

# Bayesian reasoning in the social domain

Jack Cao

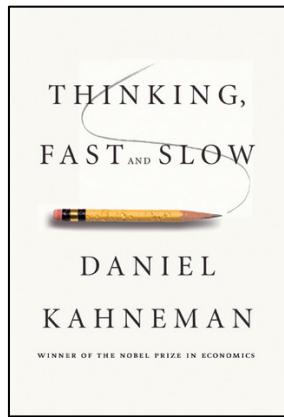
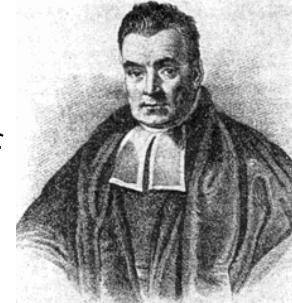
Harvard University

March 8, 2019

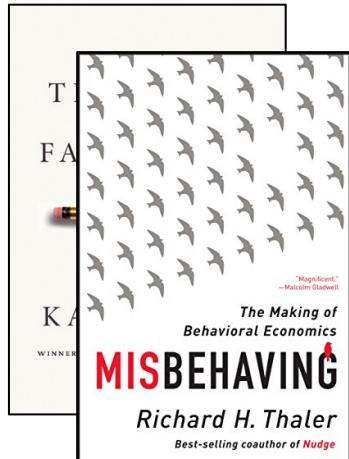
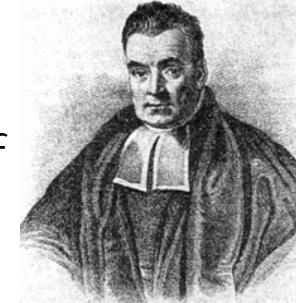
Prior Belief x Quality of New Info = Updated Belief



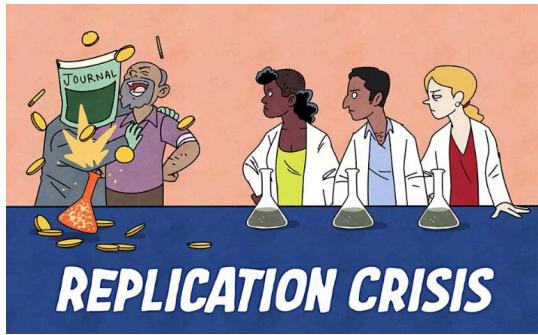
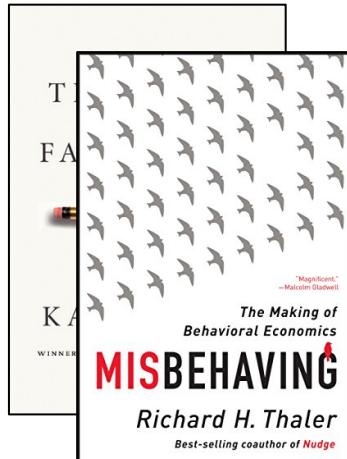
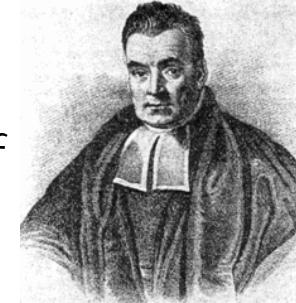
Prior Belief x Quality of New Info = Updated Belief



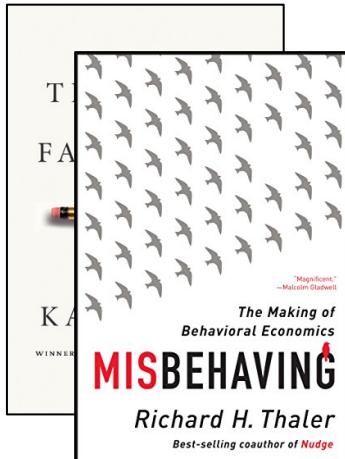
Prior Belief x Quality of New Info = Updated Belief



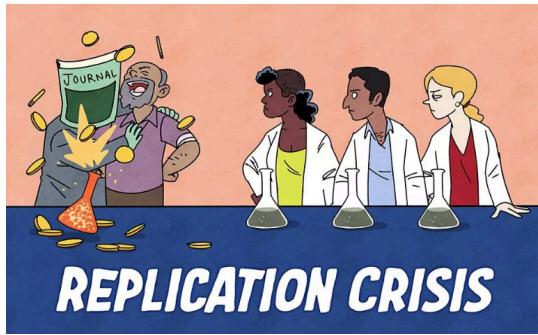
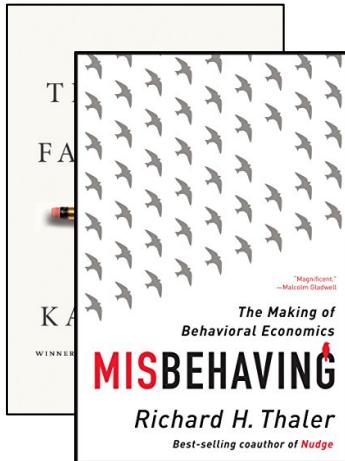
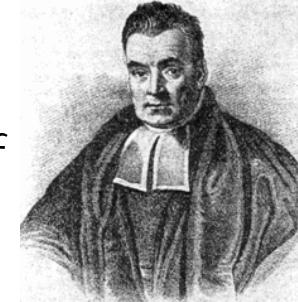
Prior Belief x Quality of New Info = Updated Belief



Prior Belief x Quality of New Info = Updated Belief



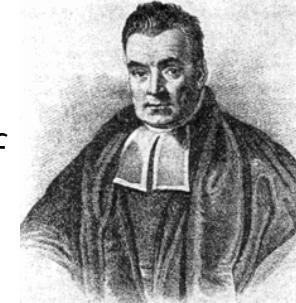
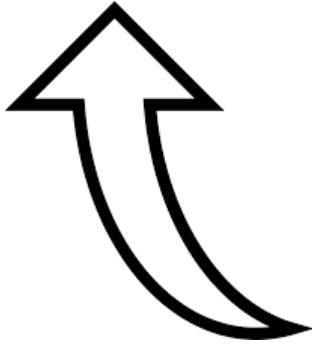
Prior Belief x Quality of New Info = Updated Belief



"The theory that would not die..."  
Sharon McGayne, Science Writer

"...arguably the most powerful mechanism created for processing data and knowledge."  
Jim Berger, Statistician

Prior Belief x Quality of New Info = Updated Belief



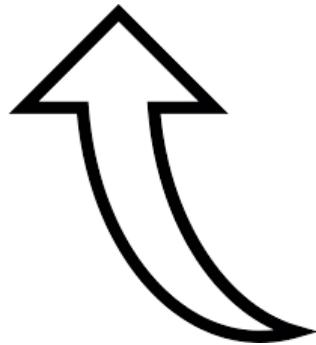
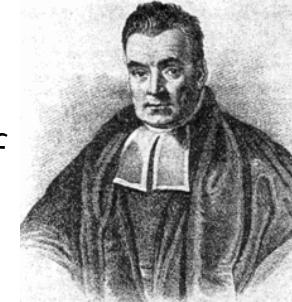
In the social domain, priors are stereotypes.

