Fairness sixpack

DEMOGRAPHIC PARITY $S \perp G$	WE'RE ALL EQUAL $T \perp G$	TARGET INDIFFERENCE $S \perp T$
EQUAL ODDS $S \perp_{T} G$	PREDICTIVE PARITY $T \perp_{S} G$	STRATIFIED INDIFFERENCE $S \perp_{G} T$

Thore Husfeldt
Lund University and ITU Copenhagen

Disclaimer

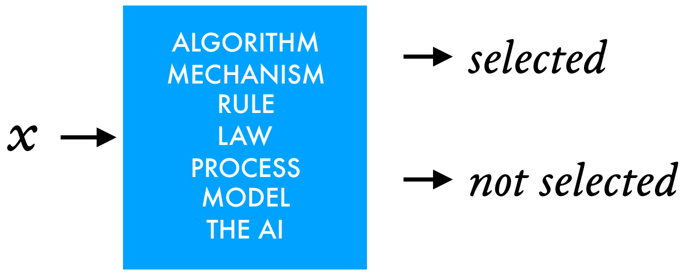
1. *No original results*
2. *I'm an algorithms person (in the classical sense), not an AI researcher*

Claimer

Honest attempt to present group fairness

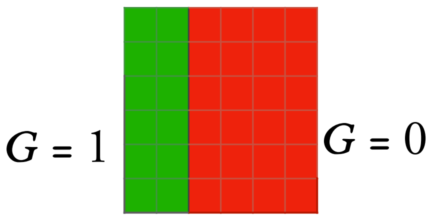
1. *minimal (as few concepts as possible)*
2. *complete & consistent (framework fully explored)*
3. *interesting (most phenomena appear)*
4. *accessible (first principles, dispassionate, value-free)*

Our Setup



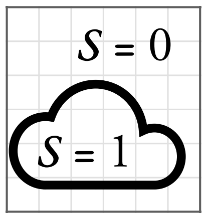
gets the job
gets the grant
goes to jail
sees the ad for diapers
is admitted to Harvard
is matched with Alice
gets the loan
pays full price
gets cheap insurance
is selected for security check
must show their passport

Group



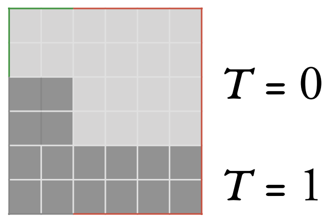
Greeks, Green.
(Romans are red.)

