PHPE 308M/PHIL 209F
Fairness

Eric Pacuit, University of Maryland

November 19, 2025

A Perfectly Fair Algorithm

Suppose that there are a bunch of coins of varying biases.

Each individual in the population is

1. randomly assigned a coin; and
2. randomly assigned to one of two rooms, A and B.

Goal: For each person, Predict whether that person's coin will land heads or tails. That is, our aim is to Predict, for each person, whether they are a heads person or a tails person.

Luckily, each coin comes labeled with its bias, with a real number in the interval $[0, 1]$ indicating its bias, or its objective chance of landing heads.

A Perfectly Fair Algorithm

For each person, take their coin and read its label.

- If the coin label says x, assign that person a risk score of x.
- if $x > 0.5$, then Predict that they are a heads person (positive)
- if $x < 0.5$, then Predict that they are a tails person (negative).
- if $x = 0.5$, then randomize prediction (but "sidestep this issue by assuming that none of the coins are labeled "0.5").

A Perfectly Fair Algorithm

For each person, take their coin and read its label.

- If the coin label says x, assign that person a risk score of x.
- if $x > 0.5$, then Predict that they are a heads person (positive)
- if $x < 0.5$, then Predict that they are a tails person (negative).
- if $x = 0.5$, then randomize prediction (but "sidestep this issue by assuming that none of the coins are labeled "0.5").

This algorithm is perfectly fair and unbiased, and in particular, it is not unfair to any people in virtue of their room membership.

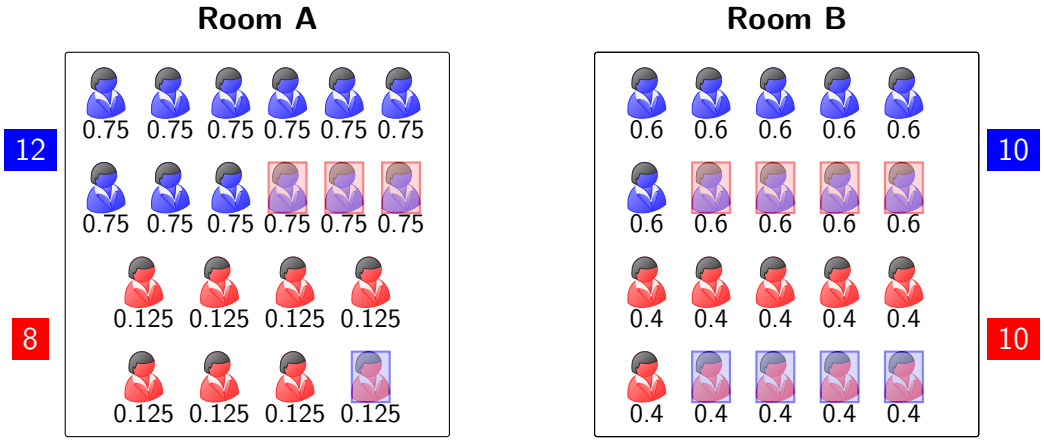
A Perfectly Fair Algorithm

- The algorithm predictions are not sensitive to individuals' room membership. And the sole feature on which its predictions are based (the labeled bias of the coin) is clearly the relevant one to focus on and is neither a proxy for, nor caused or explained by, room membership.

- Indeed, it is not just that the algorithm is in no way unfair to individuals in virtue of their membership in a certain room; there is seemingly no unfairness of any kind anywhere in this situation.

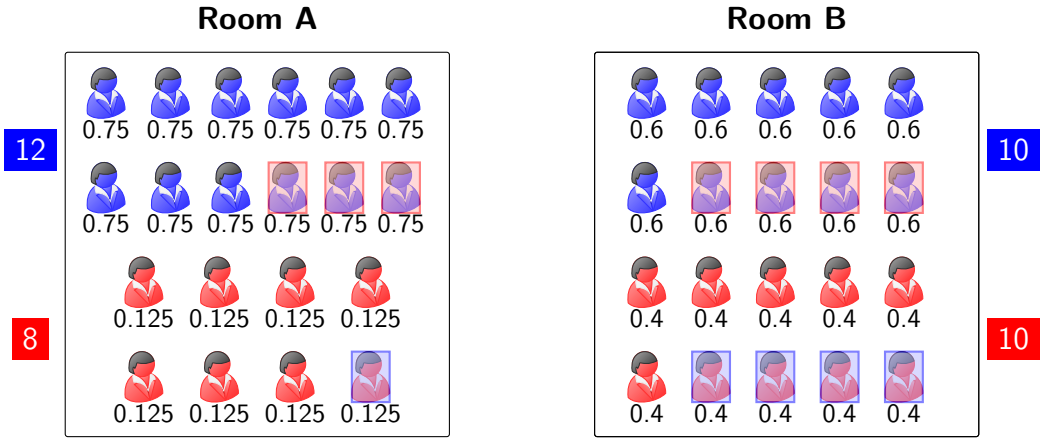
- This algorithm is uniquely optimal; no alternative can be expected to do as well or better at Predicting whether individuals are heads people or tails people.