Locksley et al. (1980)

Krosnick et al. (1990)

Jussim (2012)

Prior Belief x Quality of New Info = Updated Belief



In the social domain, priors are stereotypes.

Locksley et al. (1980)

Krosnick et al. (1990)

Jussim (2012)



The law privileges Egalitarian values  
over Bayesian principles.

# Artificial Intelligence that's more Bayesian than egalitarian



Translate

The image shows a screenshot of the Google Translate interface. At the top, there are two language selection bars. The left bar shows "English", "Spanish", "French", "Turkish - detected", and a dropdown arrow. The right bar shows "English", "Spanish", "Arabic", and a dropdown arrow. Between these bars is a double-headed arrow icon. To the right of the bars is a blue "Translate" button. Below the bars, the input text "O bir doktor.  
O bir hemşire." is displayed in a white box. An "x" icon is in the top right corner of this box. At the bottom of the input box are icons for microphone, camera, and a text entry field. To the right of the input box is a light gray sidebar with a star icon, a square icon, a microphone icon, and a share icon. In the bottom right corner of the input box, the text "28/5000" is visible.

# Artificial Intelligence that's more Bayesian than egalitarian



Translate

The image shows a screenshot of the Google Translate interface. At the top, there are two language selection bars. The left bar shows "English", "Spanish", "French", and "Turkish - detected" with a dropdown arrow. The right bar shows "English", "Spanish", "Arabic", and a dropdown arrow. Between these bars is a double-headed arrow icon. To the right of the bars is a blue "Translate" button.

In the main area, there are two text boxes. The left text box contains the Turkish sentence "O bir doktor." above "O bir hemşire." Below this sentence are two small icons: a speaker icon and a pen icon. To the right of the text box is a small "x" icon. The right text box contains the English translation "He is a doctor." above "She is a nurse." To the right of this sentence are three small icons: a star, a square, a speaker, and a share symbol. A checked checkbox icon is positioned to the right of the "She is a nurse." translation.

Below the text boxes, the character count "28/5000" is displayed.

What judgments do people make when  
Bayesian principles and egalitarian values are at stake?

What judgments do people make when  
Bayesian principles and egalitarian values are at stake?

1. People undermine their commitment to egalitarian values by making Bayesian judgments

What judgments do people make when  
Bayesian principles and egalitarian values are at stake?

1. People undermine their commitment to egalitarian values by making Bayesian judgments
2. Formalizing "egalitarian values"

What judgments do people make when  
Bayesian principles and egalitarian values are at stake?

1. People undermine their commitment to egalitarian values by making Bayesian judgments
2. Formalizing "egalitarian values"
3. Google Images as a proxy for the social environment

# What judgments do people make when Bayesian principles and egalitarian values are at stake?

A man performed surgery.

A woman performed surgery.

Who's more likely to be a doctor?

# What judgments do people make when Bayesian principles and egalitarian values are at stake?

A man performed surgery.

A woman performed surgery.

Who's more likely to be a doctor?

Man

Prior  
belief



Quality  
of new  
info

=

Updated  
belief

Woman

Prior  
Belief



Quality  
of new  
info

=

Updated  
belief

# What judgments do people make when Bayesian principles and egalitarian values are at stake?

A man performed surgery.

A woman performed surgery.

Who's more likely to be a doctor?

Man

Higher  $\times$

Quality  
of new  
info

= Updated  
belief

Woman

Lower  $\times$

Quality  
of new  
info

= Updated  
belief

# What judgments do people make when Bayesian principles and egalitarian values are at stake?

A man performed surgery.

A woman performed surgery.

Who's more likely to be a doctor?

Man

Higher  $\times$

Great,  
but not  
perfect

= Updated  
belief

Woman

Lower  $\times$

Great,  
but not  
perfect

= Updated  
belief

make a  
contribution

subscribe find a job sign in search ▾

US edition ▾

the guardian

news | opinion | sport | arts | lifestyle | more ▾

**Healthcare Network**  
Nursing in focus

Meet the nurse who will soon perform  
surgery on patients alone

Unlike other nursing roles, a surgical care practitioner is involved with the patient  
every step of the way

# What judgments do people make when Bayesian principles and egalitarian values are at stake?

A man performed surgery.

A woman performed surgery.

Who's more likely to be a doctor?

