

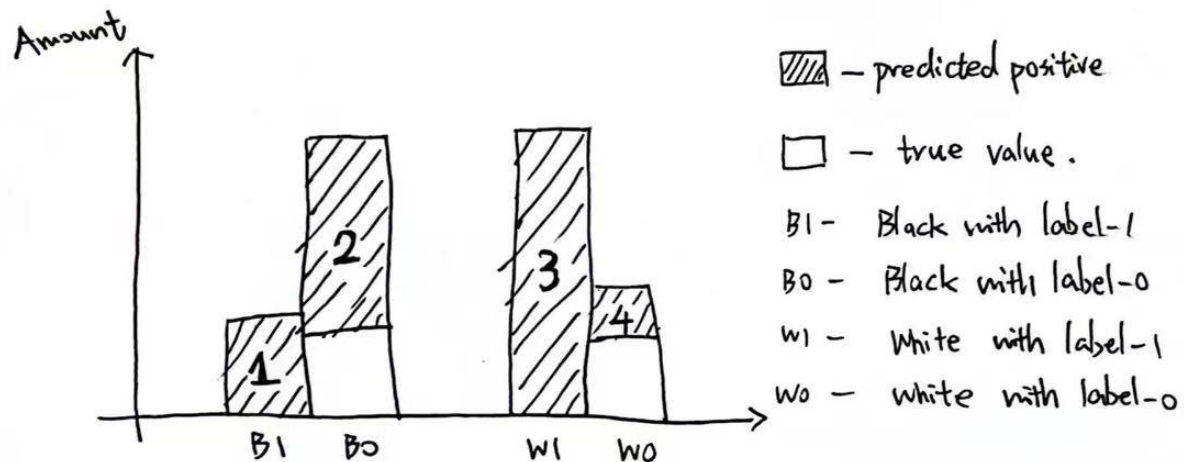
# **Concrete Examples of Fairness Definition**

# 1. Demographic Parity

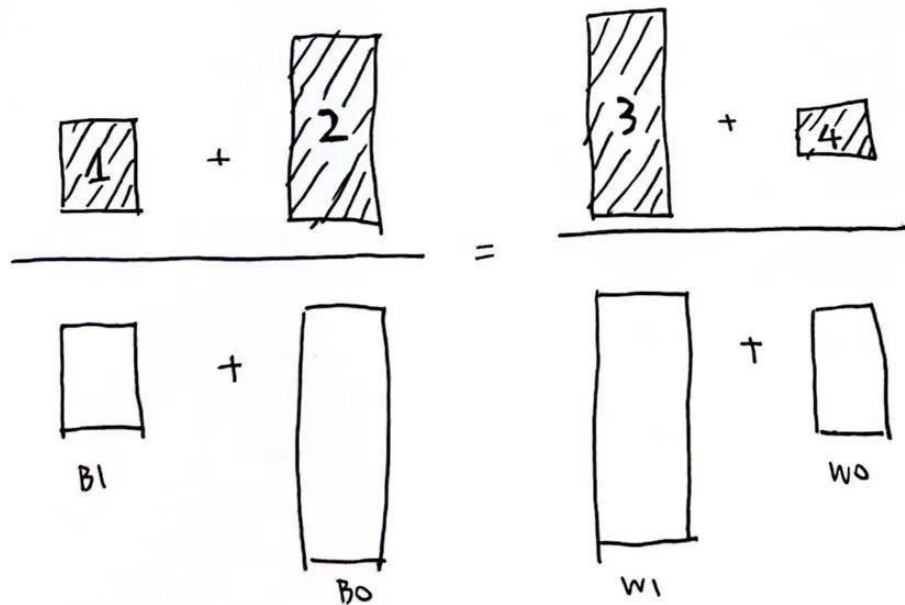
- **Formula:**  $\Pr(Y' = 1 \mid A=0) = \Pr(Y' = 1 \mid A=1)$
- **Definition:** The likelihood of a positive outcome should be the same regardless of whether the person is in the protected group. In other words, demographic parity is achieved when the probability of a certain prediction is not dependent on sensitive group membership.
- **Example:** In a loan application process, demographic parity means that the approval rate should be the same for all demographic groups when considering people with similar background. For instance, if 70% of **loan applications from Group A** (e.g., males) are approved, demographic parity requires that approximately 70% of **loan applications from Group B** (e.g., females) should also be approved.

