

CLASSIFICATION ERRORS

EMMA BEAUXIS-AUSSALET

e.m.a.l.beauxis@hva.nl

RECAP

EMMA BEAUXIS-AUSSALET

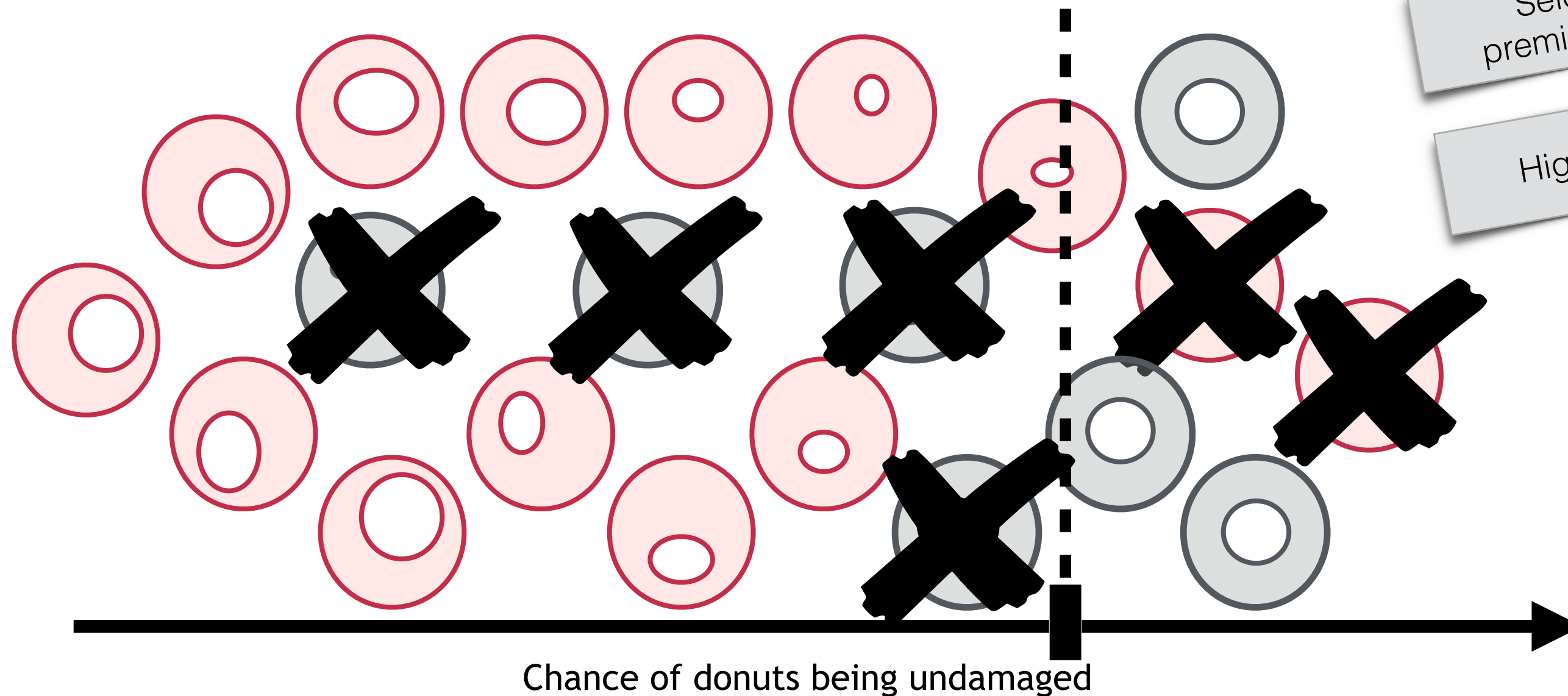
e.m.a.l.beauxis@hva.nl

TUNING THE ERRORS

damaged

undamaged

Tuning parameters can **balance errors between classes**.