Equal False-Positive Rates is Violated



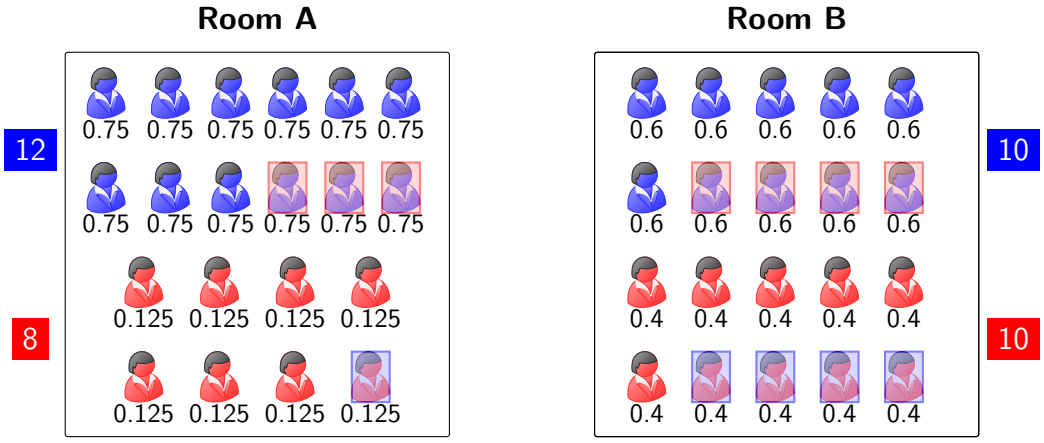
Room A		Room B
(False Pos Rate) 3/10	$\neq$	4/10 (False Pos Rate)

Equal False-Negative Rates is Violated



Room A		Room B
(False Neg Rate) 1/10	$\neq$	4/10 (False Neg Rate)

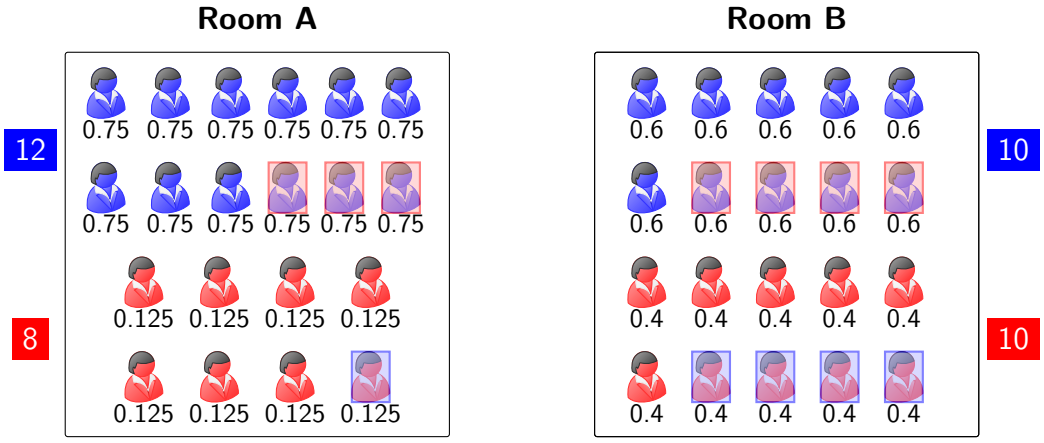
Equal Positive Predicative Value is Violated