Man

Higher  $\times$

Great,  
but not  
perfect

=

More  
likely

Woman

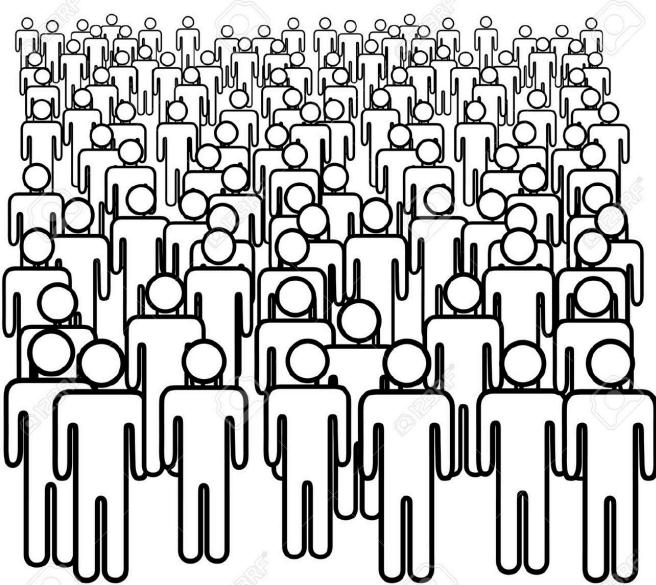
Lower  $\times$

Great,  
but not  
perfect

=

Less  
likely

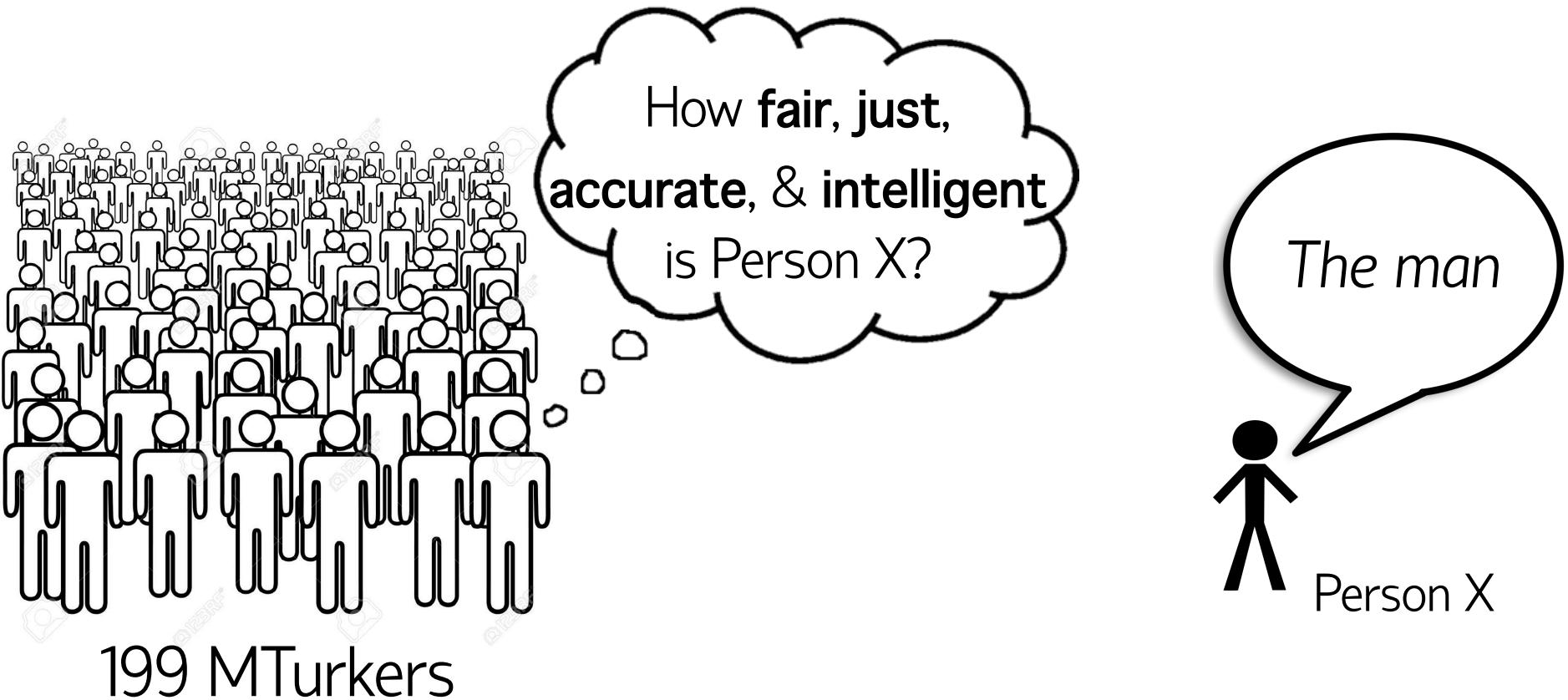
# What judgments do people make when Bayesian principles and egalitarian values are at stake?