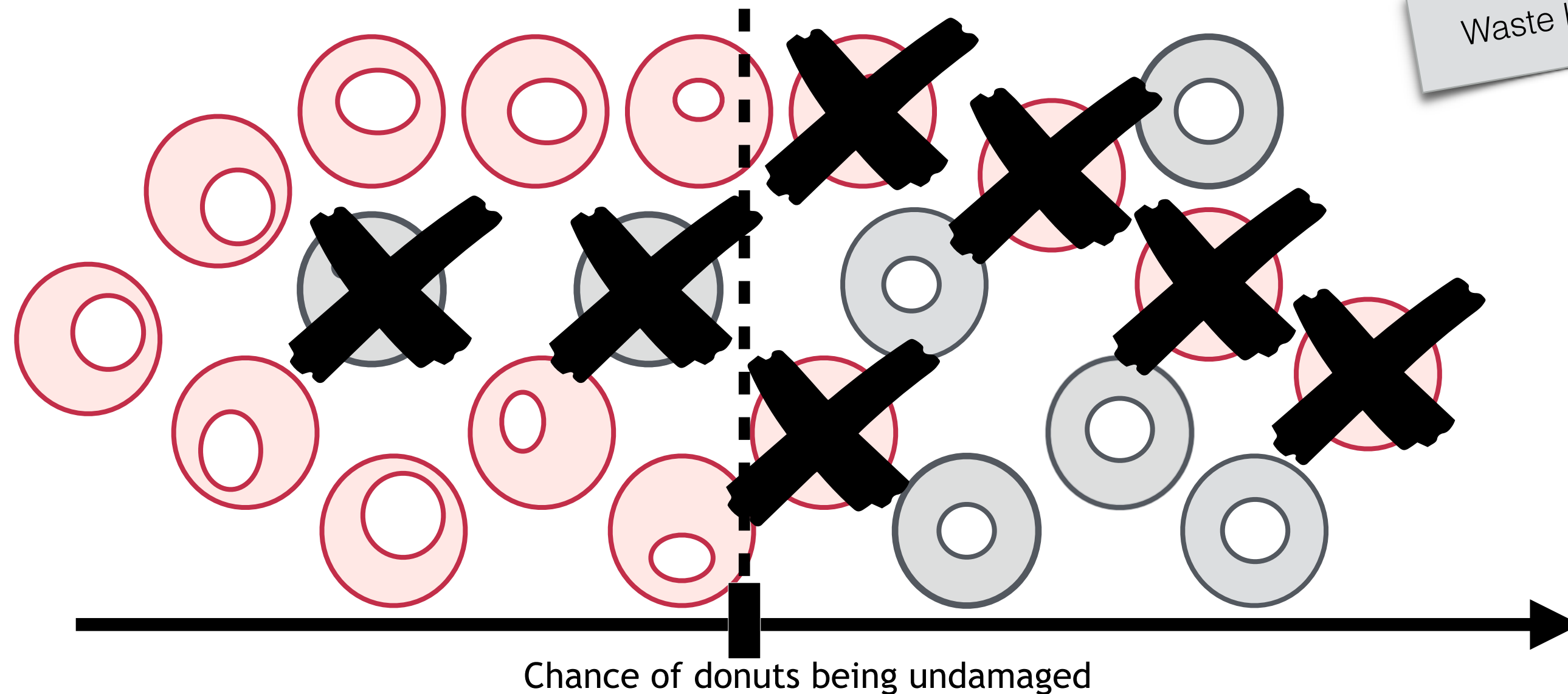


TUNING THE ERRORS

damaged

undamaged

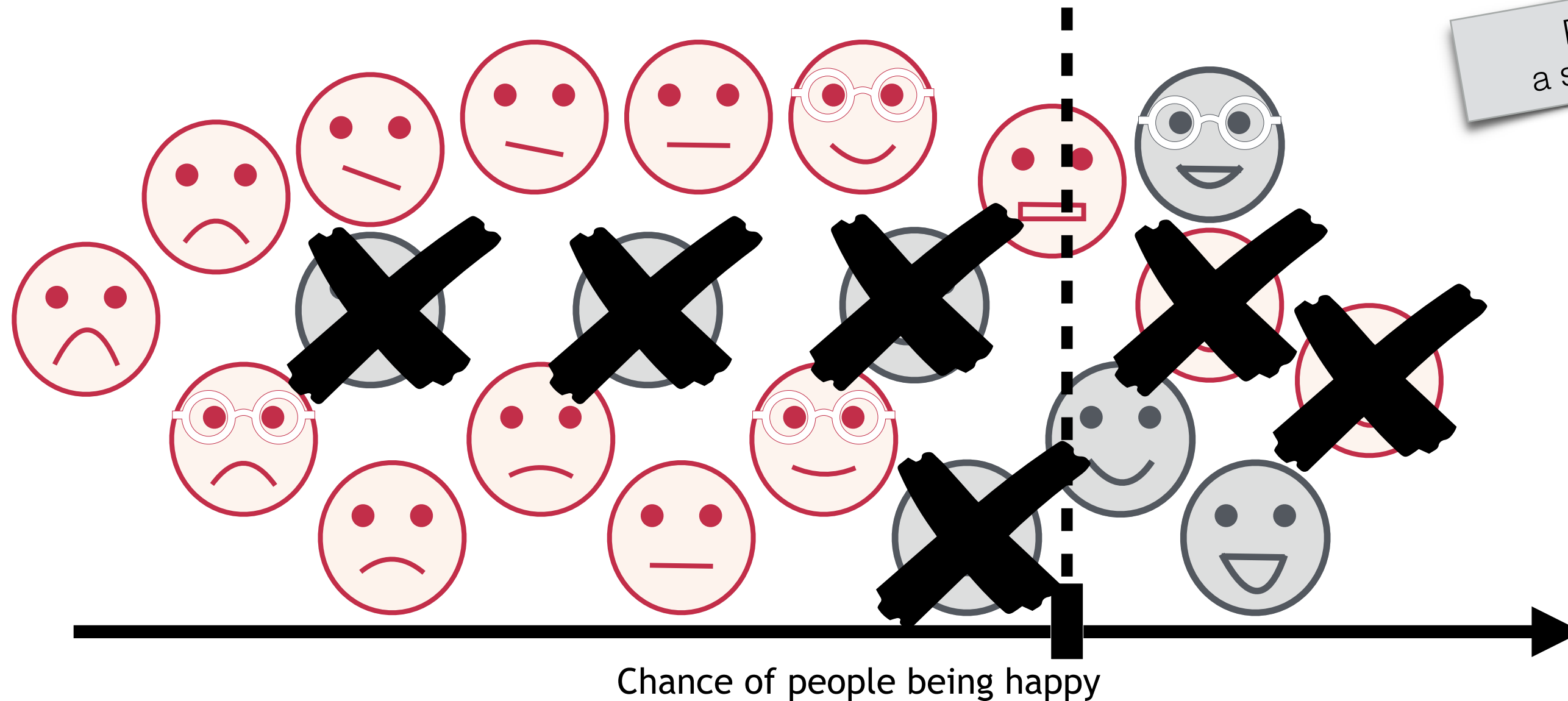
Tuning parameters can **balance errors between classes**.



TUNING THE ERRORS

sad : happy

The **tolerance to errors** depends on the use case.



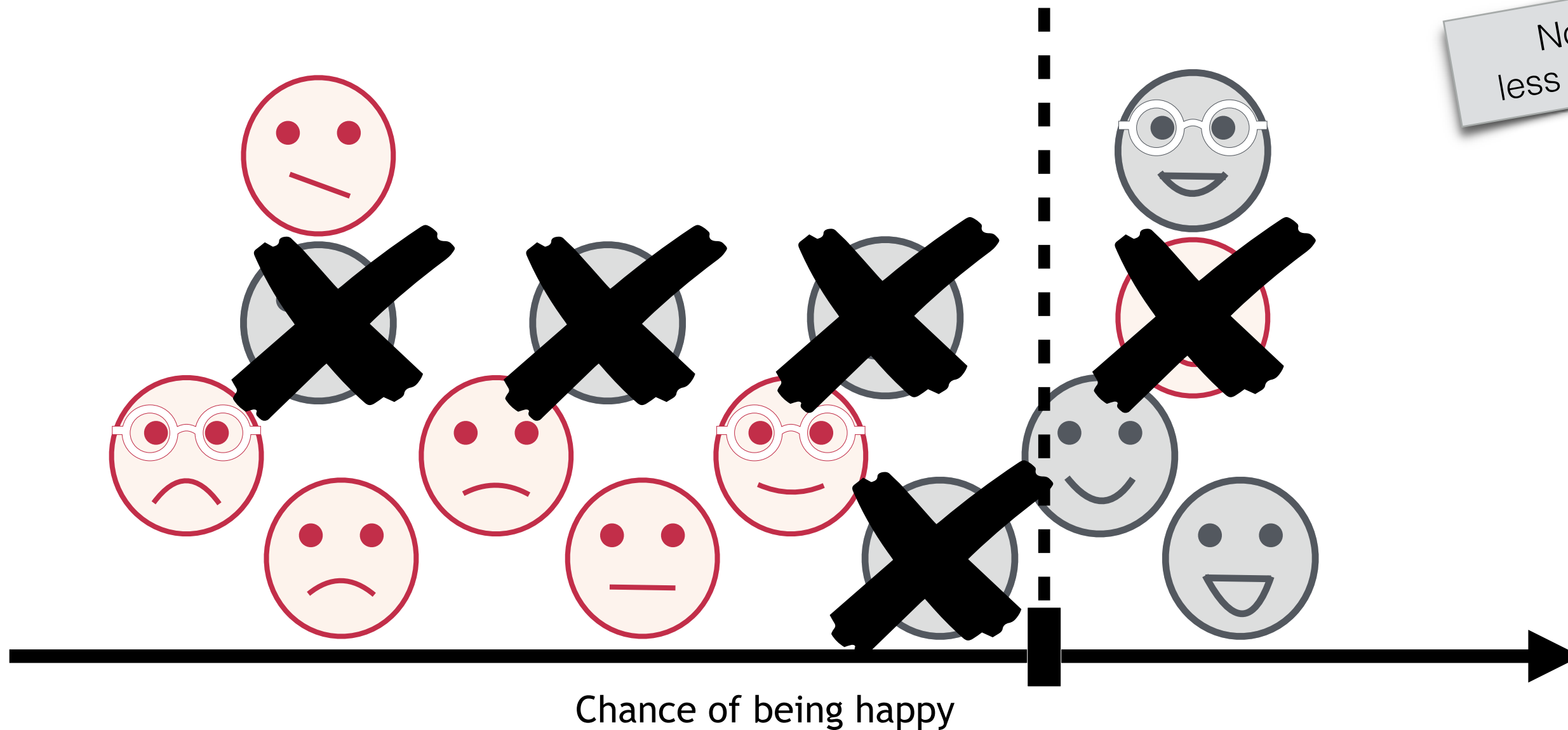
TUNING THE ERRORS

sad

happy

Beware that **class proportions** may vary over time, and affect the errors.