Room A	Room B

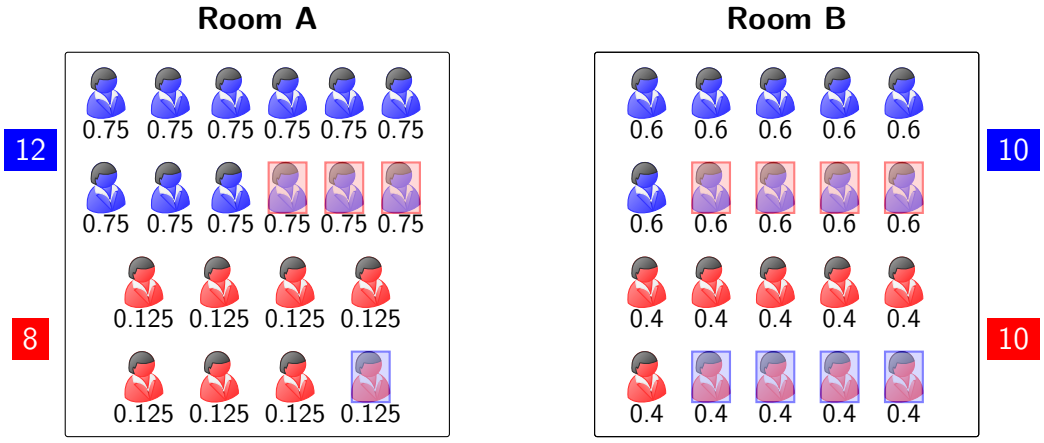
(Pos Predicative Value) $9/12 \neq 6/10$ (Pos Predicative Value)

Equal Negative Predicative Value is Violated



Room A	Room B
(Neg Predicative Value) $7/8$	$\neq$ $6/10$ (Neg Predicative Value)

Equal Overall Error Rates is Violated



Room A		Room B
(Overall Error Rate) $4/20$	$\neq$	$8/20$ (Overall Error Rate)

	Room A	Room B
Avg Score of Positives	0.6875	0.52
Avg Score of Negatives	0.3125	0.48
False Pos Rate	3/10	4/10
False Neg Rate	1/10	4/10
Pos Predicative Value	3/4	3/5
Neg Predicative Value	7/8	3/5
Ratio False Pos: False Neg	3	1
Overall Error Rate	4/20	8/20
% Predicted to be Pos	12/20	10/10
Ratio of Pred Pos:Actual Pos	12/10	10/10

- It should be clear that previous facts do not show that the Predicative algorithm was unfair or biased against any individuals in virtue of their being members of one room or the other.

- It should be clear that previous facts do not show that the Predicative algorithm was unfair or biased against any individuals in virtue of their being members of one room or the other.

- It is argued that none of the previous criteria (except Calibration Within Groups) are **necessary** for fairness.

- It should be clear that previous facts do not show that the Predicative algorithm was unfair or biased against any individuals in virtue of their being members of one room or the other.

- It is argued that none of the previous criteria (except Calibration Within Groups) are **necessary** for fairness.

- It is not claimed that the case of people, coins, and rooms is realistic or completely analogous to cases like COMPAS:

> Room membership is not socially constructed, is not the basis of historical oppression, and does not influence what features people have or how they "behave" (whether their coins land heads).

The conclusion does not depend on the example being realistic.

The conclusion does not depend on the example being realistic.

1. Simplifications and idealizations can help clarify issues by abstracting away from messy complicating factors. In real-life cases, group membership influences what features individuals have, thereby raising the thorny issue of basing predictions on "proxies" for group membership.

The conclusion does not depend on the example being realistic.

1. Simplifications and idealizations can help clarify issues by abstracting away from messy complicating factors. In real-life cases, group membership influences what features individuals have, thereby raising the thorny issue of basing predictions on "proxies" for group membership.

2. To conclude that some criterion is not necessary for fairness, all you need is a single case where fairness is satisfied but the criterion violated.

Conceptual Point

When a predictive algorithm is used to make decisions with distributional consequences or other effects that we deem unfair or unjust, this does not mean that the algorithm itself is unfair or biased against individuals in virtue of their group membership.

Conceptual Point

When a predictive algorithm is used to make decisions with distributional consequences or other effects that we deem unfair or unjust, this does not mean that the algorithm itself is unfair or biased against individuals in virtue of their group membership.