Selection



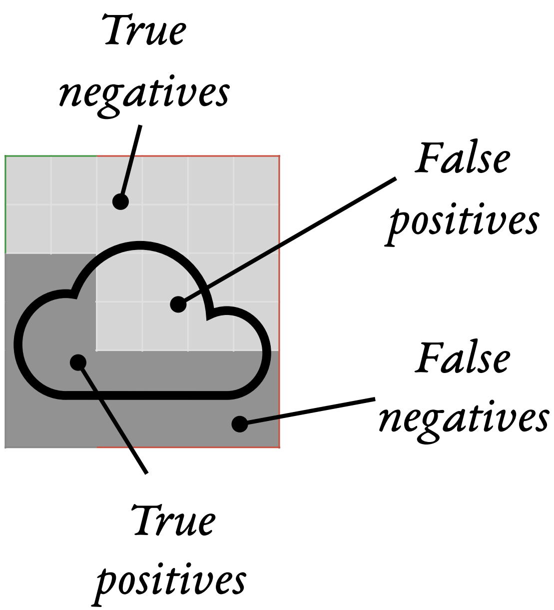
Scholarship,
Security check

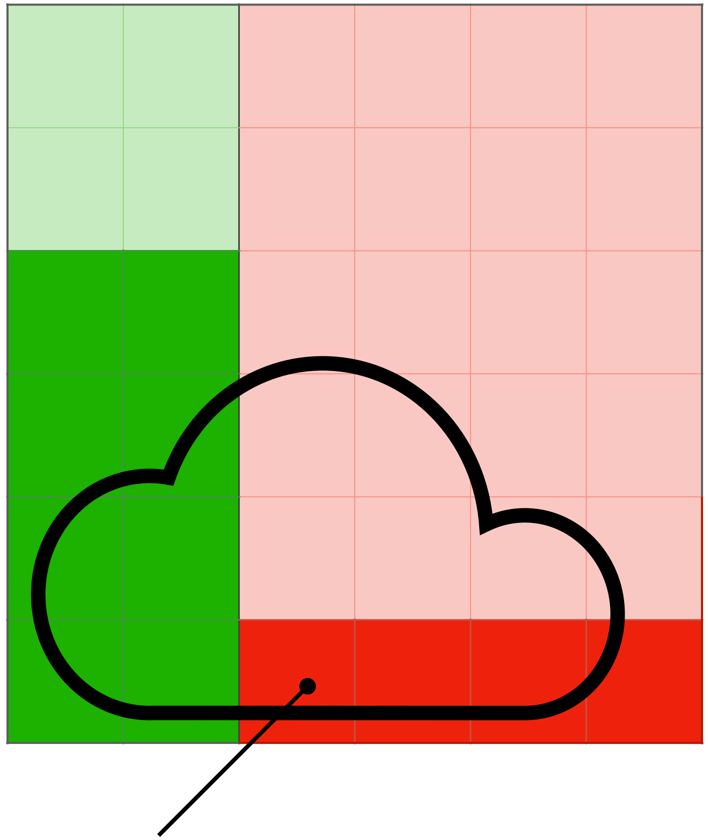
Target



Talent, Terrorist.
Drawn Tinted

Our Setup





Talented Romans selected for scholarship at Athenian School for Homerian Poetry
$G = 0, \ S = 1, \ T = 1$

Our Setup

Three random variables, $G, S, T \in \{0, 1\}$.

To make it interesting:
- individuals are different (T not constant)
- society is tribal (G not constant)
- not everybody is selected (S not constant)

Equality:
Maximum utility (select for target): $S = T$.
Maximum group bias (select for group membership): $G = S$.

Independence:
Recall: Random variables A, B are *independent* if for all a, b

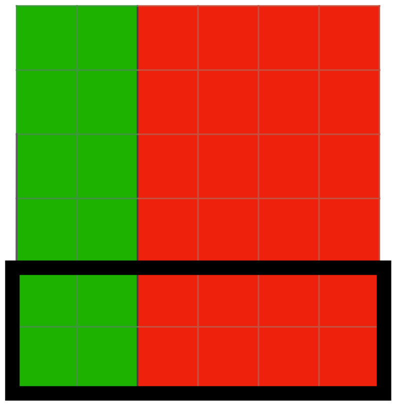
$$\Pr(A = a)\Pr(B = b) = \Pr(A = a \text{ and } B = b)$$

Demographic parity

Darlington's 4th fairness criterion • independence • group fairness • statistical parity • outcome equality • absence of disparate impact • equity

Selection is independent of group membership

$$S \perp G$$



Very widespread intuition, often the *only*. Absence of demographic parity often seen as indication of discrimination. Societies adopt and evaluate policies to achieve demographic parity.

How to achieve: quotas

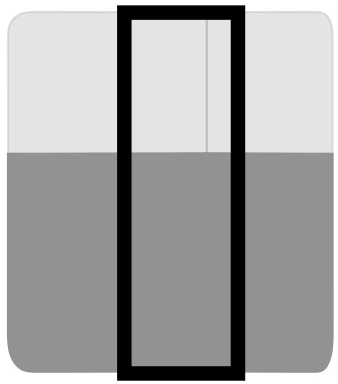
+ equal representation

– target value does not even appear

Target indifference

Selection is independent of target value

$$S \perp T$$



How to achieve: flip (weighted) coin, drawing of lots, *sortition*

Widely used (military draft, jury selection, breaking ties in elections)