Now we have
less sad customers



TUNING THE ERRORS

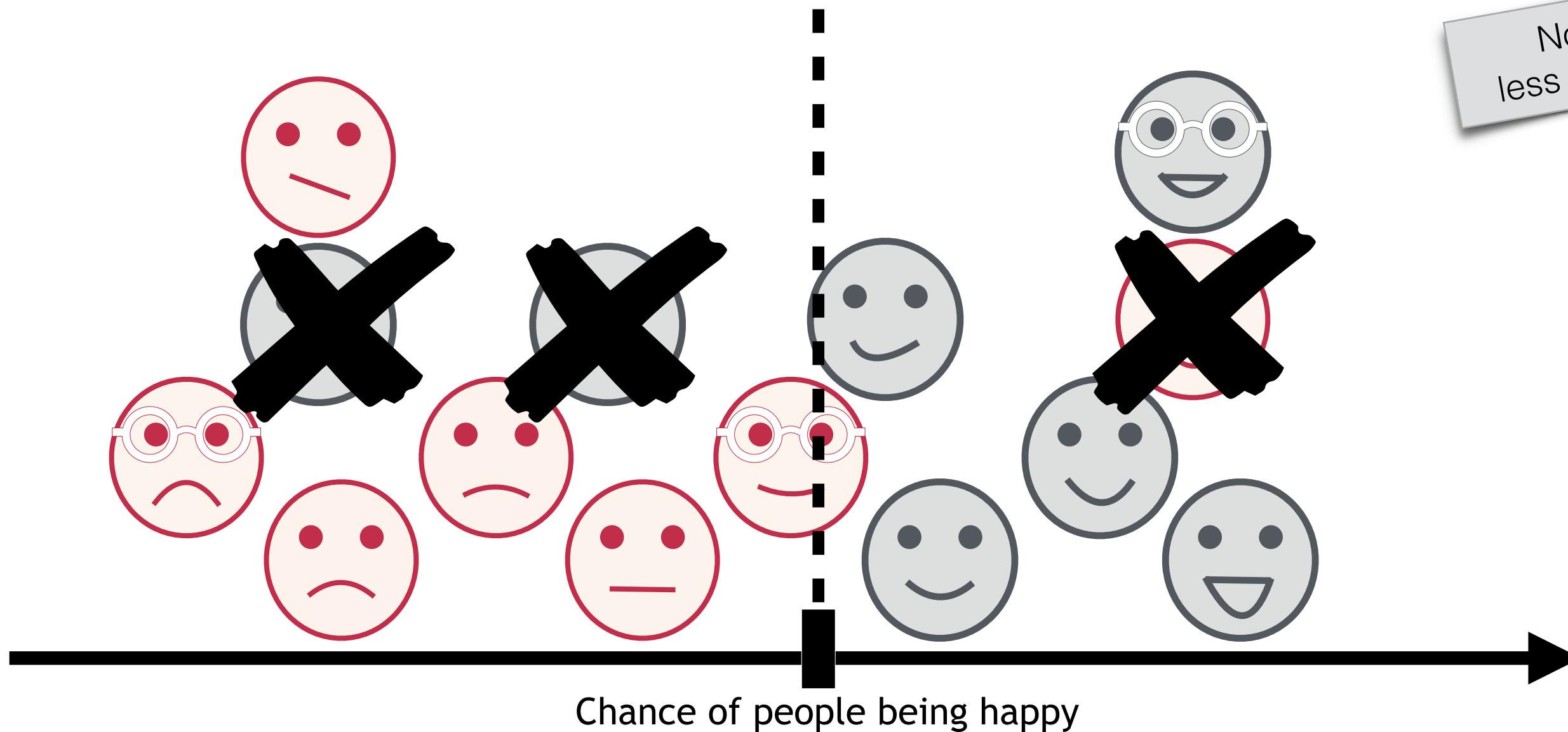
sad

:

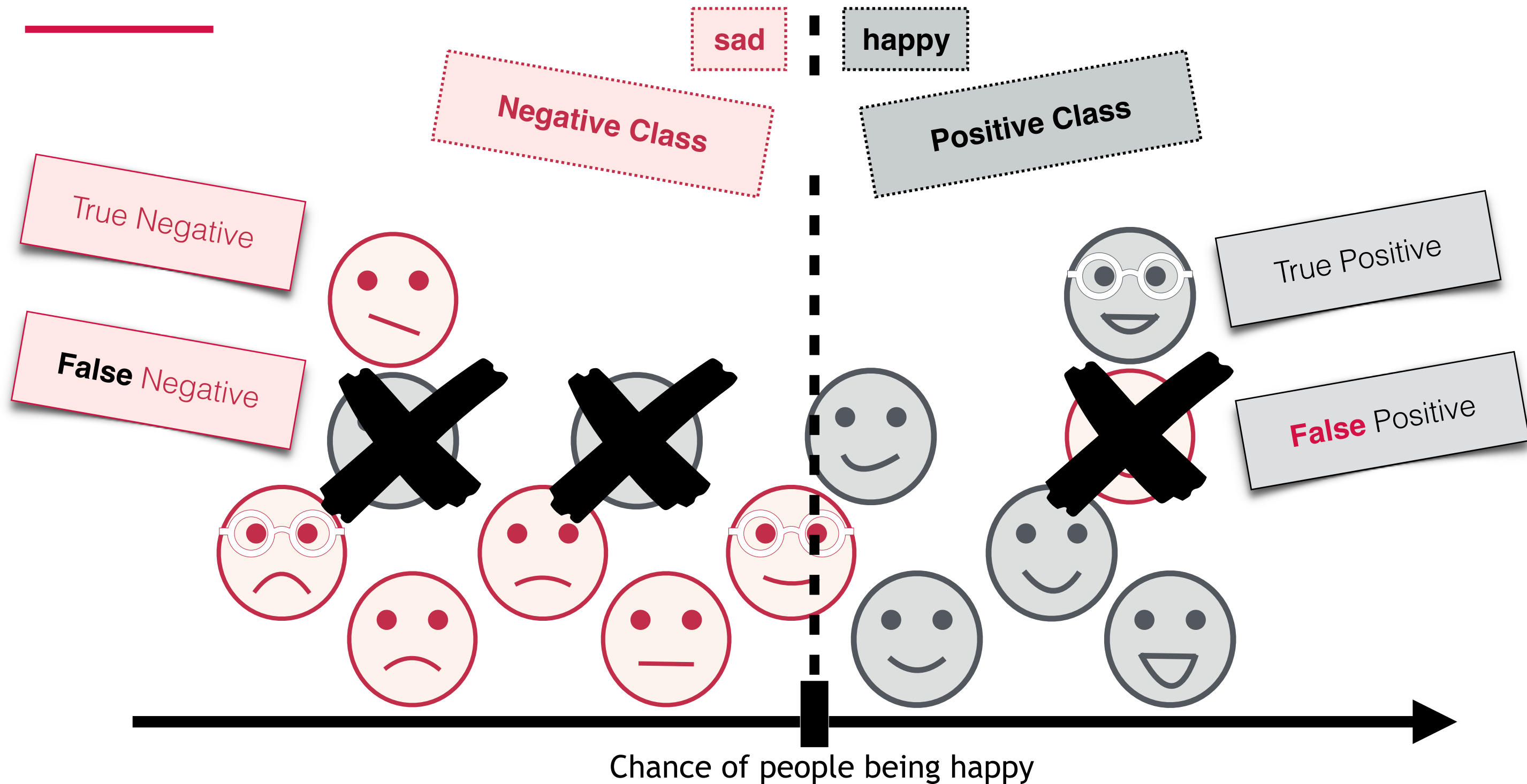
happy

Beware that **class proportions** may vary over time, and affect the errors.