The unfairness or bias could instead lie elsewhere: with the background conditions of society, with the way decisions are made on the basis of its predictions, and/or with various side effects of the use of that algorithm, such as the exacerbation of harmful stereotypes.

Practical Point

The best response may sometimes be not to modify the predictive algorithm itself, but to instead intervene elsewhere,

- by changing the background conditions of society (e.g., through reparations, criminal justice reforms, or changes in the tax code),
- by modifying how we act on the basis of the algorithm's predictions (e.g., by adopting different risk thresholds for different groups, above which we deny bail, or reject a loan application, and so on), or
- by attempting to mitigate the other negative side effects of the algorithm's use.

Example: Traffic and Inequality

Suppose we face two problems: traffic and inequality.

We are deciding whether to adopt congestion pricing, which reduces traffic through extra fees for driving in the city during rush hours.

Example: Traffic and Inequality

Suppose we face two problems: traffic and inequality.

We are deciding whether to adopt congestion pricing, which reduces traffic through extra fees for driving in the city during rush hours.

One might worry that this is unfair to poorer people, who may have to drive farther to work and (since they earn less) would pay a higher percentage of their incomes on congestion fees.

Example: Traffic and Inequality

Suppose we face two problems: traffic and inequality.

We are deciding whether to adopt congestion pricing, which reduces traffic through extra fees for driving in the city during rush hours.

One might worry that this is unfair to poorer people, who may have to drive farther to work and (since they earn less) would pay a higher percentage of their incomes on congestion fees.

In response, one might be tempted to abandon congestion pricing altogether, or to shift to a more complicated scheme which exempts low-income drivers.

Example: Traffic and Inequality

But a better solution is available: institute the original congestion pricing scheme along with, say, a reduction in the income tax for all lower earners.

Example: Traffic and Inequality

But a better solution is available: institute the original congestion pricing scheme along with, say, a reduction in the income tax for all lower earners.

We have multiple goals (reducing congestion and reducing inequality), but we also have multiple points where we can intervene. We shouldn't think that fairness demands that the congestion pricing be scrapped or that lower earners be exempted.

We shouldn't ask the congestion pricing scheme to do all the work, addressing congestion and inequality at the same time.

Example: Traffic and Inequality

Of course, if it is politically or otherwise infeasible to enact this optimal policy, where congestion and inequality are addressed simultaneously but separately, it may be second best to enact the more complex congestion pricing scheme that tries to address congestion and inequality at the same time.

But we should not be misled into thinking that fairness itself requires this second-best solution.

Similarly with predictive algorithms.

We have multiple aims: fair and accurate predictions, as well as just decisions and a just overall society. And we should not put excessive responsibility on the predictive algorithm itself for achieving these multiple ends.

We should, of course, ensure that the predictive algorithm achieves the first aim. But insofar as we can, we should use additional interventions elsewhere in the system to achieve the others.

Summary

The argument is that no statistical criteria, except perhaps Calibration Within Groups, is a necessary condition on fairness for predictive algorithms.

Summary

The argument is that no statistical criteria, except perhaps Calibration Within Groups, is a necessary condition on fairness for predictive algorithms.

- But how we should actually design predictive algorithms depends on more than just the fairness of the algorithm itself.

Summary

The argument is that no statistical criteria, except perhaps Calibration Within Groups, is a necessary condition on fairness for predictive algorithms.

- But how we should actually design predictive algorithms depends on more than just the fairness of the algorithm itself.
- In some cases, we may be able to get the results we want by just ensuring the fairness of the algorithm while making suitable interventions elsewhere.

Summary

The argument is that no statistical criteria, except perhaps Calibration Within Groups, is a necessary condition on fairness for predictive algorithms.

- But how we should actually design predictive algorithms depends on more than just the fairness of the algorithm itself.
- In some cases, we may be able to get the results we want by just ensuring the fairness of the algorithm while making suitable interventions elsewhere.
- But in other cases, we ought to design the algorithm so as to achieve certain distributional and other results.