+ no corruption, "sanitizing effect of ignorance," no one to blame, no boasting

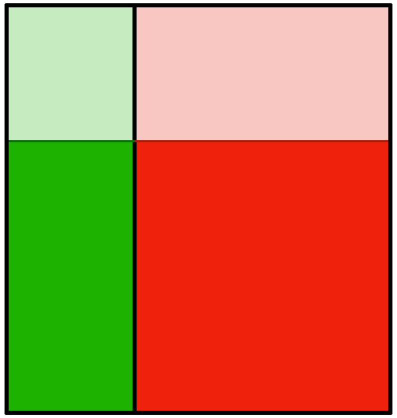
– no utility, lack of agency-induced dignity, lack of motivation, uncertainty, undermines control

We're all equal

equal base rates • no group differences

Target value is independent of group membership

$$T \perp G$$

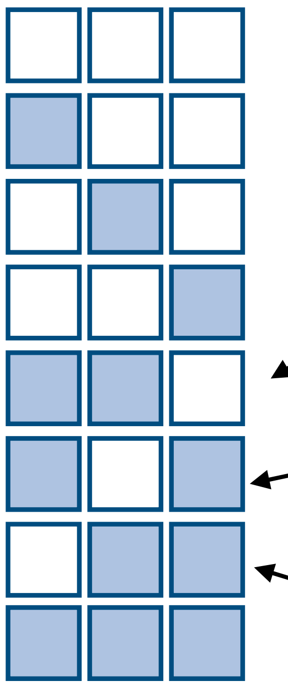


Underlying assumption in many social theories.

How to achieve: out of scope.
(Not a property of selector, but instead claim about population.)

Notions based on independence

OUTCOME	REALITY	UTILITY
DEMOGRAPHIC PARITY	WE'RE ALL EQUAL	TARGET INDIFFERENCE
$S \perp G$	$T \perp G$	$S \perp T$



	Equal outcomes	*No group differences*	*Flawless prediction*
Perfect world:	$S \perp G$	$T \perp G$	$S = T$
Vonnegut dystopia	$S \perp G$	$T = G$	$S \perp T$
Inefficient prejudice	$S = G$	$T \perp G$	$S \perp T$

Conditional independence

$$A \perp B \mid C \quad \text{or} \quad (A \perp B) \mid C \quad \text{or} \quad A \underset{C}{\perp} B$$

Def.: Random variables A, B are *conditionally independent* given C, if for all a, b, c

$$\Pr(A = a \mid C = c) \Pr(B = b \mid C = c) = \Pr(A = a \text{ and } B = b \mid C = c)$$

(Knowledge about whether C occurs "makes A and B independent".)

Ex.: Claire has two coins (one fair, the other always comes up "heads"). She picks one and lets both Alice and Bob flip it.
Then A ⊥ B is false, but A ⊥ B | C is true.

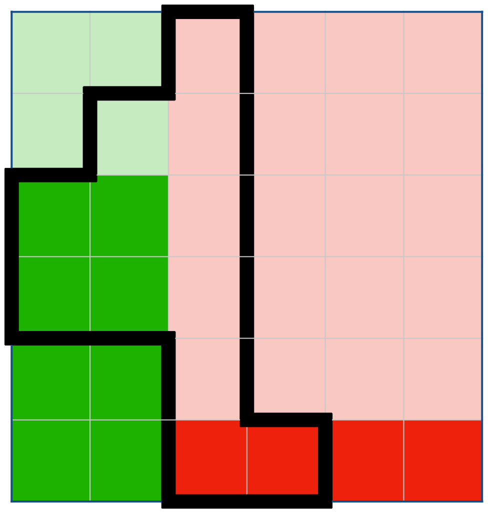
Equal odds

separation • no disparate (mis)treatment

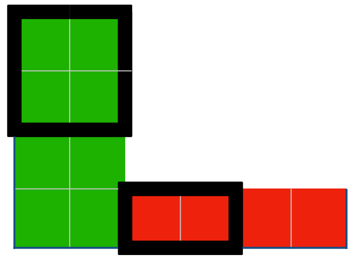
Conditioned on target value, demographic parity holds.