Now we have
less sad customers



TERMINOLOGY



ERROR AND BIAS

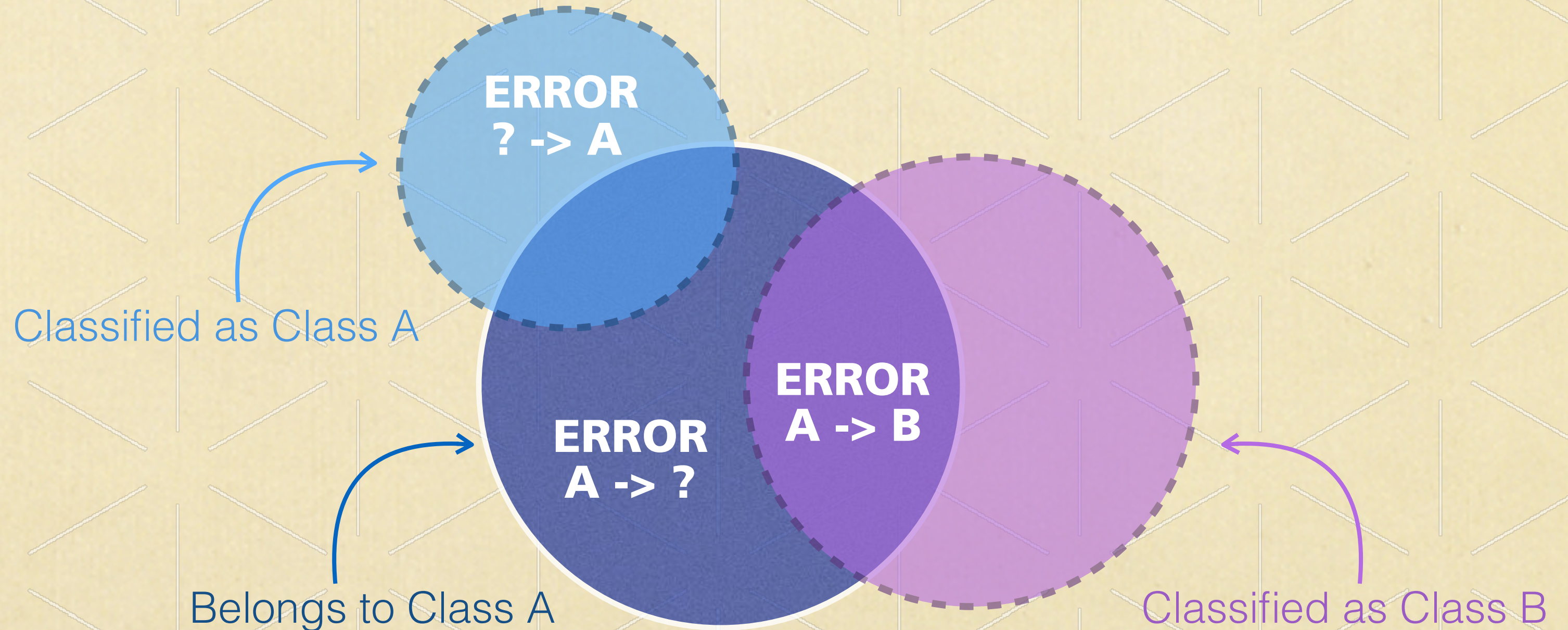
EMMA BEAUXIS-AUSSALET

e.m.a.l.beauxis@hva.nl

ERROR

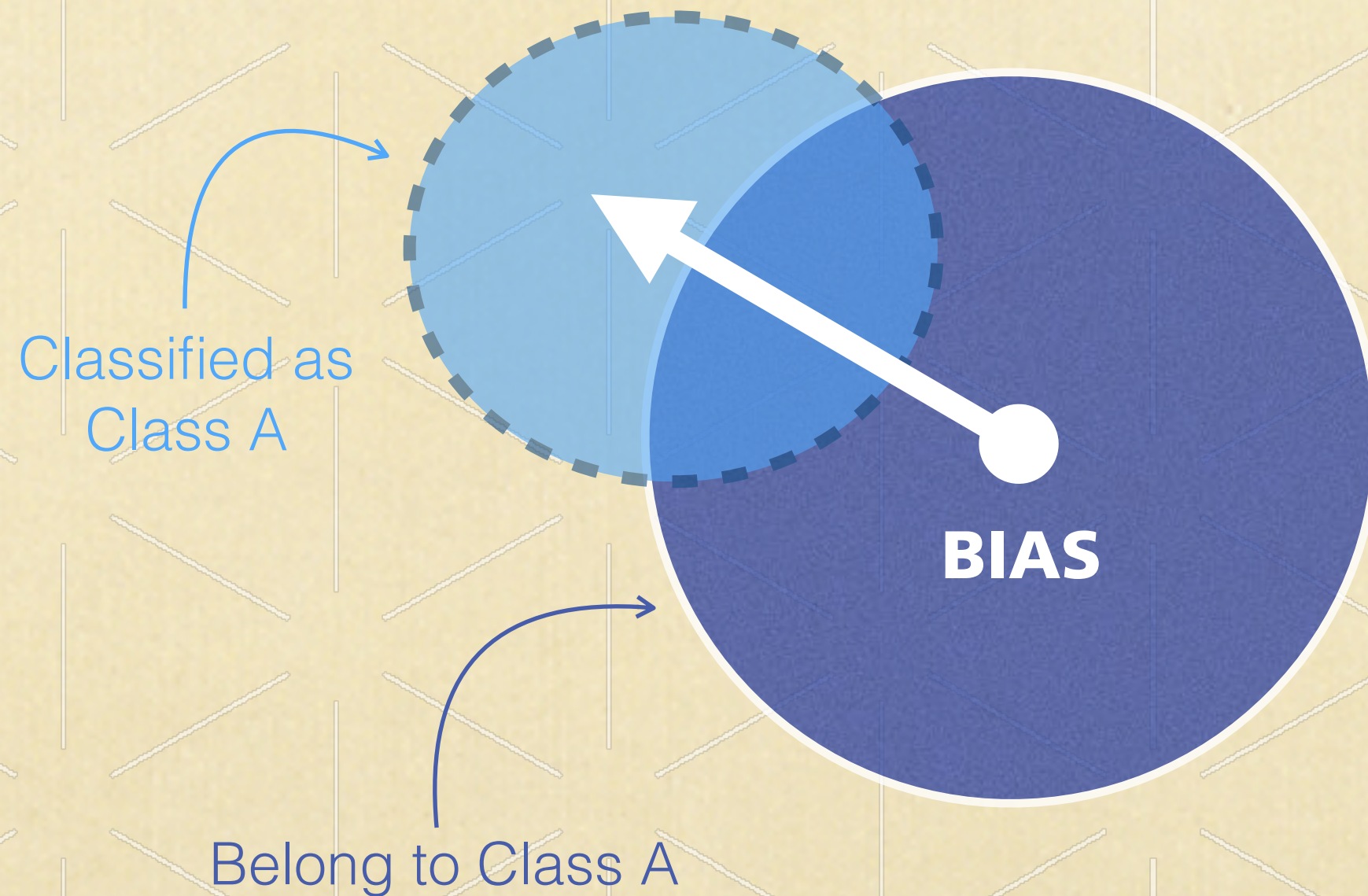
What belongs to Class A is classified as Class B.

It is **directional**: objects from Class A are misclassified into Class B.