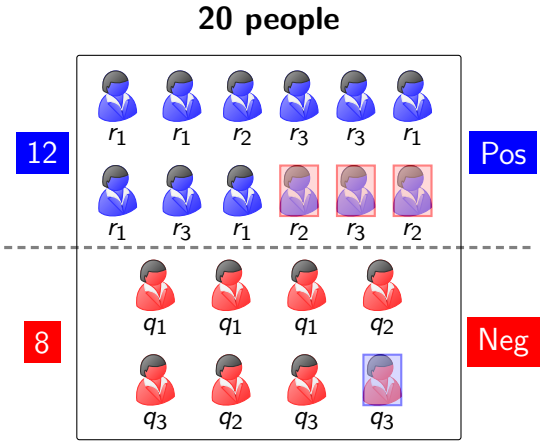
Summary

The argument is that no statistical criteria, except perhaps Calibration Within Groups, is a necessary condition on fairness for predictive algorithms.

- But how we should actually design predictive algorithms depends on more than just the fairness of the algorithm itself.
- In some cases, we may be able to get the results we want by just ensuring the fairness of the algorithm while making suitable interventions elsewhere.
- But in other cases, we ought to design the algorithm so as to achieve certain distributional and other results.
- How to go about this, however, will depend both on ethical considerations and on complex, multidimensional empirical factors not reducible to a simple formula.

Benjamin Eva (2022). *Algorithmic Fairness and Base Rate Tracking*. Philosophy & Public Affairs, 50(2), pp. 239 - 266.

Calibration



risk score	proportion Pos
r_1	1.0
r_2	1/3
r_3	3/4
q_1	0
q_2	0
q_3	1/3

Calibration Within Groups (Strong): For each possible risk score, the (expected) percentage of individuals assigned that risk score who are actually positive is the same for each relevant group and is equal to that risk score.

Calibration Within Groups (Strong): For each possible risk score, the (expected) percentage of individuals assigned that risk score who are actually positive is the same for each relevant group and is equal to that risk score.

Calibration Within Groups (Weak): For each possible risk score, the (expected) percentage of individuals assigned that risk score who are actually positive is the same for each relevant group.

Suppose that an insurance company assigns risk scores to drivers. The output of the prediction is given in the following table, where there is the same number of drivers in each group:

Age	Credit score	Base rate	Risk score
Young	Good	$\frac{3}{80}$	$\frac{1}{20}$
Young	Bad	$\frac{3}{80}$	$\frac{1}{10}$
Old	Good	$\frac{1}{40}$	$\frac{1}{20}$
Old	Bad	$\frac{1}{20}$	$\frac{1}{10}$

Suppose that an insurance company assigns risk scores to drivers. The output of the prediction is given in the following table, where there is the same number of drivers in each group:

Age	Credit score	Base rate	Risk score
Young	Good	$\frac{3}{80}$	$\frac{1}{20}$
Young	Bad	$\frac{3}{80}$	$\frac{1}{10}$
Old	Good	$\frac{1}{40}$	$\frac{1}{20}$
Old	Bad	$\frac{1}{20}$	$\frac{1}{10}$

Does this data give evidence that the algorithm is *unfair* (i.e., biased against old/young drivers)?

Suppose that an insurance company assigns risk scores to drivers. The output of the prediction is given in the following table, where there is the same number of drivers in each group:

Age	Credit score	Base rate	Risk score
Young	Good	$\frac{3}{80}$	$\frac{1}{20}$
Young	Bad	$\frac{3}{80}$	$\frac{1}{10}$
Old	Good	$\frac{1}{40}$	$\frac{1}{20}$
Old	Bad	$\frac{1}{20}$	$\frac{1}{10}$

- The overall base rate for young and old drivers is 3/80
- Calibration is violated since a risk score of 1/20 is associated with a base rate of 3/80 for young drivers and 1/40 for old drivers.
- The average risk score for young and old drivers is 3/40

Calibration is not necessary for fairness

Even though the algorithm violates calibration within groups, it is not actually unfair because the violations "cancel out":

- Old drivers with good credit are unfairly disadvantaged compared to young drivers with good credit.
- Young drivers with bad credit are unfairly disadvantaged compared to old drivers with bad credit.

Overall: Both age groups have the same average risk score (3/40), so no systematic age bias

Calibration is not necessary for fairness

At what level of "group" should we evaluate fairness?

- Fine-grained level: Young/good vs Old/good (calibration violated, so it is unfair)
- Coarse-grained level: Young vs Old overall (average scores are equal, so it is fair)

Redlining

Imagine a bank that wants to discriminate against black loan applicants, and suppose that black applicants tend to live in zip codes with higher than average default rates, although, within any given zip code, black applicants actually have the same average default rate as other applicants from the same area.