Intuition: no error type disproportionally affects any group.
IOW, your group membership does not influence
your chance of (de)selection.

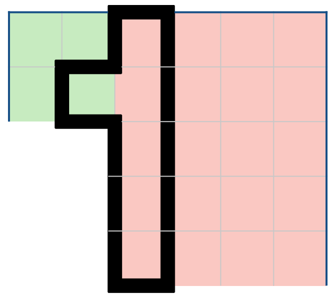
$$S \perp_{T} G$$



T = 1



T = 0

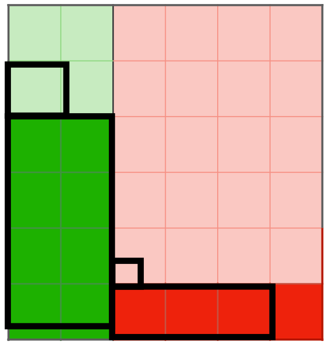


Predictive parity

sufficiency • Cleary model • predictive rate parity • employer's fairness • equally good/bad prediction • well-calibration within groups

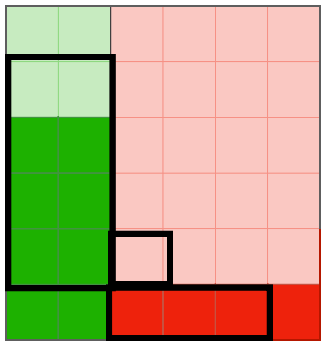
Conditioned on selection, we're all equal.
For teacher at selective school, base rates look equal.
(E.g., diploma is equally predictive for all graduates.)

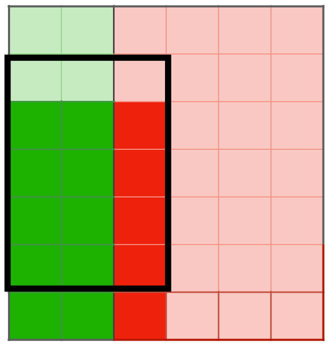
$$T \underset{S}{\perp} G$$



Eeasier to visualise: "positive predictive parity"

$$T \underset{S=1}{\perp} G$$



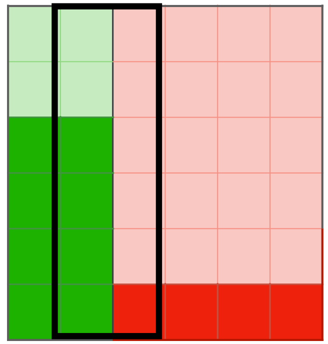


Stratified indifference

Target indifference conditioned on group membership

Within groups, selection is independent of target values.