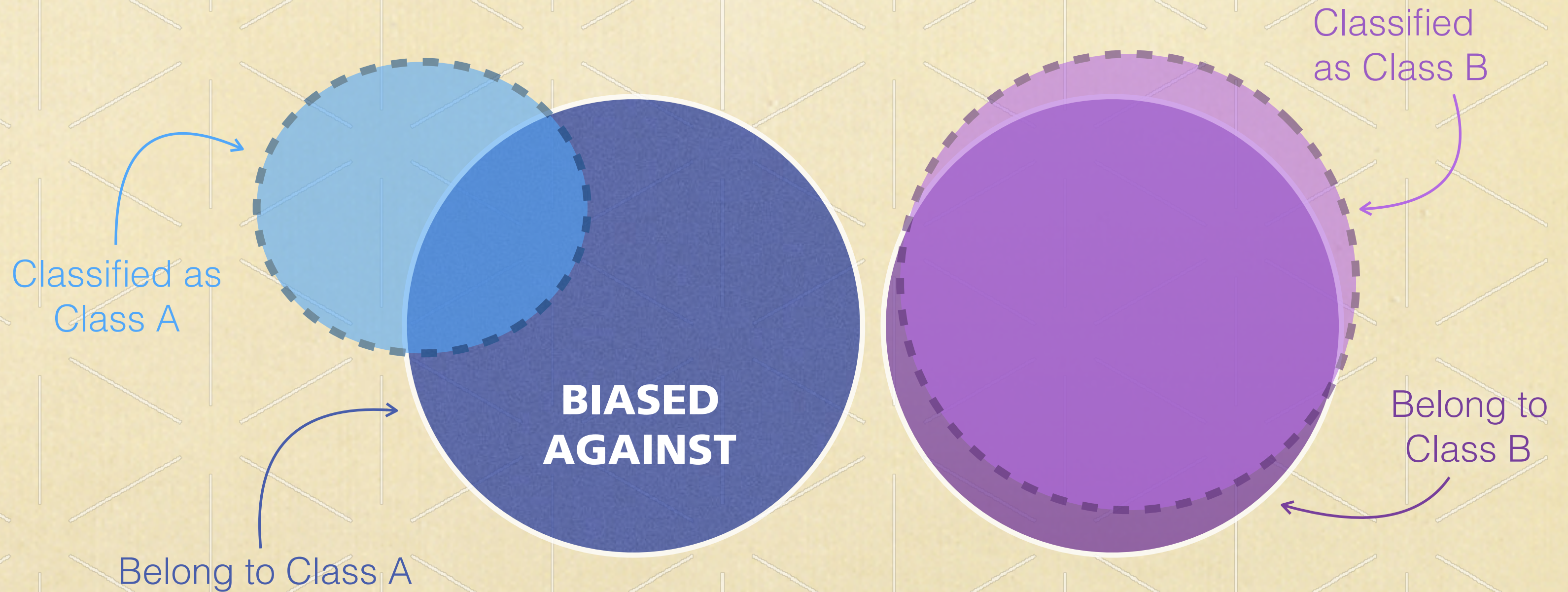
BIAS (1)

Errors systematically occur **for one class**.



BIAS (2)

Errors systematically differ **between classes**.



BIAS (3)

Errors systematically differ **between groups**.



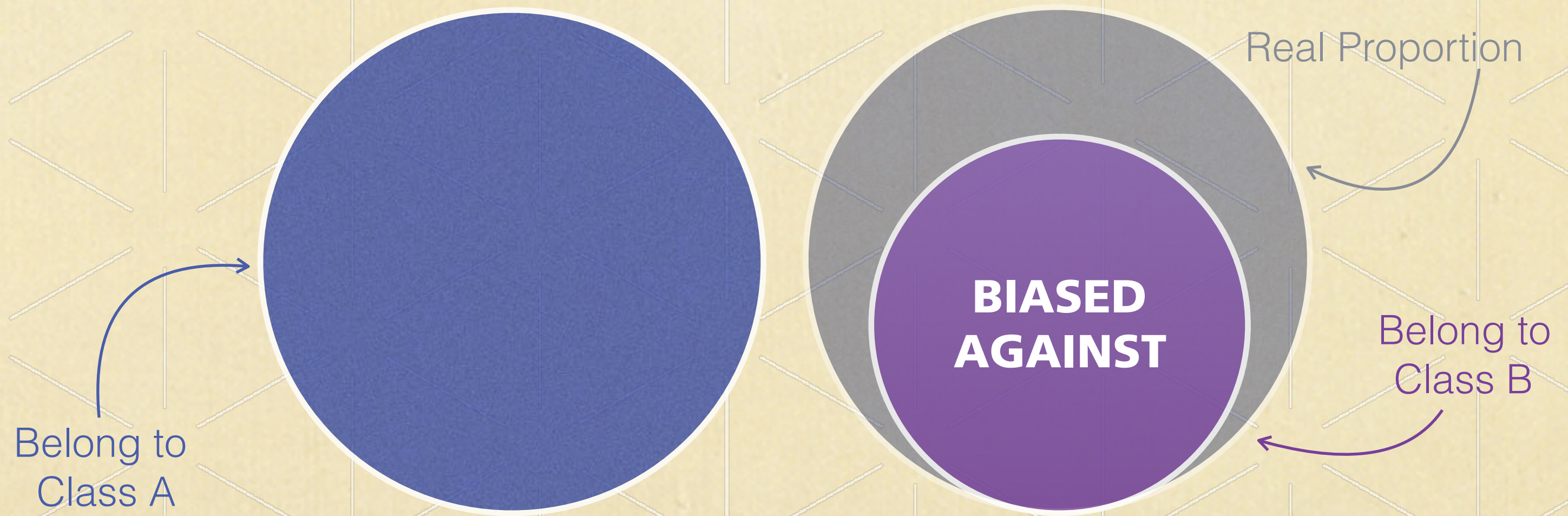
BIAS (4)

Groups are **under-represented** (either in test or training set).
Classification models are more uncertain about them.



BIAS (5)

Classes are **under-represented** (either in test or training set).
Classification models are more uncertain about them.



MEASURE ERROR

EMMA BEAUXIS-AUSSALET

e.m.a.l.beauxis@hva.nl

BEWARE OF **ERROR RATES**

We may forget what **denominators** represent.
(experts or non-experts)

$$\textit{Error Rate} = \frac{\textit{Number of Error}}{\text{ON WHAT?!}}$$

THAT'S VERY IMPORTANT...

[1] Hoffrage, Krauss, Martignon, and Gigerenzer (2015)

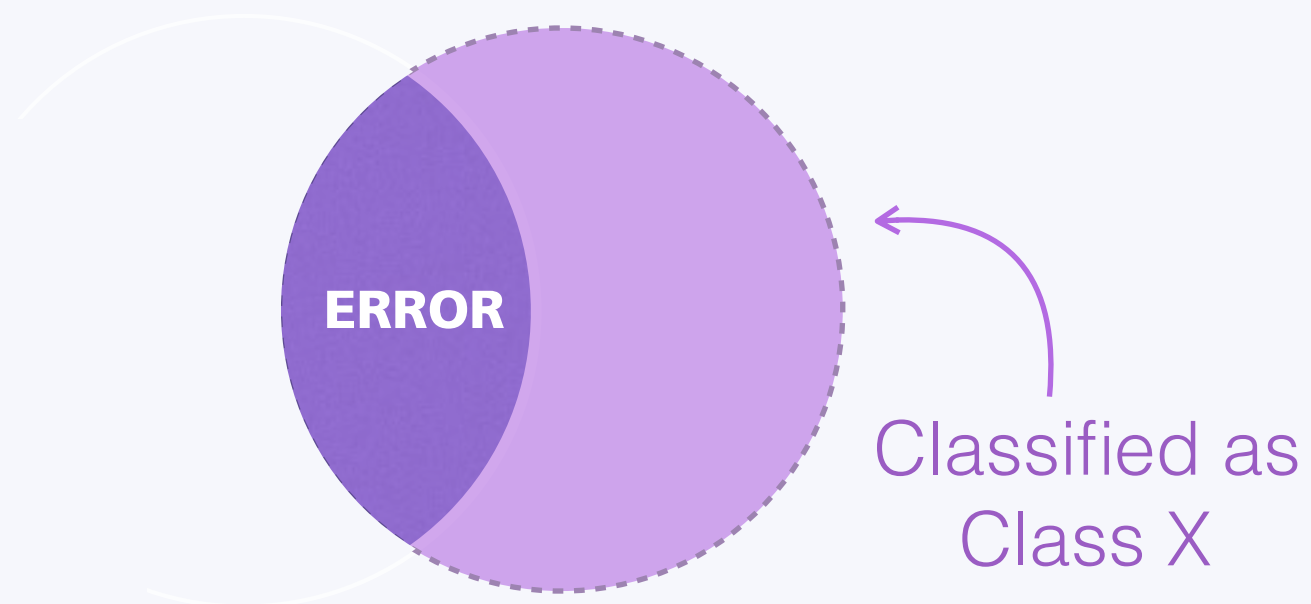
[2] Khan, Breslav, Glueck, and Hornbæk (2015)