Redlining

Imagine a bank that wants to discriminate against black loan applicants, and suppose that black applicants tend to live in zip codes with higher than average default rates, although, within any given zip code, black applicants actually have the same average default rate as other applicants from the same area.

The bank can achieve its discriminatory agenda by assigning risk scores to applicants based purely on their zip code, and ignoring other relevant factors like income, credit history, and so on.

This is an idealized illustration of a real historical phenomena called "redlining," which lenders used to avoid giving mortgages to minority applicants in the 1930s.

Redlining 1

Race	Zip	Credit	Number	Default rate	Risk score
White	TR10	Good	90	$\frac{1}{10}$	$\frac{1}{4}$
White	TR10	Bad	30	$\frac{1}{5}$	$\frac{1}{4}$
White	TR11	Good	40	$\frac{1}{10}$	$\frac{3}{4}$
White	TR11	Bad	40	$\frac{1}{5}$	$\frac{3}{4}$
Black	TR10	Good	60	$\frac{1}{10}$	$\frac{1}{4}$
Black	TR10	Bad	20	$\frac{1}{5}$	$\frac{1}{4}$
Black	TR11	Good	60	$\frac{1}{10}$	$\frac{3}{4}$
Black	TR11	Bad	60	$\frac{1}{5}$	$\frac{3}{4}$

Redlining: A Calibrated but Unfair Algorithm

The algorithm assigns risk scores based ONLY on zip code:

- All TR10 residents are assigned a risk score of $1/4$
- All TR11 residents are assigned a risk score of $3/4$

Redlining: A Calibrated but Unfair Algorithm

The algorithm assigns risk scores based ONLY on zip code:

- All TR10 residents are assigned a risk score of $1/4$
- All TR11 residents are assigned a risk score of $3/4$

The algorithm completely ignores credit scores, even though credit score is a perfect predictor of default risk:

- Good credit $\rightarrow$ default rate of $1/10$ (regardless of race or zip)
- Bad credit $\rightarrow$ default rate of $1/5$ (regardless of race or zip)

Does This Satisfy Calibration Within Groups?

Redlining 1

Race	Zip	Credit	Number	Default rate	Risk score
White	TR10	Good	90	$\frac{1}{10}$	$\frac{1}{4}$
White	TR10	Bad	30	$\frac{1}{5}$	$\frac{1}{4}$
White	TR11	Good	40	$\frac{1}{10}$	$\frac{3}{4}$
White	TR11	Bad	40	$\frac{1}{5}$	$\frac{3}{4}$
Black	TR10	Good	60	$\frac{1}{10}$	$\frac{1}{4}$
Black	TR10	Bad	20	$\frac{1}{5}$	$\frac{1}{4}$
Black	TR11	Good	60	$\frac{1}{10}$	$\frac{3}{4}$
Black	TR11	Bad	60	$\frac{1}{5}$	$\frac{3}{4}$

Check risk score 1/4 (all TR10 residents):

- White applicants: 90 good + 30 bad = 120 total
  - Defaults: $90 \times \frac{1}{10} + 30 \times \frac{1}{5} = 15$
  - Actual rate: $\frac{15}{120} = \frac{1}{8}$

Does This Satisfy Calibration Within Groups?

Redlining 1

Race	Zip	Credit	Number	Default rate	Risk score
White	TR10	Good	90	$\frac{1}{10}$	$\frac{1}{4}$
White	TR10	Bad	30	$\frac{1}{5}$	$\frac{1}{4}$
White	TR11	Good	40	$\frac{1}{10}$	$\frac{3}{4}$
White	TR11	Bad	40	$\frac{1}{5}$	$\frac{3}{4}$
Black	TR10	Good	60	$\frac{1}{10}$	$\frac{1}{4}$
Black	TR10	Bad	20	$\frac{1}{5}$	$\frac{1}{4}$
Black	TR11	Good	60	$\frac{1}{10}$	$\frac{3}{4}$
Black	TR11	Bad	60	$\frac{1}{5}$	$\frac{3}{4}$

Check risk score 1/4 (all TR10 residents):

- White applicants: 90 good + 30 bad = 120 total
  - Defaults: $90 \times \frac{1}{10} + 30 \times \frac{1}{5} = 15$
  - Actual rate: $\frac{15}{120} = \frac{1}{8}$
- Black applicants: 60 good + 20 bad = 80 total
  - Defaults: $60 \times \frac{1}{10} + 20 \times \frac{1}{5} = 10$
  - Actual rate: $\frac{10}{80} = \frac{1}{8}$

Does This Satisfy Calibration Within Groups?