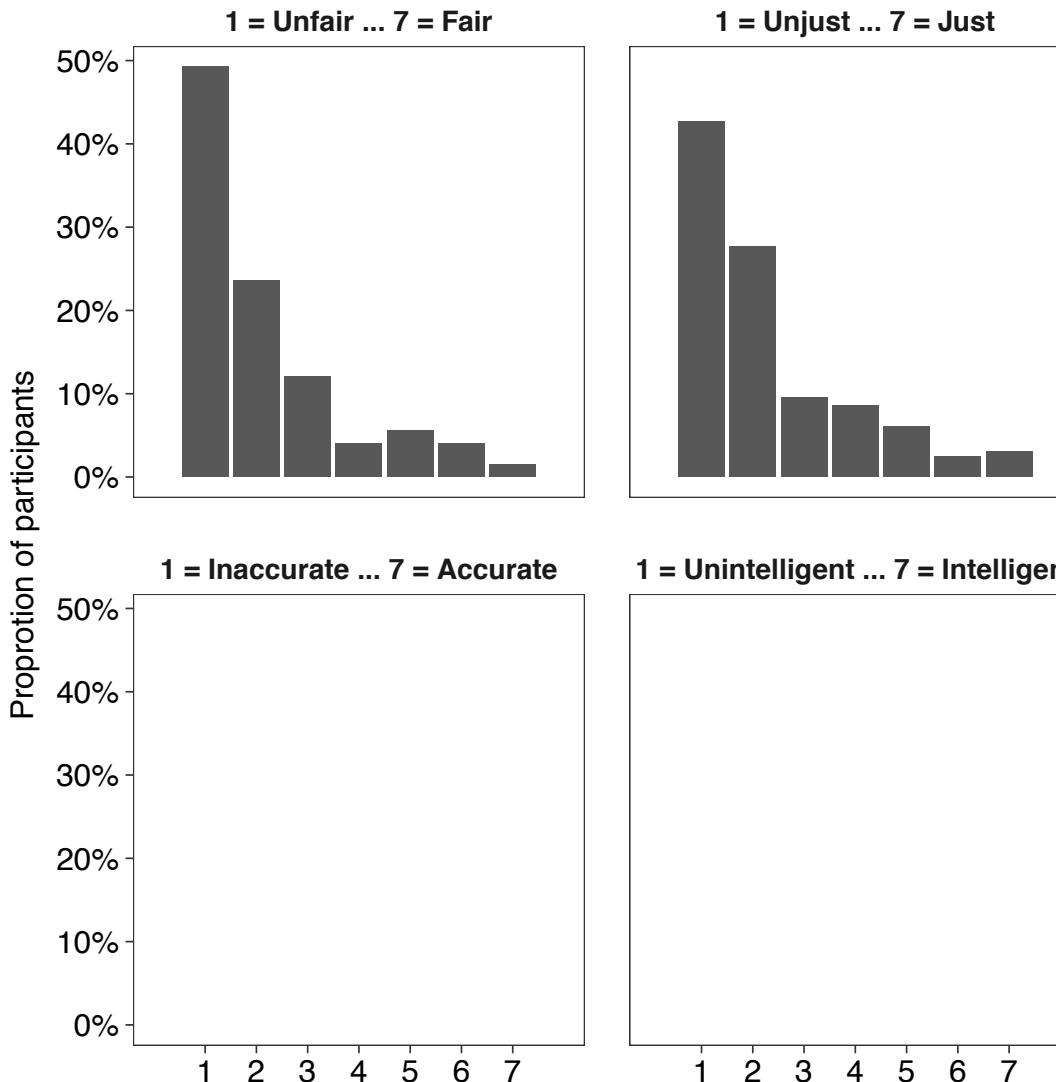
199 MTurkers



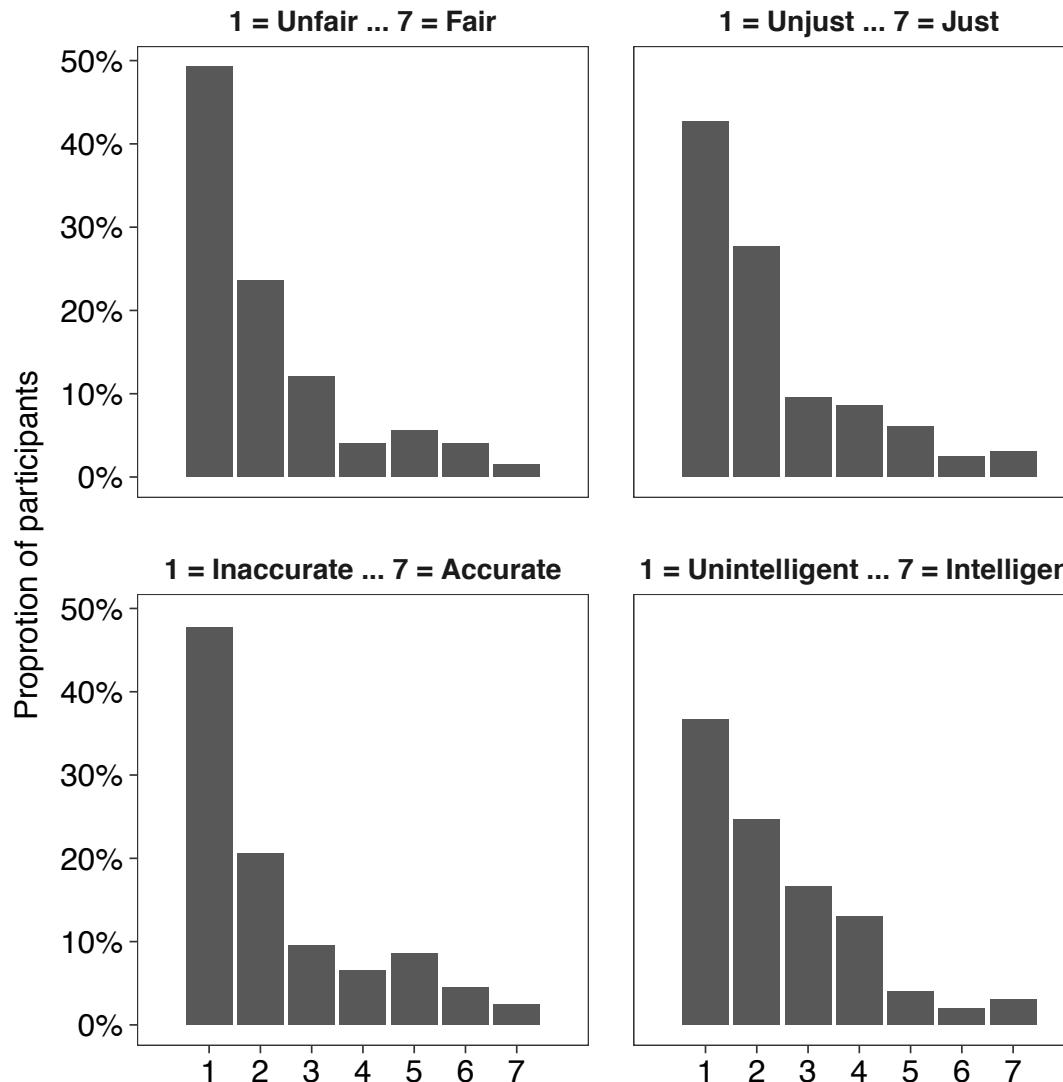
# What judgments do people make when Bayesian principles and egalitarian values are at stake?