[3] Beauxis-Aussalet, van Doorn, and Hardman (2018)

BEWARE OF PRECISION

If **class proportions** vary, precision varies.

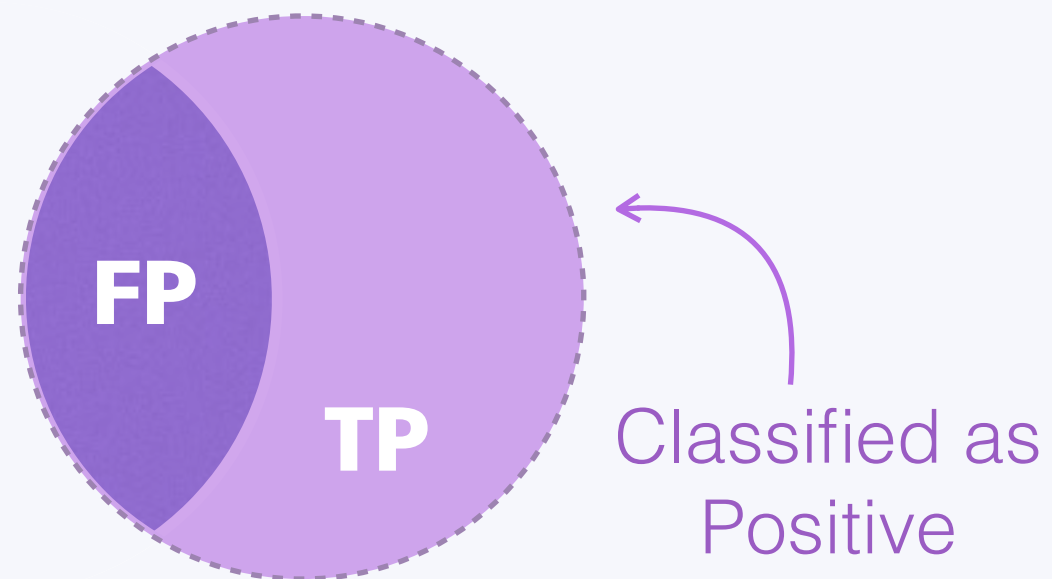
$$\text{Precision} = \frac{\text{Number of Objects Correctly Classified as } X}{\text{Total Number of Objects Classified as } X}$$



BEWARE OF PRECISION

If **class proportions** vary, precision varies.

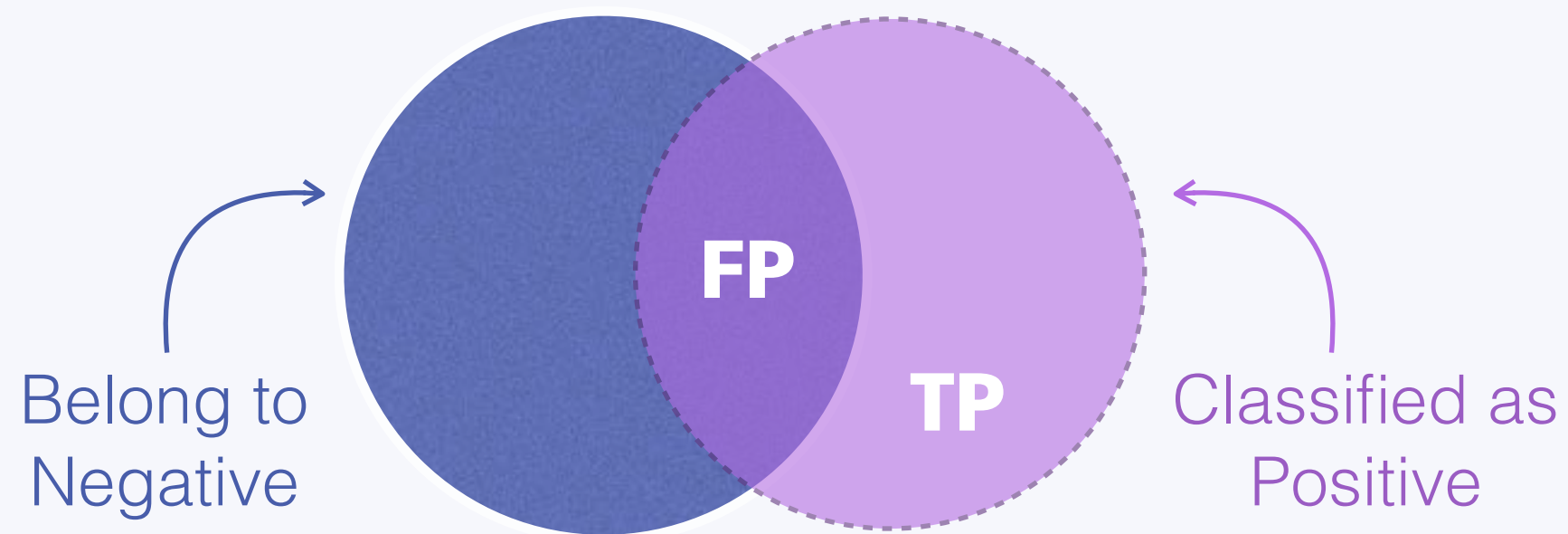
$$\textit{Precision} = \frac{TP}{TP + FP}$$



BEWARE OF PRECISION

If **class proportions** vary, precision varies.

