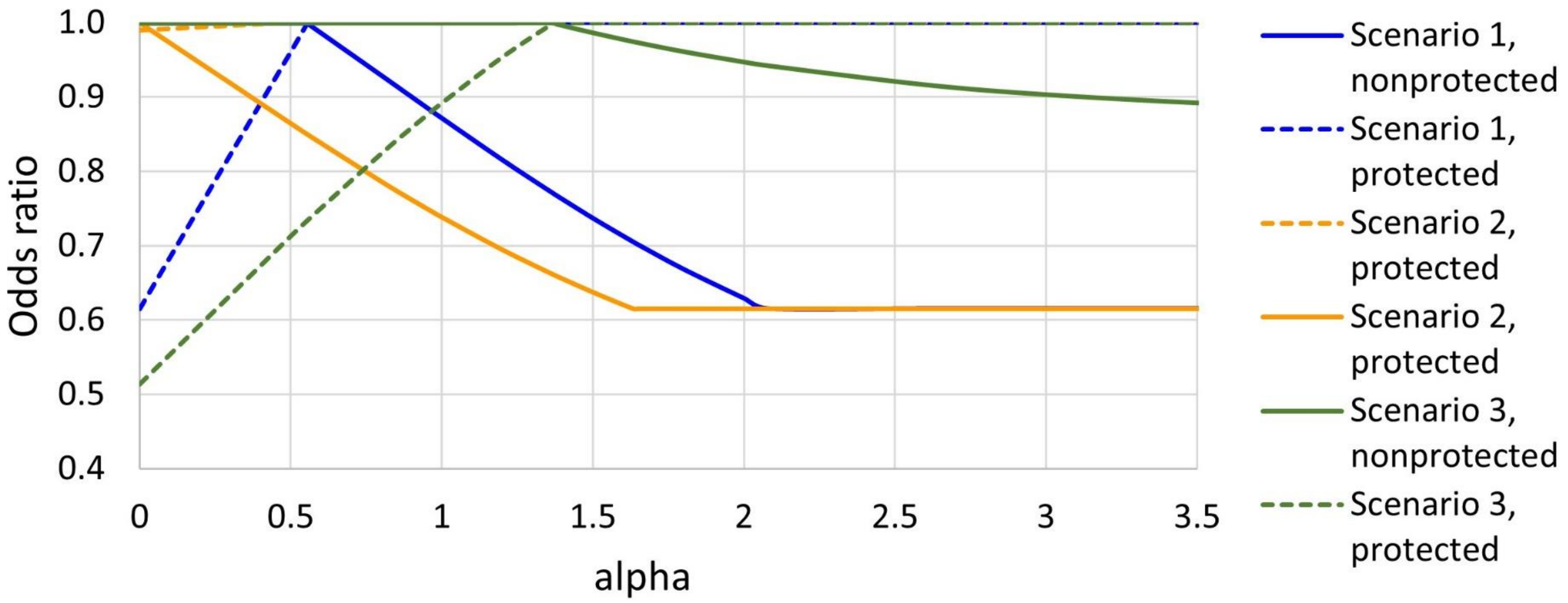
- Assume majority is 65% qualified, protected group 50% qualified.
- Overall selection rate = **0.25** < overall qualification rate of 0.6



- Even **less fair than demographic parity**.
- Sometimes viewed as **easier to defend** than demographic parity.

Equalized Odds

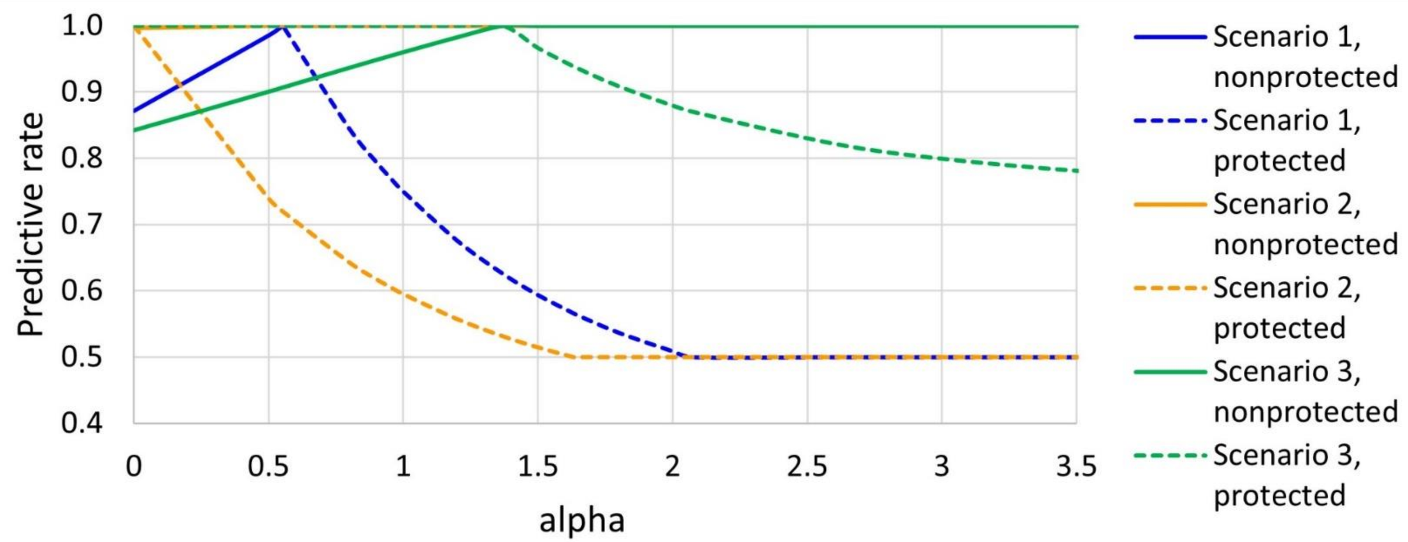
- Overall selection rate = **0.6** = overall qualification rate



- Only an **accuracy maximizing** solution (odds ratio = 1) yields equalized odds. **Fairness not a factor**.
- Nearly all odds ratios = 1 when selecting **more** individuals than are qualified.