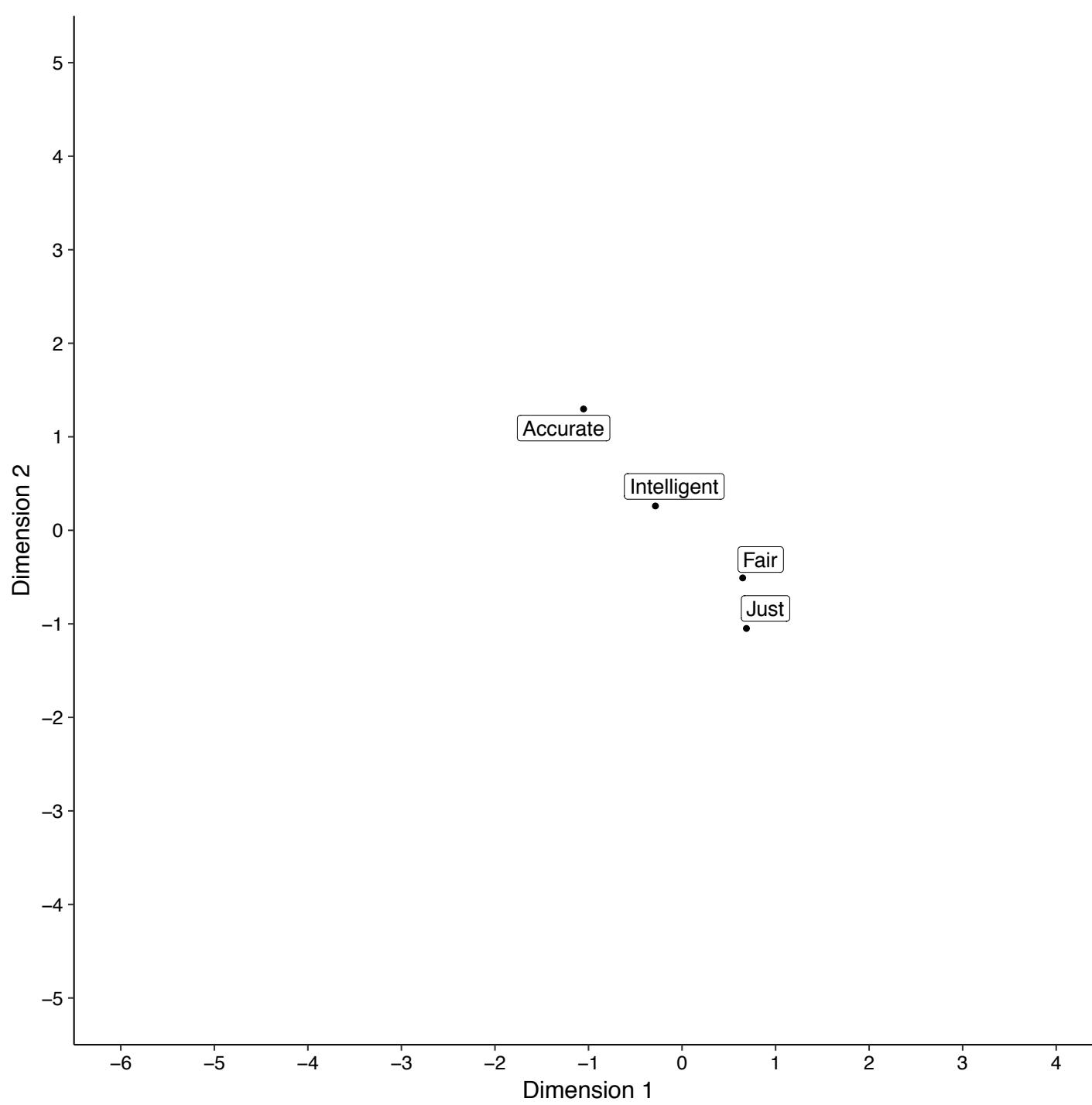


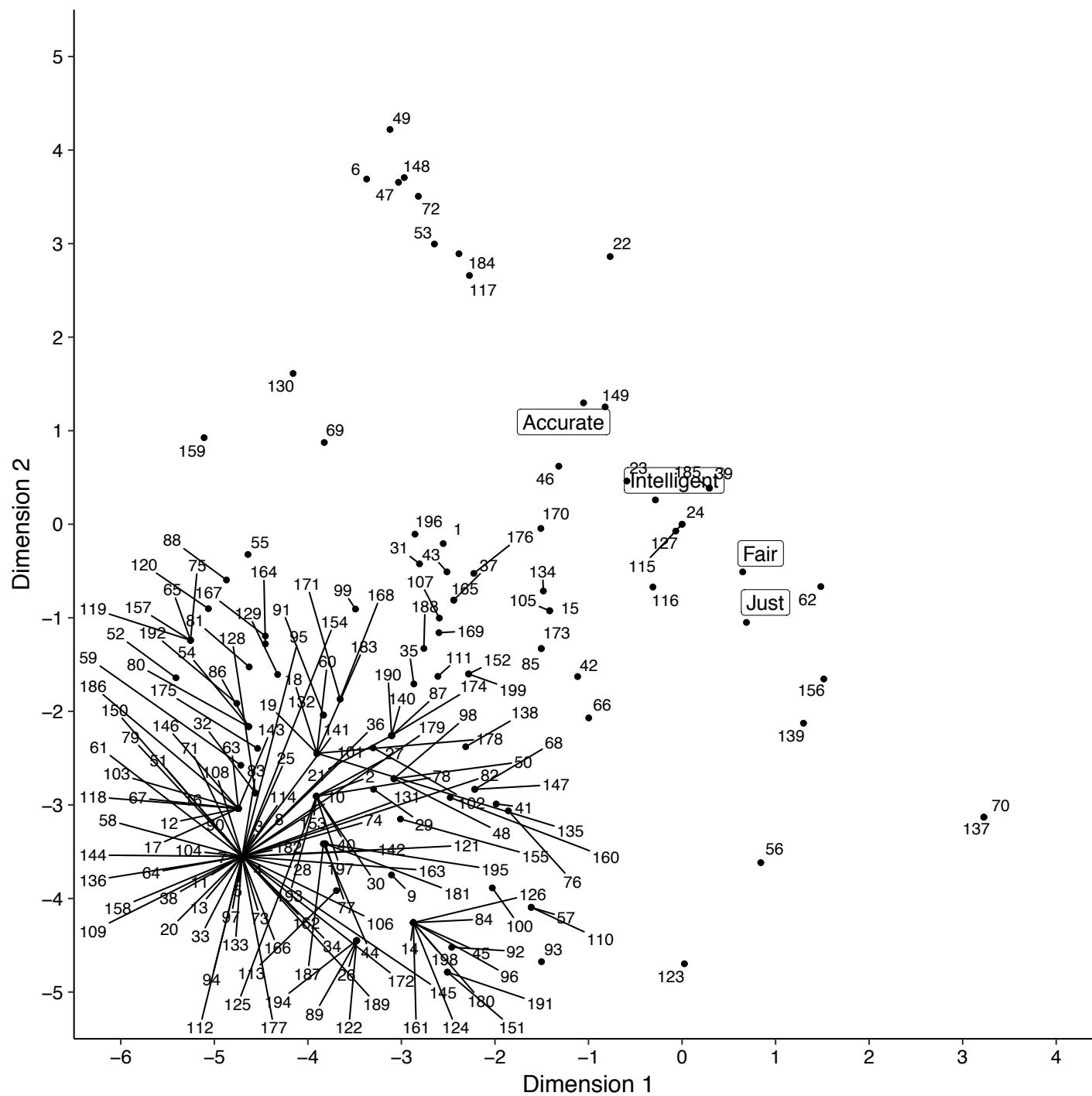
# What judgments do people make when Bayesian principles and egalitarian values are at stake?