$$S \perp_{G} T$$

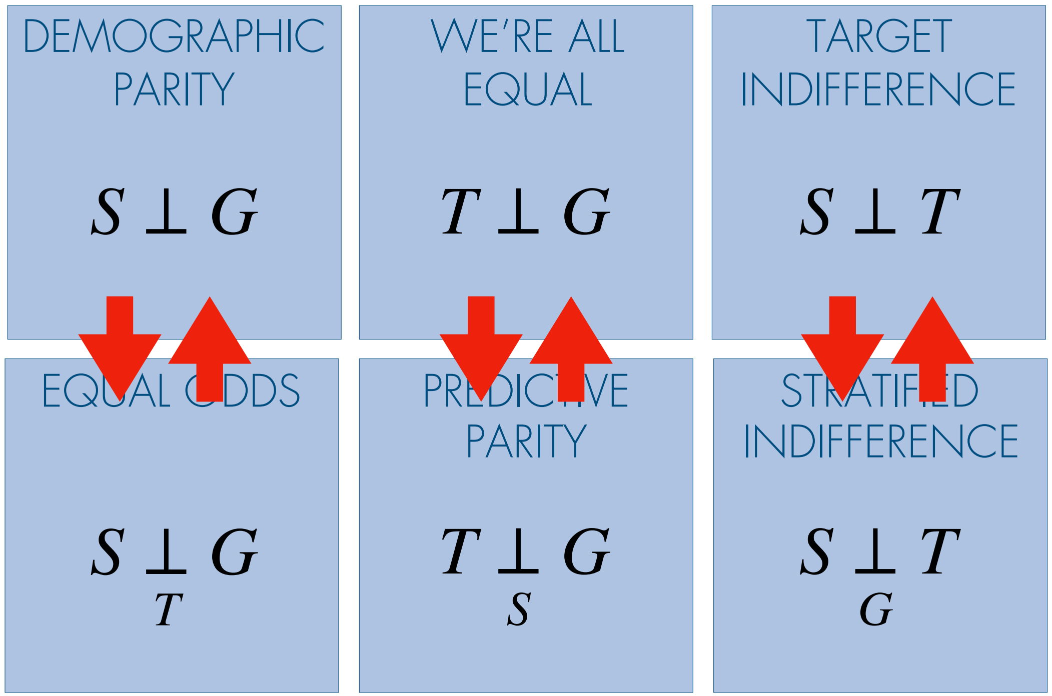


Example: Stratified sortition, such as military draft.

Fairness sixpack

DEMOGRAPHIC PARITY $S \perp G$	WE'RE ALL EQUAL $T \perp G$	TARGET INDIFFERENCE $S \perp T$
EQUAL ODDS $S \perp G \mid T$	PREDICTIVE PARITY $T \perp G \mid S$	STRATIFIED INDIFFERENCE $S \perp T \mid G$

Independence versus Conditional Independence

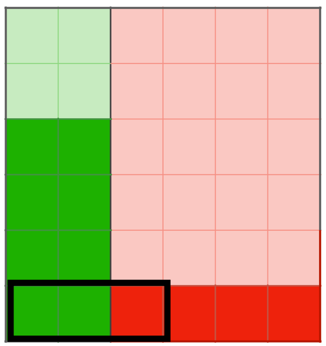


Q: Which notion is weaker?

Independence versus Conditional Independence

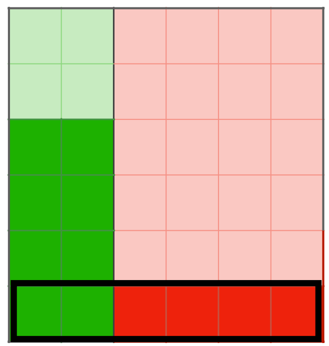
$$A \underset{C}{\perp} B \overset{?}{\Leftrightarrow} A \perp B$$

$S \underset{T}{\perp} G$ **but not** $S \perp G$



equal odds achieved, demographic parity violated

$S \perp G$ **but not** $S \underset{T}{\perp} G$



demographic parity achieved, equal odds violated

Fairness sixpack

DEMOGRAPHIC PARITY $S \perp G$	WE'RE ALL EQUAL $T \perp G$	TARGET INDIFFERENCE $S \perp T$
EQUAL ODDS $S \underset{T}{\perp} G$	PREDICTIVE PARITY $T \underset{S}{\perp} G$	STRATIFIED INDIFFERENCE $S \underset{G}{\perp} T$

Q: which ones would you like?

Assume…

DEMOGRAPHIC PARITY $S \perp G$	WE'RE ALL EQUAL $T \perp G$	TARGET INDIFFERENCE $S \perp T$
EQUAL ODDS $S \perp G \mid T$	PREDICTIVE PARITY $T \perp G \mid S$	STRATIFIED INDIFFERENCE $S \perp T \mid G$

Q: Is this even possible? If so, what happens?