Redlining 1

Race	Zip	Credit	Number	Default rate	Risk score
White	TR10	Good	90	$\frac{1}{10}$	$\frac{1}{4}$
White	TR10	Bad	30	$\frac{1}{5}$	$\frac{1}{4}$
White	TR11	Good	40	$\frac{1}{10}$	$\frac{3}{4}$
White	TR11	Bad	40	$\frac{1}{5}$	$\frac{3}{4}$
Black	TR10	Good	60	$\frac{1}{10}$	$\frac{1}{4}$
Black	TR10	Bad	20	$\frac{1}{5}$	$\frac{1}{4}$
Black	TR11	Good	60	$\frac{1}{10}$	$\frac{3}{4}$
Black	TR11	Bad	60	$\frac{1}{5}$	$\frac{3}{4}$

Check risk score 3/4 (all TR11 residents):

- White applicants: 40 good + 40 bad = 80 total
  - Defaults: $40 \times \frac{1}{10} + 40 \times \frac{1}{5} = 12$
  - Actual rate: $\frac{12}{80} = \frac{3}{20}$

Does This Satisfy Calibration Within Groups?

Redlining 1

Race	Zip	Credit	Number	Default rate	Risk score
White	TR10	Good	90	$\frac{1}{10}$	$\frac{1}{4}$
White	TR10	Bad	30	$\frac{1}{5}$	$\frac{1}{4}$
White	TR11	Good	40	$\frac{1}{10}$	$\frac{3}{4}$
White	TR11	Bad	40	$\frac{1}{5}$	$\frac{3}{4}$
Black	TR10	Good	60	$\frac{1}{10}$	$\frac{1}{4}$
Black	TR10	Bad	20	$\frac{1}{5}$	$\frac{1}{4}$
Black	TR11	Good	60	$\frac{1}{10}$	$\frac{3}{4}$
Black	TR11	Bad	60	$\frac{1}{5}$	$\frac{3}{4}$

Check risk score 3/4 (all TR11 residents):

- White applicants: 40 good + 40 bad = 80 total
  - Defaults: $40 \times \frac{1}{10} + 40 \times \frac{1}{5} = 12$
  - Actual rate: $\frac{12}{80} = \frac{3}{20}$
- Black applicants: 60 good + 60 bad = 120 total
  - Defaults: $60 \times \frac{1}{10} + 60 \times \frac{1}{5} = 18$
  - Actual rate: $\frac{18}{120} = \frac{3}{20}$

Calibration is Satisfied!

Calibration Within Groups (Weak): For each risk score, the actual default rate is the same across racial groups.

Risk Score	White Actual Rate	Black Actual Rate
1/4	1/8	1/8
3/4	3/20	3/20

Calibration is Satisfied!

Calibration Within Groups (Weak): For each risk score, the actual default rate is the same across racial groups.

Risk Score	White Actual Rate	Black Actual Rate
1/4	1/8	1/8
3/4	3/20	3/20

The algorithm satisfies weak calibration within groups.

Calibration is Satisfied!

Calibration Within Groups (Weak): For each risk score, the actual default rate is the same across racial groups.

Risk Score	White Actual Rate	Black Actual Rate
1/4	1/8	1/8
3/4	3/20	3/20

The algorithm satisfies weak calibration within groups.

But is it fair?