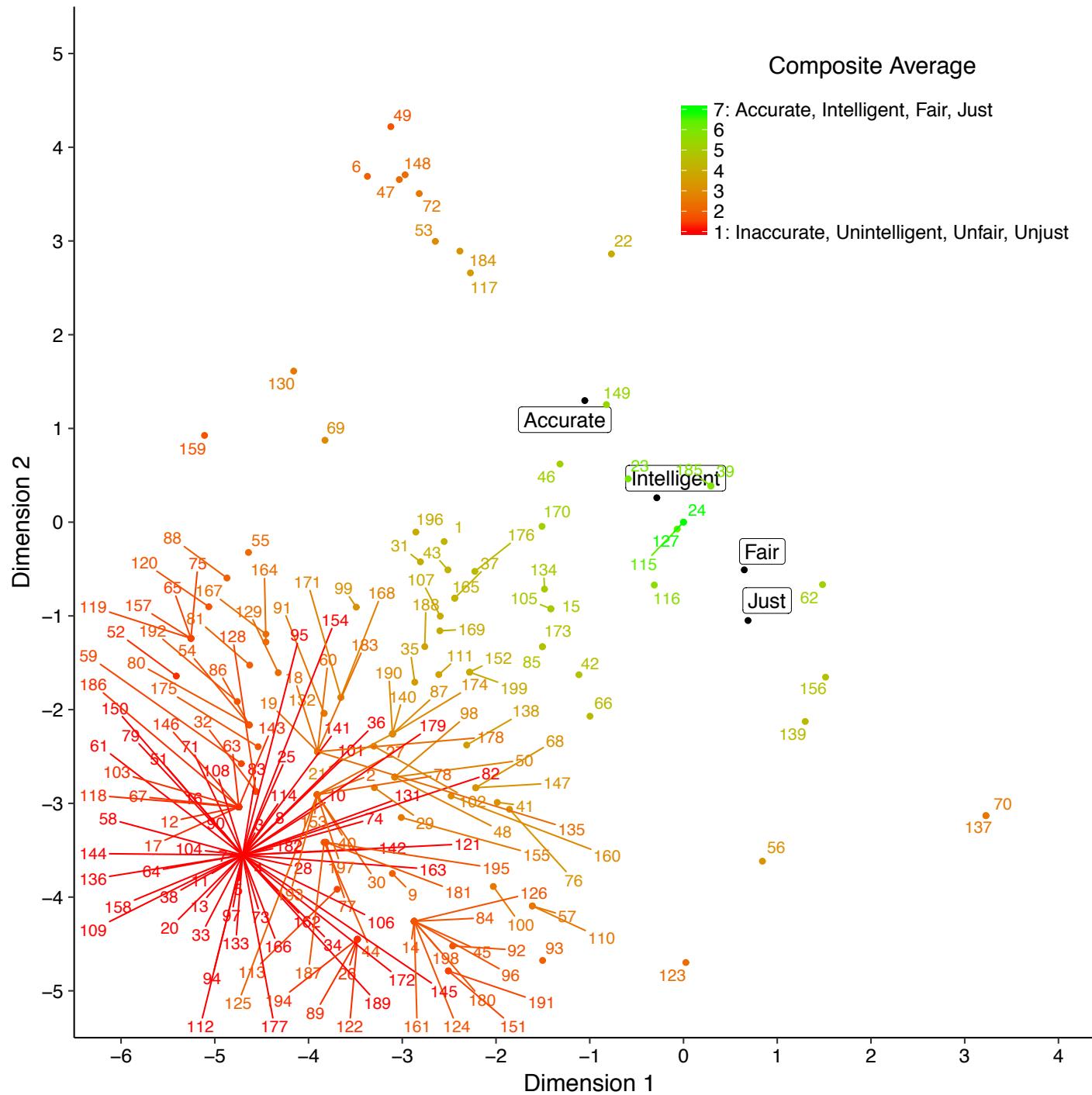


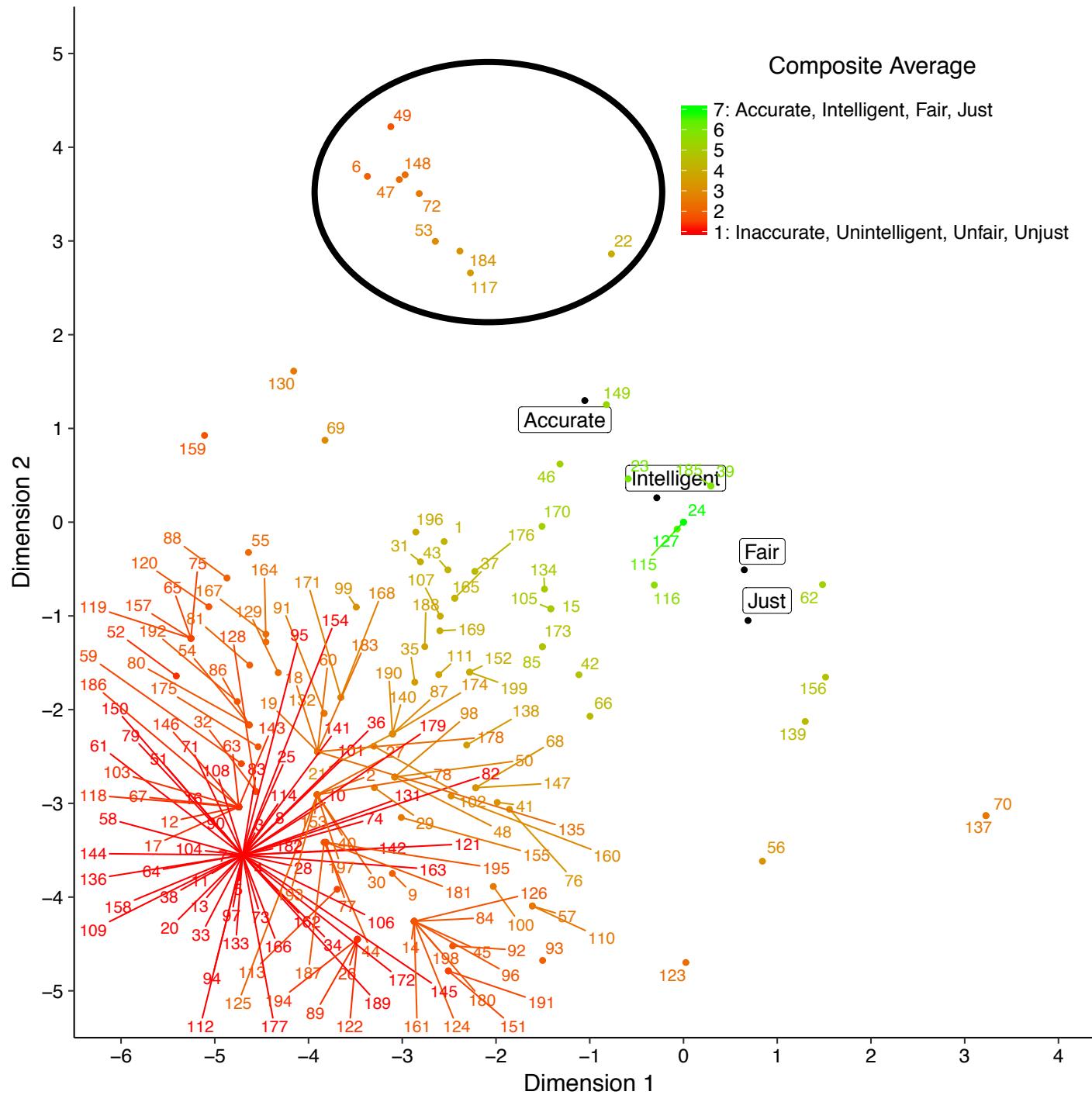
# What judgments do people make when Bayesian principles and egalitarian values are at stake?