Prop.

$$A \perp B \;\wedge\; A \underset{C}{\perp} B \quad\Rightarrow\quad B \perp C \;\wedge\; B \perp C$$

Proof. (Half a page of simple discrete probability.)

If

DEMOGRAPHIC PARITY

$S \perp G$

EQUAL ODDS

$S \underset{T}{\perp} G$

then

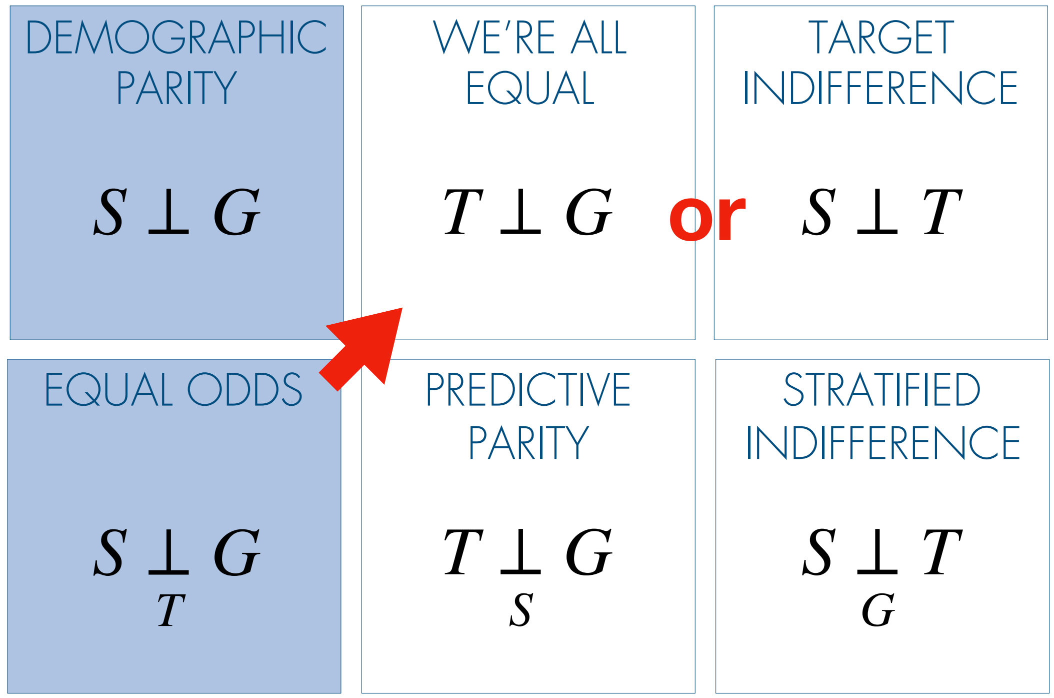
WE'RE ALL EQUAL

$T \perp G$

or

TARGET INDIFFERENCE

$S \perp T$


DEMOGRAPHIC PARITY
$S \perp G$
WE'RE ALL EQUAL
$T \perp G$
or
TARGET INDIFFERENCE
$S \perp T$
EQUAL ODDS
$S \perp_T G$
PREDICTIVE PARITY
$T \perp_S G$
STRATIFIED INDIFFERENCE
$S \perp_G T$

What about...