Base Rate Tracking

Redlining 1

Race	Zip	Credit	Number	Default rate	Risk score
White	TR10	Good	90	$\frac{1}{10}$	$\frac{1}{4}$
White	TR10	Bad	30	$\frac{1}{5}$	$\frac{1}{4}$
White	TR11	Good	40	$\frac{1}{10}$	$\frac{3}{4}$
White	TR11	Bad	40	$\frac{1}{5}$	$\frac{3}{4}$
Black	TR10	Good	60	$\frac{1}{10}$	$\frac{1}{4}$
Black	TR10	Bad	20	$\frac{1}{5}$	$\frac{1}{4}$
Black	TR11	Good	60	$\frac{1}{10}$	$\frac{3}{4}$
Black	TR11	Bad	60	$\frac{1}{5}$	$\frac{3}{4}$

- Average risk score for white applicants is

$$(90 * \frac{1}{4} + 30 * \frac{1}{4} + 40 * \frac{3}{4} + 40 * \frac{3}{4}) / 200 = 9/20.$$

Base Rate Tracking

Redlining 1

Race	Zip	Credit	Number	Default rate	Risk score
White	TR10	Good	90	$\frac{1}{10}$	$\frac{1}{4}$
White	TR10	Bad	30	$\frac{1}{5}$	$\frac{1}{4}$
White	TR11	Good	40	$\frac{1}{10}$	$\frac{3}{4}$
White	TR11	Bad	40	$\frac{1}{5}$	$\frac{3}{4}$
Black	TR10	Good	60	$\frac{1}{10}$	$\frac{1}{4}$
Black	TR10	Bad	20	$\frac{1}{5}$	$\frac{1}{4}$
Black	TR11	Good	60	$\frac{1}{10}$	$\frac{3}{4}$
Black	TR11	Bad	60	$\frac{1}{5}$	$\frac{3}{4}$

- Average risk score for black applicants is

$$(60 * \frac{1}{4} + 20 * \frac{1}{4} + 60 * \frac{3}{4} + 60 * \frac{3}{4})/200 = 11/20.$$

Base Rate Tracking

Redlining 1

Race	Zip	Credit	Number	Default rate	Risk score
White	TR10	Good	90	$\frac{1}{10}$	$\frac{1}{4}$
White	TR10	Bad	30	$\frac{1}{5}$	$\frac{1}{4}$
White	TR11	Good	40	$\frac{1}{10}$	$\frac{3}{4}$
White	TR11	Bad	40	$\frac{1}{5}$	$\frac{3}{4}$
Black	TR10	Good	60	$\frac{1}{10}$	$\frac{1}{4}$
Black	TR10	Bad	20	$\frac{1}{5}$	$\frac{1}{4}$
Black	TR11	Good	60	$\frac{1}{10}$	$\frac{3}{4}$
Black	TR11	Bad	60	$\frac{1}{5}$	$\frac{3}{4}$

- Average default rate for white applicants is

$$(90 * \frac{1}{10} + 30 * \frac{1}{5} + 40 * \frac{1}{10} + 40 * \frac{1}{5}) / 200 = 27/200.$$

Base Rate Tracking

Redlining 1

Race	Zip	Credit	Number	Default rate	Risk score
White	TR10	Good	90	$\frac{1}{10}$	$\frac{1}{4}$
White	TR10	Bad	30	$\frac{1}{5}$	$\frac{1}{4}$
White	TR11	Good	40	$\frac{1}{10}$	$\frac{3}{4}$
White	TR11	Bad	40	$\frac{1}{5}$	$\frac{3}{4}$
Black	TR10	Good	60	$\frac{1}{10}$	$\frac{1}{4}$
Black	TR10	Bad	20	$\frac{1}{5}$	$\frac{1}{4}$
Black	TR11	Good	60	$\frac{1}{10}$	$\frac{3}{4}$
Black	TR11	Bad	60	$\frac{1}{5}$	$\frac{3}{4}$

- Average default rate for black applicants is

$$(60 * \frac{1}{10} + 20 * \frac{1}{5} + 60 * \frac{1}{10} + 60 * \frac{1}{5}) / 200 = 28/200.$$

Base Rate Tracking

	Black	White	Difference
Average Risk Score	11/20	9/20	2/20
Average Default Rate	28/200	27/200	1/200

The difference between the average risk scores of the two groups is 20 times as great as the difference between their actual default rates.

Base Rate Tracking

If an algorithm assigns one group a higher average risk score than another, that discrepancy has to be justified by a corresponding discrepancy between the base rates of those two groups, and the magnitudes of those discrepancies should be equivalent.

Base Rate Tracking

In slogan form: an algorithm should only treat one groups as much more risky than another if it really is much more risky.

Base Rate Tracking: The difference between the average risk scores assigned to the relevant groups should be equal to the difference between the (expected) base rates of those groups.