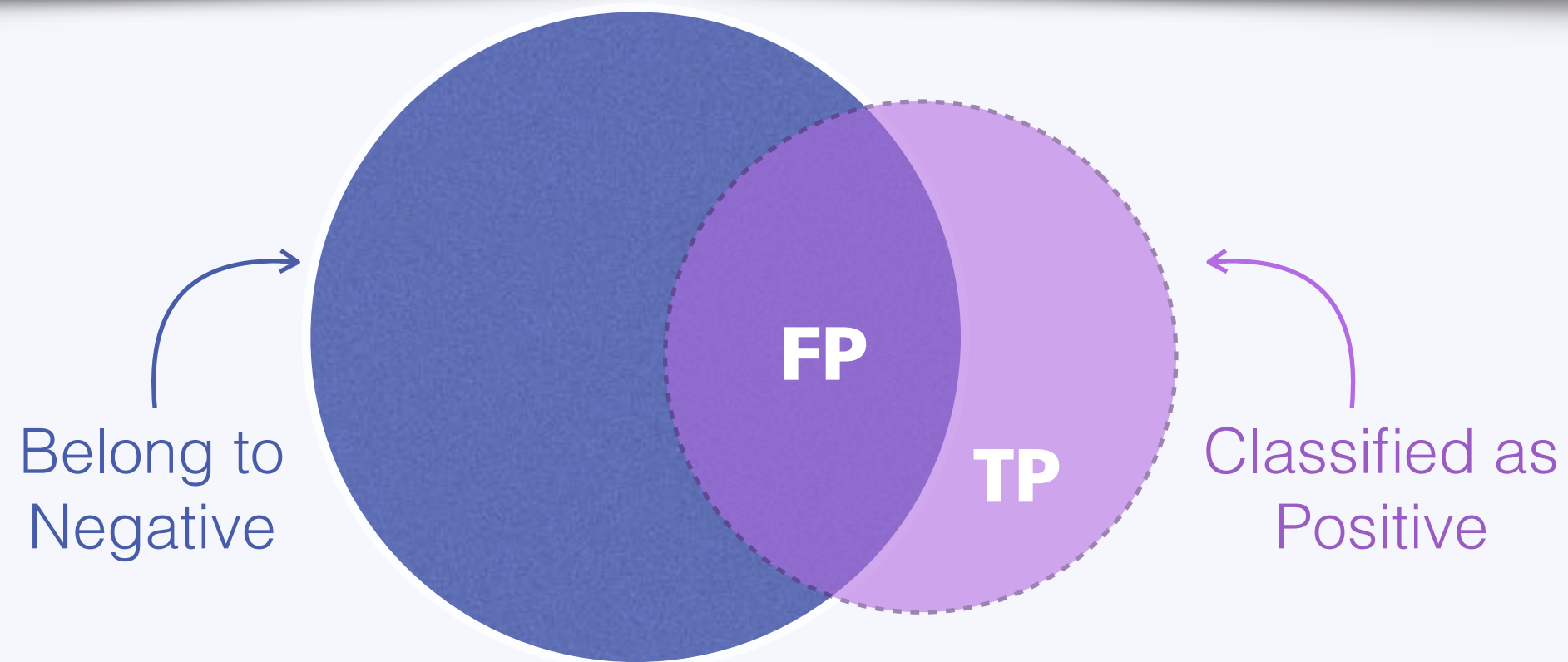
$$\textit{Precision} = \frac{TP}{TP + FP} = \frac{\text{purple crescent}}{\text{purple circle}}$$



BEWARE OF PRECISION

If **class proportions** vary, precision varies.

$$\textit{Precision} = \frac{TP}{TP + FP} = \frac{\text{purple crescent}}{\text{purple circle}}$$



BEWARE OF PRECISION

If **class proportions** vary, precision varies.

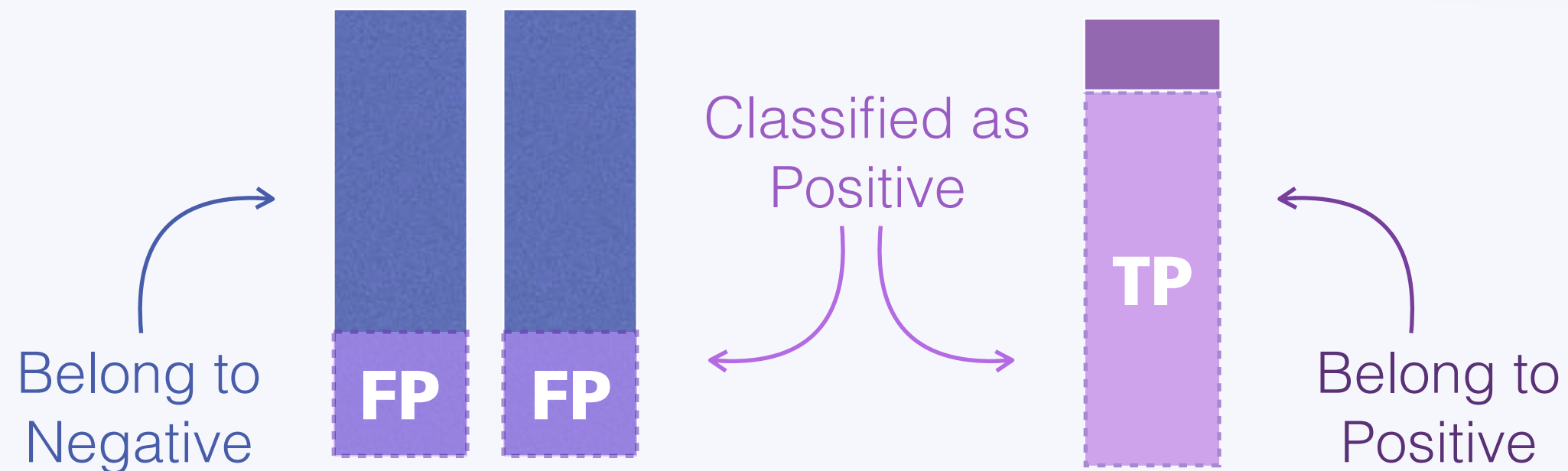
$$\textit{Precision} = \frac{TP}{TP + FP} = \frac{\text{[purple box]}}{\text{[purple box] + [blue box]}}$$



BEWARE OF PRECISION

If **class proportions** vary, precision varies.

$$\textit{Precision} = \frac{TP}{TP + FP} = \frac{\text{[purple box]}}{\text{[purple box] + [blue box] + [blue box]}}$$



BEWARE OF PRECISION

If **class proportions** vary, precision varies.

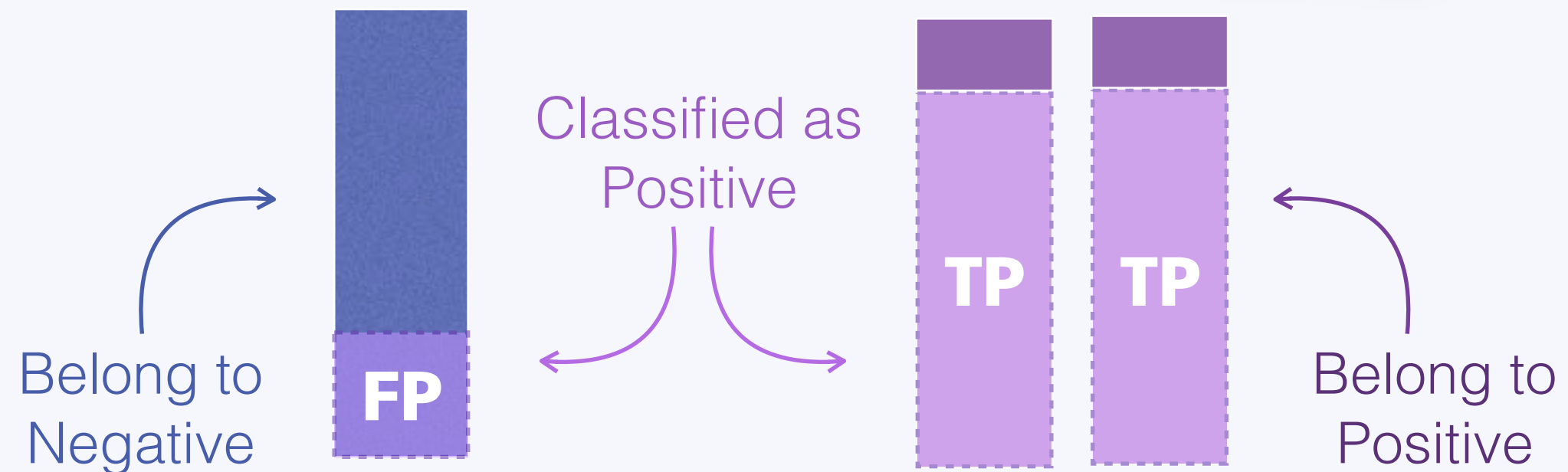
$$\textit{Precision} = \frac{TP}{TP + FP} = \frac{\text{[purple box]}}{\text{[purple box] + [blue box]}}$$