DEMOGRAPHIC PARITY $S \perp G$	WE'RE ALL EQUAL $T \perp G$	TARGET INDIFFERENCE $S \perp T$
EQUAL ODDS $S \perp G \mid T$	PREDICTIVE PARITY $T \perp G \mid S$	STRATIFIED INDIFFERENCE $S \perp T \mid G$

$$A \perp B \ \wedge\ A \underset{B}{\perp} C \quad \Rightarrow \quad A \perp C$$

DEMOGRAPHIC PARITY $S \perp G$	WE'RE ALL EQUAL $T \perp G$	TARGET INDIFFERENCE $S \perp T$
EQUAL ODDS $S \underset{T}{\perp} G$	PREDICTIVE PARITY $T \underset{S}{\perp} G$	STRATIFIED INDIFFERENCE $S \underset{G}{\perp} T$

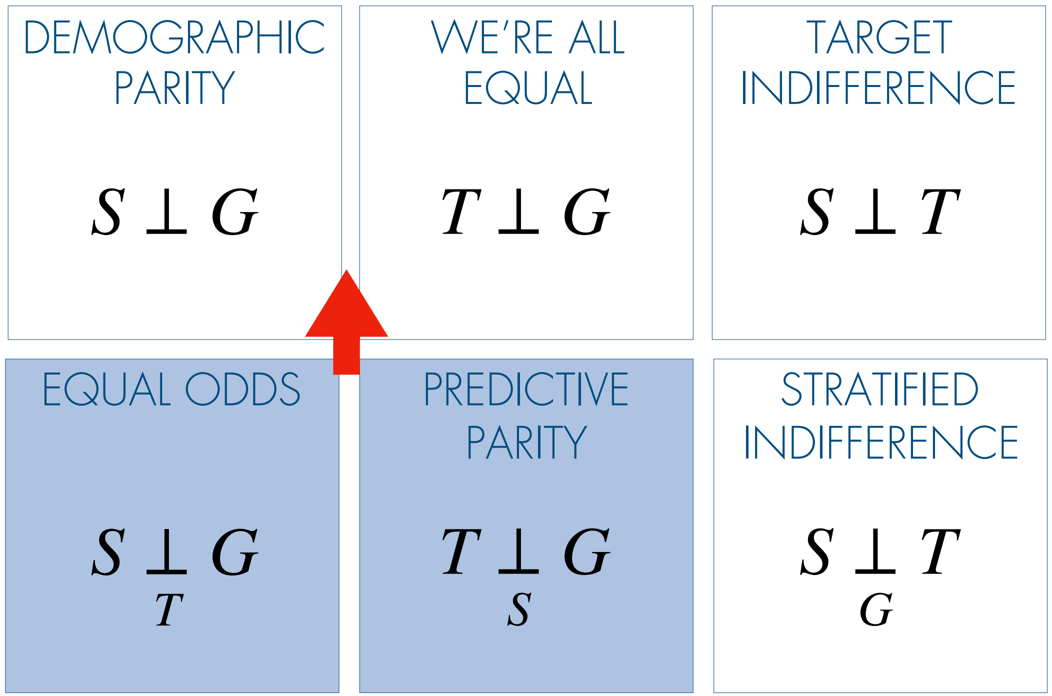
What about...

DEMOGRAPHIC PARITY $S \perp G$	WE'RE ALL EQUAL $T \perp G$	TARGET INDIFFERENCE $S \perp T$
EQUAL ODDS $S \perp_{T} G$	PREDICTIVE PARITY $T \perp_{S} G$	STRATIFIED INDIFFERENCE $S \perp_{G} T$

Proposition. *If* $(A \perp B) | C$ and $(A \perp C) | B$ *then* $A \perp B$ and $A \perp C$.

(Except for pathological cases such as C = 1 etc.)

Proof. (Half a page.)



Fairness sixpack

DEMOGRAPHIC PARITY $S \perp G$	WE'RE ALL EQUAL $T \perp G$	TARGET INDIFFERENCE $S \perp T$
EQUAL ODDS $S \perp_{T} G$	PREDICTIVE PARITY $T \perp_{S} G$	STRATIFIED INDIFFERENCE $S \perp_{G} T$