# Demographic Parity

$$Pr(Y'=1 | A=0) = Pr(Y'=1 | A=1)$$



- Overall proportion of predicted positive.
- Both positive and negative columns.
- Neglect the shadowed ratio of each columns.

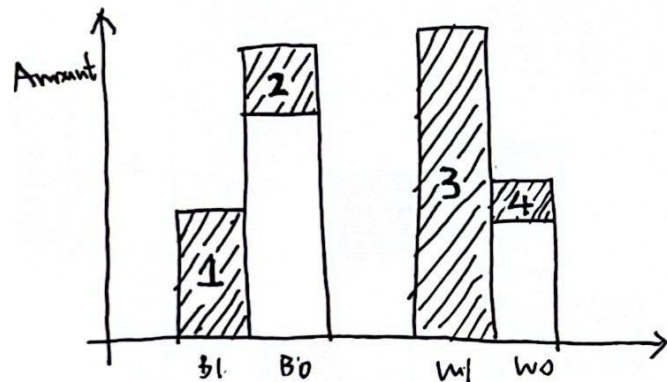


"The shadow area has equal proportion in the total number of sample among groups."

## 2. Equalized Odds

- **Formula:**  $\Pr(Y' = 1 \mid A=0, Y=y) = \Pr(Y' = 1 \mid A=1, Y=y)$
- **Definition:** This means that the probability of a person in the positive class being correctly assigned a positive outcome and the probability of a person in a negative class being incorrectly assigned a positive outcome should both be the same for the protected and unprotected group members.
- **Example:** In the criminal justice system, equalized odds means that the accuracy of predicting recidivism (likelihood of reoffending) should be similar across different groups. If the **TPR** for **Group A** (e.g., white people) is 0.8 (80%) and the **FPR** is 0.1 (10%), equalized odds requires that the **TPR and FPR** for **Group B** (e.g., black people) should also be comparably 0.8 and 0.1, respectively.

Equalized Odds  $\Pr(Y'=1 | A=0, Y=y) = \Pr(Y'=1 | A=1, Y=y)$



- predicted positive.

- true value.

B0 - Black with label-0

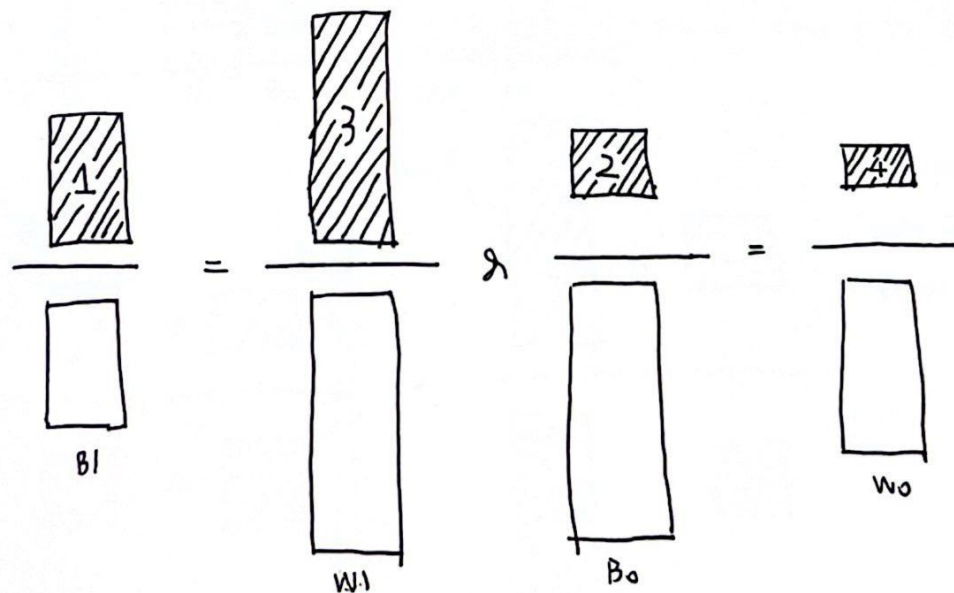
B1 - Black with label-1

W0 - White with label-0

W1 - White with label-1

- Proportion of predicted positive to true positive and true negative.

- Consider both positive and negative columns.