Predictive Rate Parity

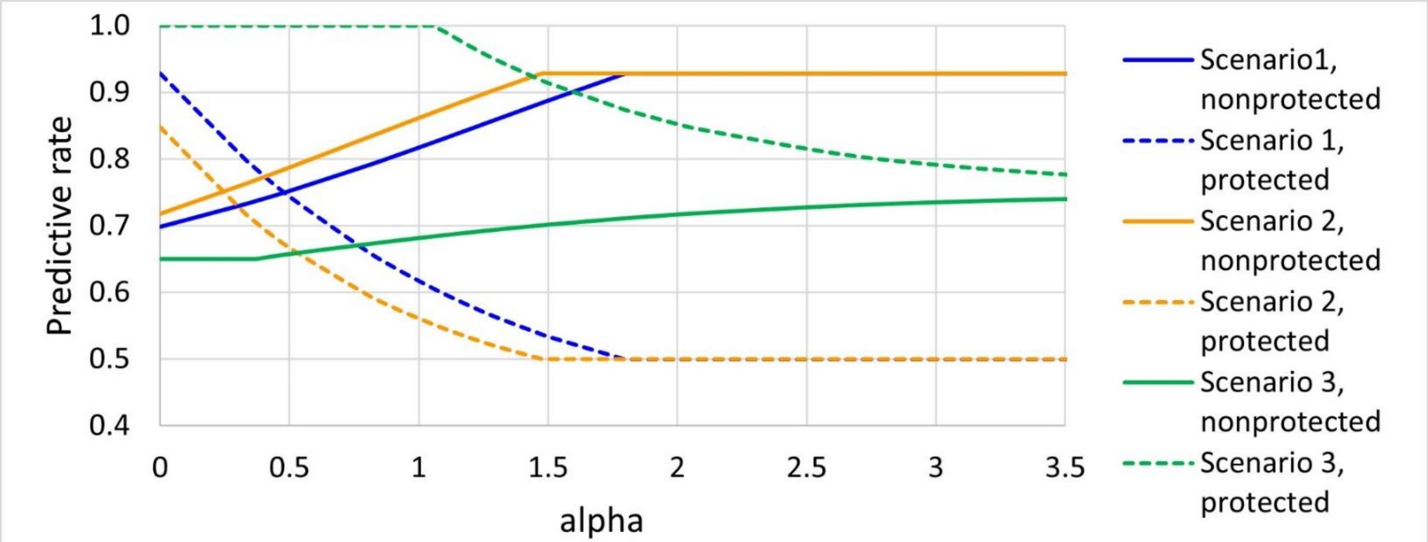
- Overall selection rate = **0.6** = overall qualification rate



- Higher predictive rates = **smaller** selection rates for protected group.
- Only an **accuracy maximizing** solution (pred rate = 1) yields predictive rate parity. **Fairness not a factor**.

Predictive Rate Parity

- Overall selection rate = **0.8** > overall qualification rate



- Nearly all predictive rates = 1 when selecting **fewer** individuals than are qualified.
- Predictive rate parity is a meaningful parity measure only when selecting **more** individuals than are qualified.

Conclusions

- Accounting for **welfare**
  - Alpha fairness (for suitable α) can normally result in **any** of the 3 types of parity, but usually when $\alpha < 1$.
  - **Significant disparity** (favoring the protected group) is often necessary to achieve fairness.
  - Achieving parity is generally **less fair than proportional fairness**
    - Even though proportional fairness is something of an **industry standard** in engineering.

Conclusions

- **Assessing parity metrics**
  - Implications of alpha fairness depend heavily on **how many individuals are selected** relative to number qualified.
  - **Equalized odds** is a meaningful fairness measure only when selecting **fewer** individuals than are qualified.
  - **Equalized odds** is **less fair** (measured by α) than **demographic parity**.
    - Which is consistent with the possibility that it is **easier to defend** on ethical grounds.
  - **Predictive rate parity** is meaningful only when selecting **more** individuals than are qualified, which may be **unrealistic**.

Conclusions

- **Parole** example
  - **Discrimination** occurs when conditions for parole are **stricter** for the minority group.
    - That is, when the minority group has a **lower odds ratio**, or a **higher predictive rate**.
  - Regarding COMPAS:
  - **Equalized odds** is relevant only if COMPAS paroles **fewer** prisoners than are qualified
    - That is, fewer than are expected to say out of prison.
  - Its ability to **achieve predictive rate parity** is an **advantage** if it paroles **more** prisoners than are qualified…
    - …perhaps in order to achieve parity without **tightening** conditions for the **majority** group.

Conclusions

- **Multiple protected groups**
  - Parity for **all groups**, even when possible, does not correspond to alpha fairness **for any α.**
    - Unless the groups are very similar.
  - However, alpha fairness for a given α can achieve a desired degree of fairness across the population as a whole
    - and in so doling, treat each group **“fairly”** in view of **its specific circumstances**.


Questions or
comments?