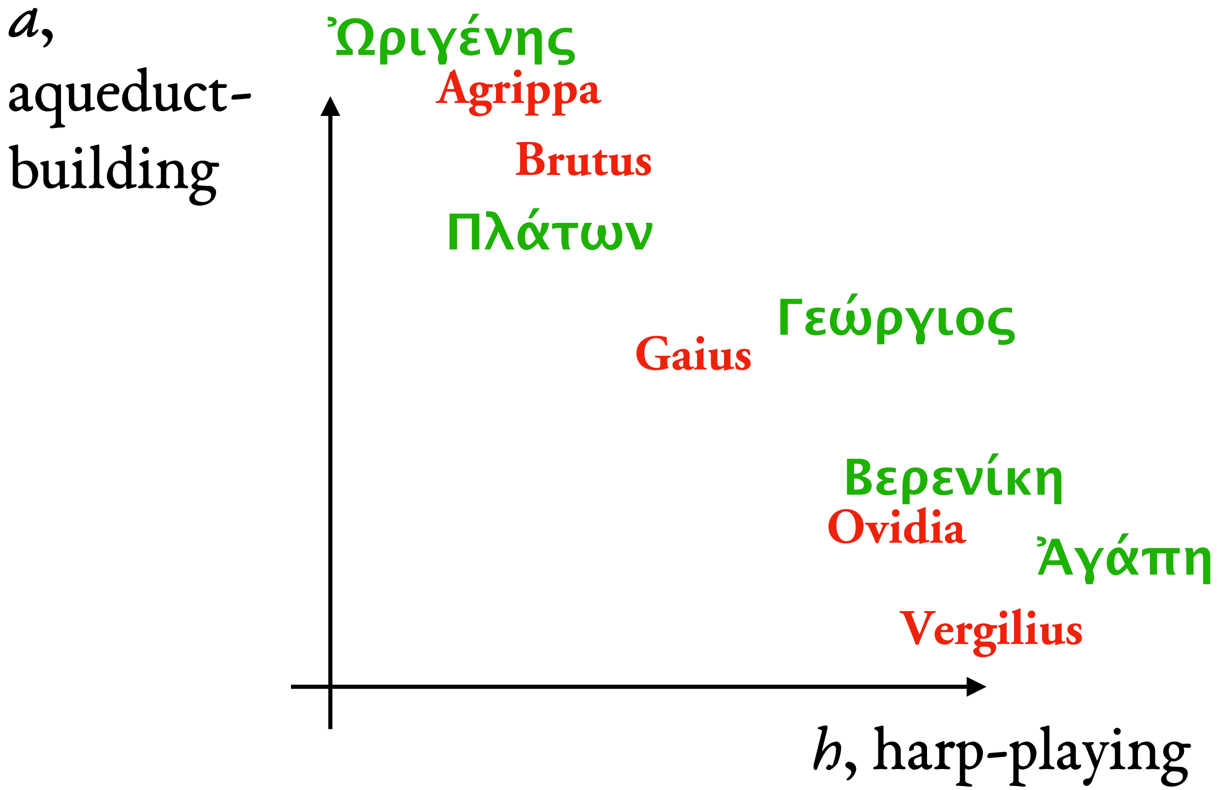
TL;DR: You can have one.

Blindness Objection

"Q: The *good* reason for this at all is to detect injustice, in particular group bias.
So why not detect injustice in *process*, rather than *outcome?*
Just prevent selector from access to group membership."

Everything correlates.

More interestingly, assume we know $x = (g, a, h)$



Summary & Conclusion

1. *Told most of the story within conceptually tiny framework—far from comprehensive!*
2. *It's complicated (but outrage is easy)*
3. *No surprise, similar impossibility results in decision theory & statistics (Arrow's theorem, Simpsons's paradox, ...)*
4. *Group-based fairness intuitions necessarily contradict each other in populations with unequal base rates*

More & Discussion

1. *(At least) 21 group fairness definitions in the literature*
2. *Nice (and less cynical) e-book by Hardt et al.*
3. *Various conferences (FAT)*
4. *US discourse has strong focus on demographic parity; do European fairness notions align with this?*

—thanks for listening!