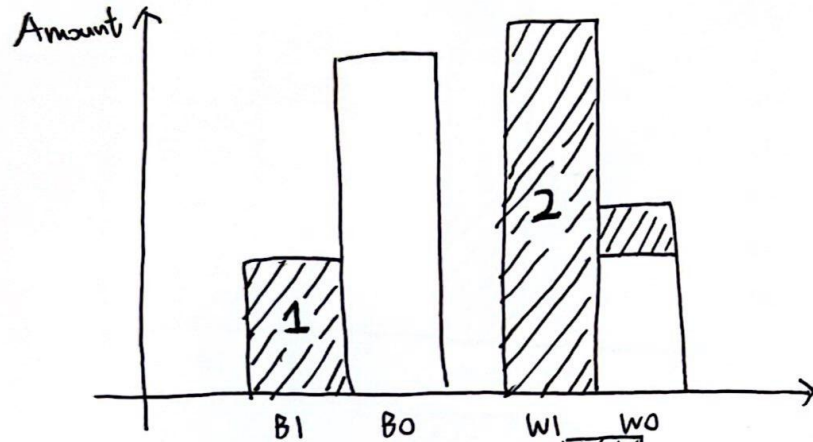
"The shadow areas of each column should have equal proportion."

### 3. Equalized Opportunity

- **Formula:**  $\Pr(Y' = 1 \mid A=0, Y=1) = \Pr(Y' = 1 \mid A=1, Y=1)$
- **Definition:** The probability of a person in a positive class being assigned to a positive outcome should be equal for both protected and unprotected group members. In other words, the equal opportunity definition states that the protected and unprotected groups should have equal true positive rates.
- **Example:** In the context of college admissions, equalized opportunity means that the acceptance rates for different racial or ethnic groups should be similar when considering students with similar academic qualifications. For instance, if the true positive rate for **admitting students of Group A** (e.g., American) is 0.7 (70%), equalized opportunity requires that the true positive rate for **admitting students of Group B** (e.g., Asian) should also be 0.7.

# Equalized Opportunity

$$\Pr(Y'=1|A=0, Y=1) = \Pr(Y'=1|A=1, Y=1)$$



▨ - predicted positive

□ - true value

B1 - Black with label-1

B0 - Black with label-0

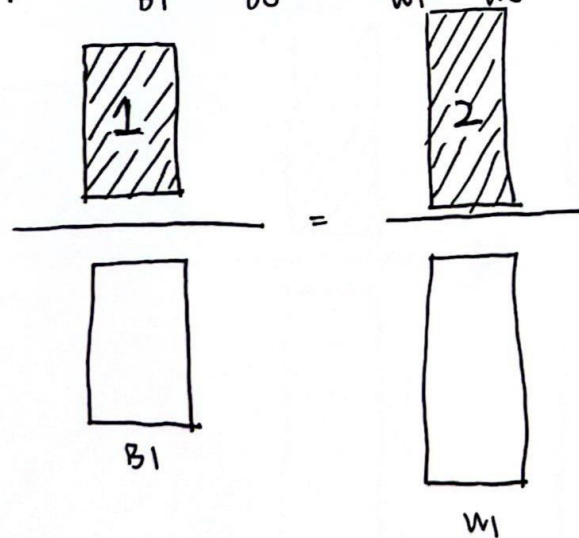
W1 - white with label-1

W0 - white with label-0

- Proportion of predicted positive to true positive.

- Consider only positive column B1 & W1.

- Neglect the right negative column B0 & W0.