BEWARE OF PRECISION

If **class proportions** vary, precision varies.

$$\textit{Precision} = \frac{TP}{TP + FP} = \frac{\text{[2 purple boxes]}}{\text{[2 purple boxes] + [1 dark purple box]}}$$

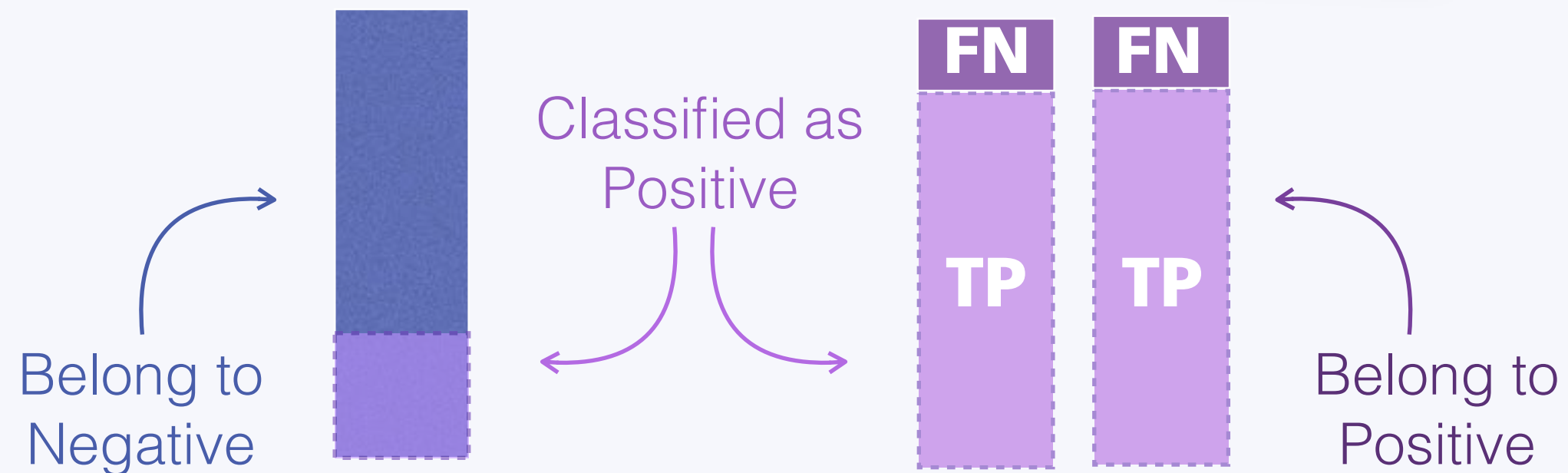


BEWARE OF CLASS PROPORTIONS

Some error rates remain unchanged by class proportions.

e.g., those of ROC curves: TP Rate (a.k.a. Recall) and FP Rate.

$$TP\ Rate = \frac{TP}{TP + FN} = \frac{\text{[purple box]}}{\text{[purple box] + [dark purple box]}}$$



BEWARE OF CLASS PROPORTIONS

Some error rates remain unchanged by class proportions.

e.g., those of ROC curves: TP Rate (a.k.a. Recall) and FP Rate.

$$TP\ Rate = \frac{TP}{TP + FN} = \frac{\text{[purple box]}}{\text{[purple box] + [dark purple box]}}$$



BEWARE OF ACCURACY

It fails to indicate the errors for specific classes or groups.

It varies with class proportions. Biases cannot be identified.

$$\text{Accuracy} = \frac{\text{Number of Correct Classifications}}{\text{Total Number of Classifications}} = \frac{\text{[purple box]} + \text{[purple box]} + \text{[dark purple box]} + \text{[blue box]}}{\text{[purple box]} + \text{[purple box]} + \text{[dark purple box]} + \text{[blue box]}}$$

Example: We achieve 98% accuracy, but one class is entirely misclassified (a small class $\leq 2\%$ of all objects).

ERROR IN END-RESULTS

EMMA BEAUXIS-AUSSALET

e.m.a.l.beauxis@hva.nl

HOW MANY ERRORS IN THE END-RESULTS?

Precision = $\frac{TP}{TP + FP} = 0.95$

FP Rate = $\frac{FP}{FP + TN} = 0.1$

	Positive	Negative
Elements assigned to each class	1000	800

HOW MANY ERRORS IN THE END-RESULTS?

$$\text{Precision} = \frac{TP}{TP + FP} = 0.95$$

$$\text{FP Rate} = \frac{FP}{FP + TN} = 0.1$$

	Positive	Negative
Elements assigned to each class	1000	800

$$\text{TP} = \text{Precision} \times 1000$$

$$= \frac{TP}{TP + FP} \times TP + FP$$

$$\text{TP} = 0.95 \times 1000 = 950$$

$$\text{FP} = 1000 - 950 = 50$$

HOW MANY ERRORS IN THE END-RESULTS?

$$\text{Precision} = \frac{TP}{TP + FP} = 0.95$$

$$\text{FP Rate} = \frac{FP}{FP + TN} = 0.1$$

	Positive	Negative
Elements assigned to each class	1000	800

$$\text{TP} = \text{Precision} \times 1000$$

$$= \frac{TP}{TP + FP} \times TP + FP$$

$$\text{TP} = 0.95 \times 1000 = 950$$

$$\text{FP} = 1000 - 950 = 50$$

$$\text{FP Rate} = \frac{FP}{FP + TN} = 0.1$$

$$\frac{50}{50 + \text{TN}} = 0.1$$

$$\frac{50 + \text{TN}}{50} = \frac{1}{0.1}$$

$$50 + \text{TN} = \frac{50}{0.1}$$

$$\text{TN} = \frac{50}{0.1} - 50 = 450$$

$$\text{FN} = 800 - 450 = 350$$

HOW MANY ERRORS IN THE END-RESULTS?

$$\text{Precision} = \frac{TP}{TP + FP} = 0.95$$

$$\text{FP Rate} = \frac{FP}{FP + TN} = 0.1$$

	Positive	Negative
Elements assigned to each class	1000	800

$$\text{TP} = \text{Precision} \times 1000$$

$$= \frac{TP}{TP + FP} \times TP + FP$$

$$\text{TP} = 0.95 \times 1000 = 950$$

$$\text{FP} = 1000 - 950 = 50$$

$$\text{FP Rate} = \frac{FP}{FP + TN} = 0.1$$

$$\frac{50}{50 + \text{TN}} = 0.1$$

$$\frac{50 + \text{TN}}{50} = \frac{1}{0.1}$$

$$50 + \text{TN} = \frac{50}{0.1}$$

$$\text{TN} = \frac{50}{0.1} - 50 = 450$$

$$\text{FN} = 800 - 450 = 350$$

HOW MANY ERRORS IN THE END-RESULTS?

$$\text{Precision} = \frac{TP}{TP + FP} = 0.95$$

$$\text{FP Rate} = \frac{FP}{FP + TN} = 0.1$$

$$\text{FP Rate} = \frac{FP}{FP + TN} = 0.1$$

50

Elements assigned
to each class

**Beware low error rates
may mean high numbers of errors
in practice!**

$$\text{TP} = \text{Precision} \times 1000$$

$$50 + \text{TN} = \frac{50}{0.1}$$

$$= \frac{TP}{TP + FP} \times TP + FP$$

$$\text{TP} = 0.95 \times 1000 = 950$$

$$\text{TN} = \frac{50}{0.1} - 50 = 450$$

$$\text{FP} = 1000 - 950 = 50$$

$$\text{FN} = 800 - 450 = 350$$