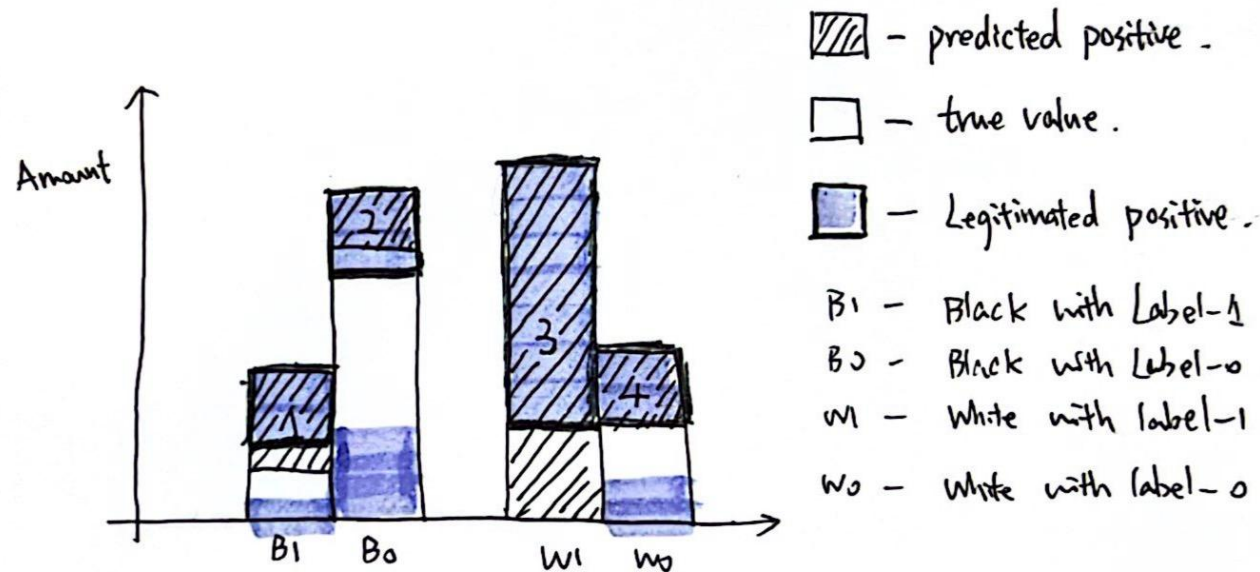
"The shadow areas of positive true value have  $\neq$  equal proportion among groups."

## 4. Conditional Statistical Parity

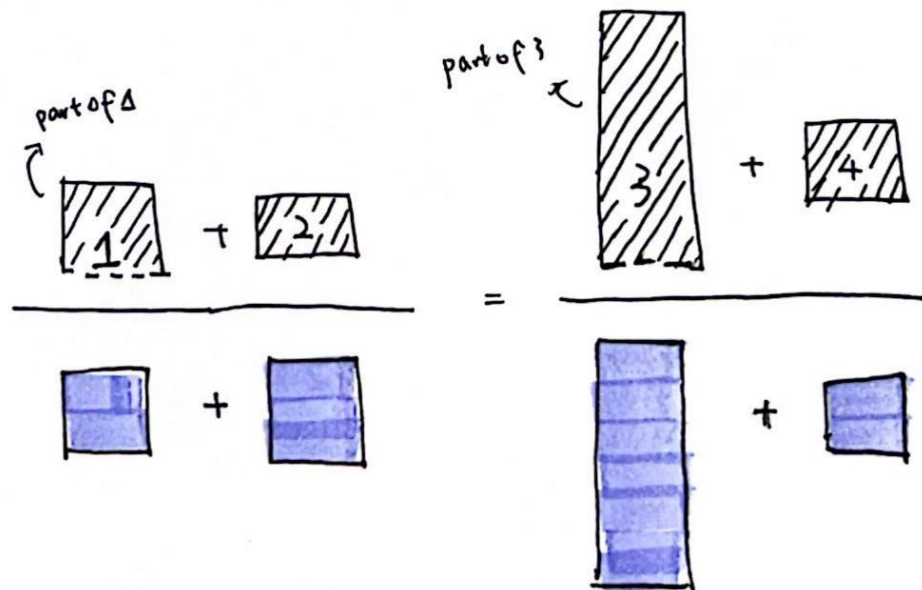
- **Formula:**  $\Pr(Y'=1 \mid L=1, A=0) = \Pr(Y'=1 \mid L=1, A=1)$
- **Definition:** The **legitimate factors L** refers to a group of attributes that may affect the prediction outcome other than the sensitive attribute. It states that people in both protected and unprotected groups should have equal probability of being assigned to a positive outcome given a set of legitimate factors L.
- **Example:** In the hiring process, conditional statistical parity means that the hiring rate should be the same for all demographic groups when considering individuals with similar **qualifications**. For instance, if 50% of **qualified applicants from Group A** (e.g., males) are hired, conditional statistical parity requires that approximately 50% of **qualified applicants from Group B** (e.g., females) should also be hired.



Conditional Statistical Parity  $\Pr(Y'=1 | L=1, A=0) = \Pr(Y'=1 | L=1, A=1)$



- Overall proportion of predicted positive to the legitimated positive.
- Both positive and negative columns.



"The proportion of the shadow area to the colored area is equal among groups."