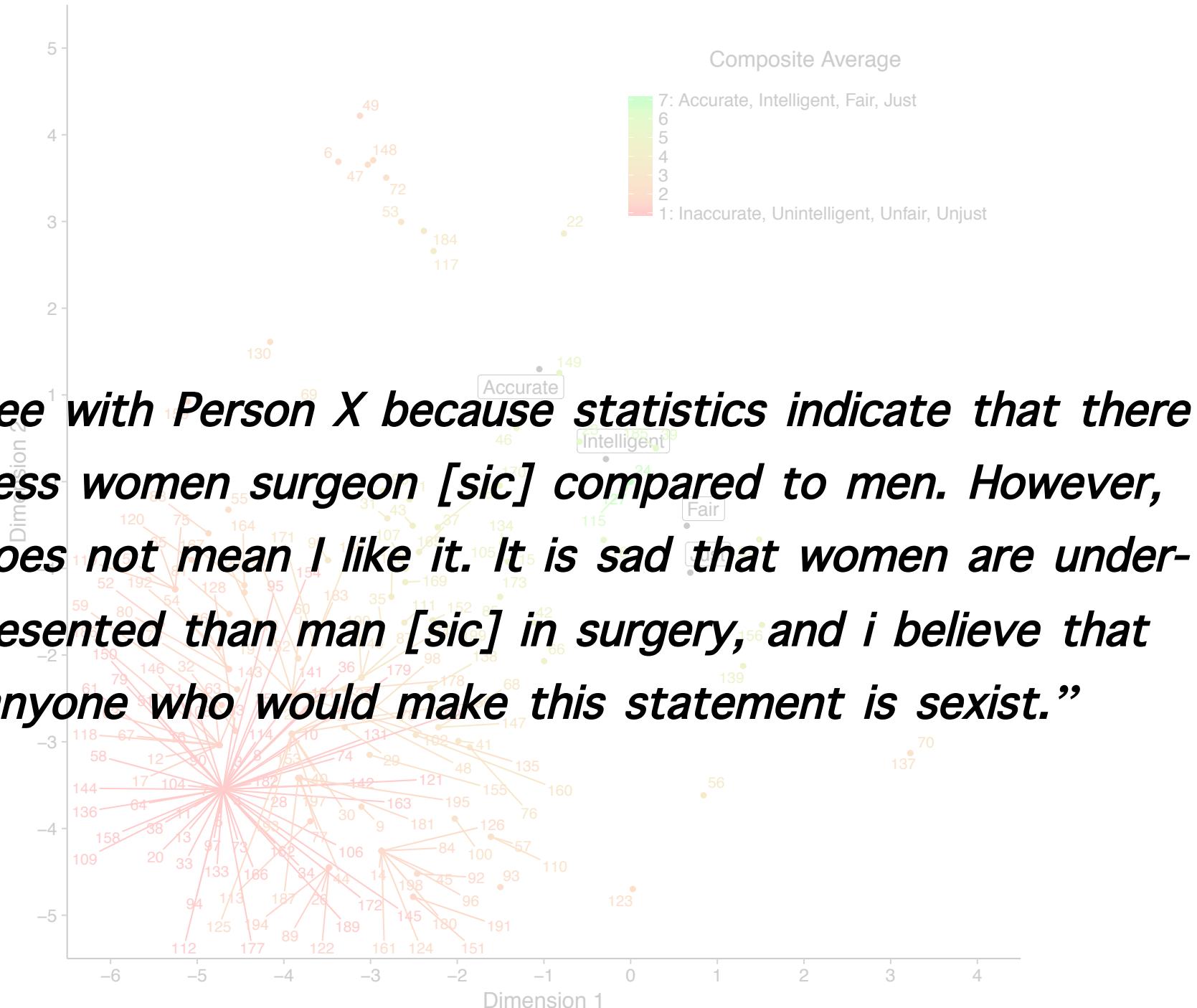




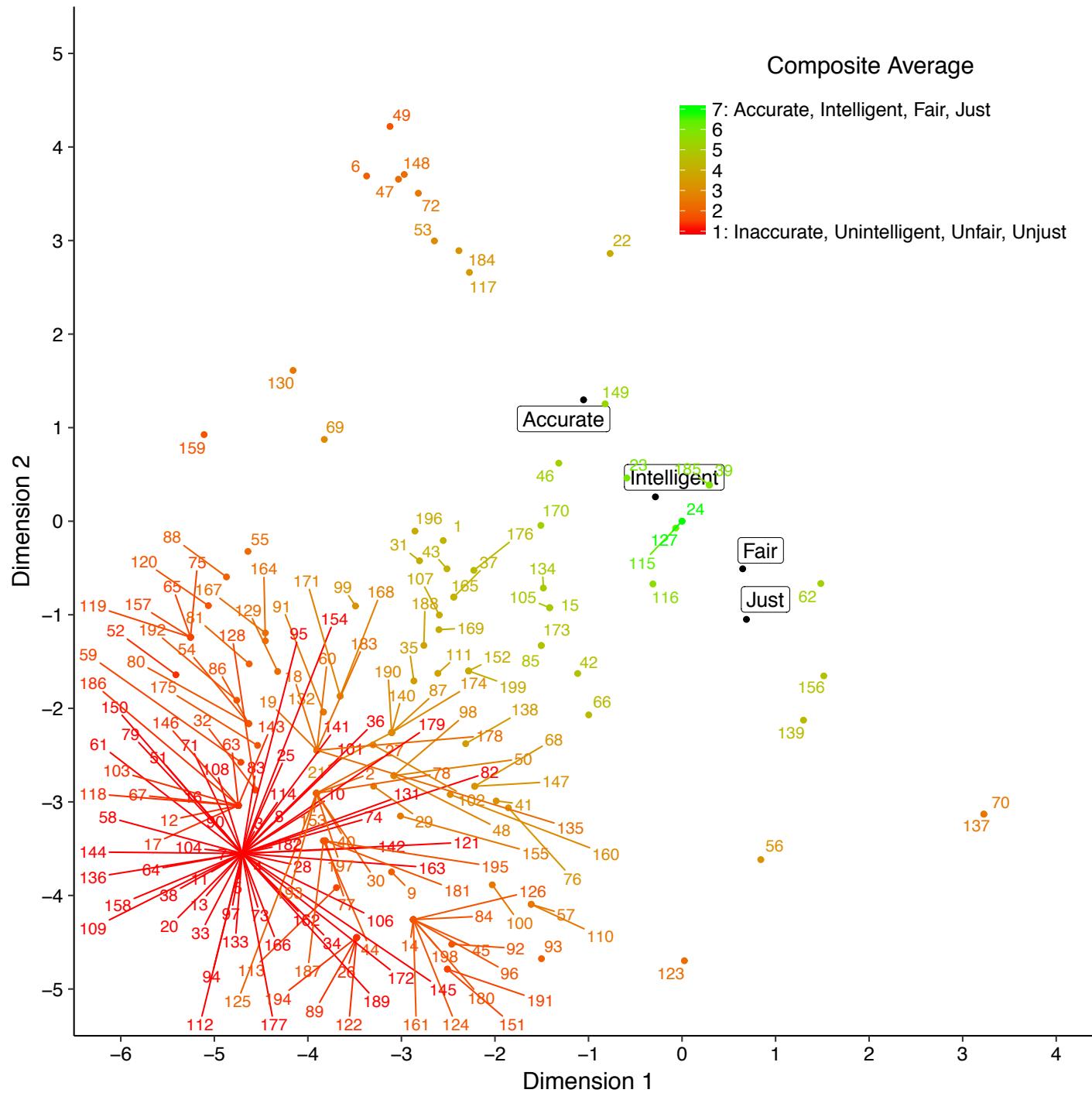






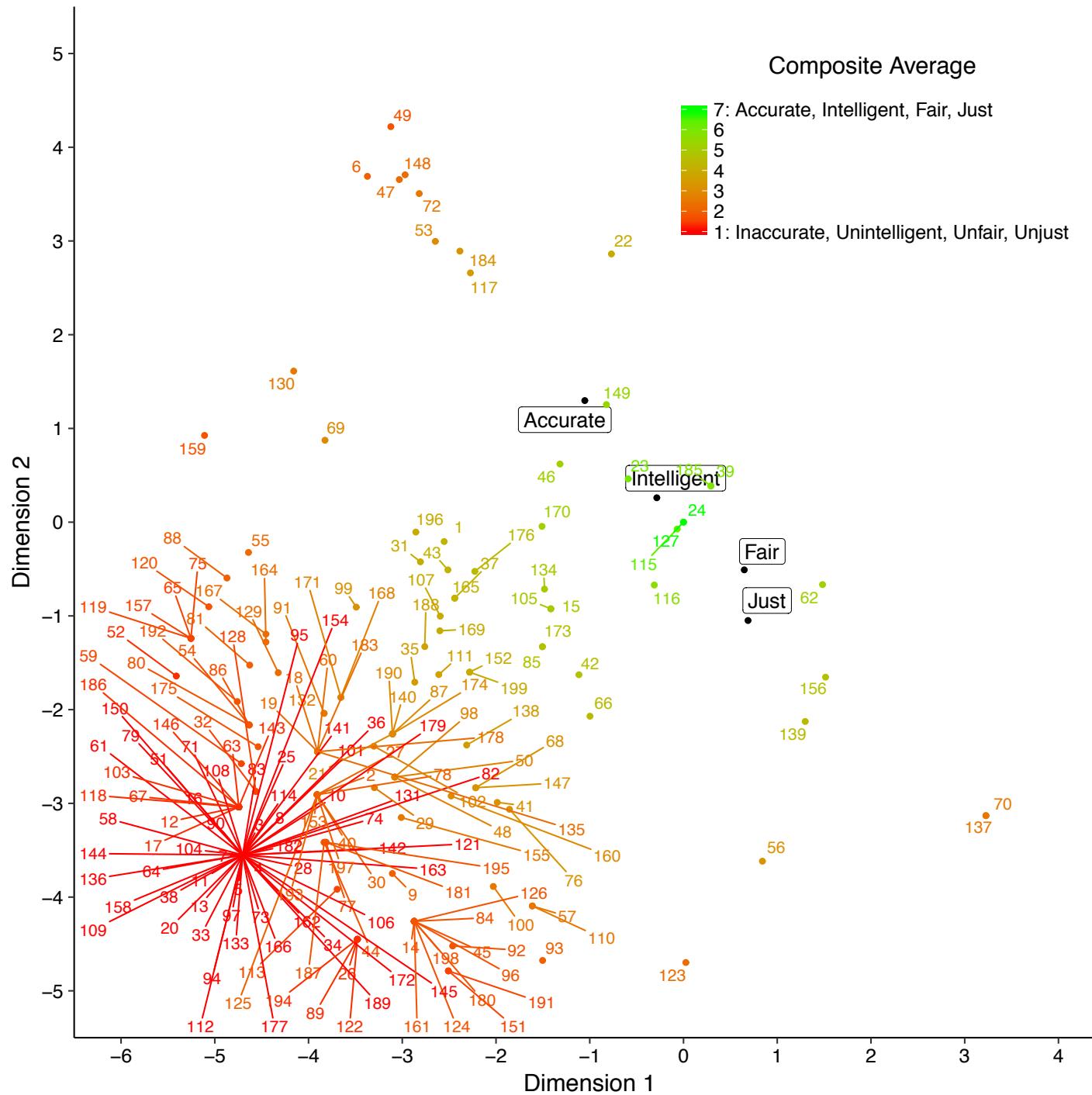


*“I agree with Person X because statistics indicate that there are less women surgeon [sic] compared to men. However, this does not mean I like it. It is sad that women are under-represented than man [sic] in surgery, and i believe that anyone who would make this statement is sexist.”*

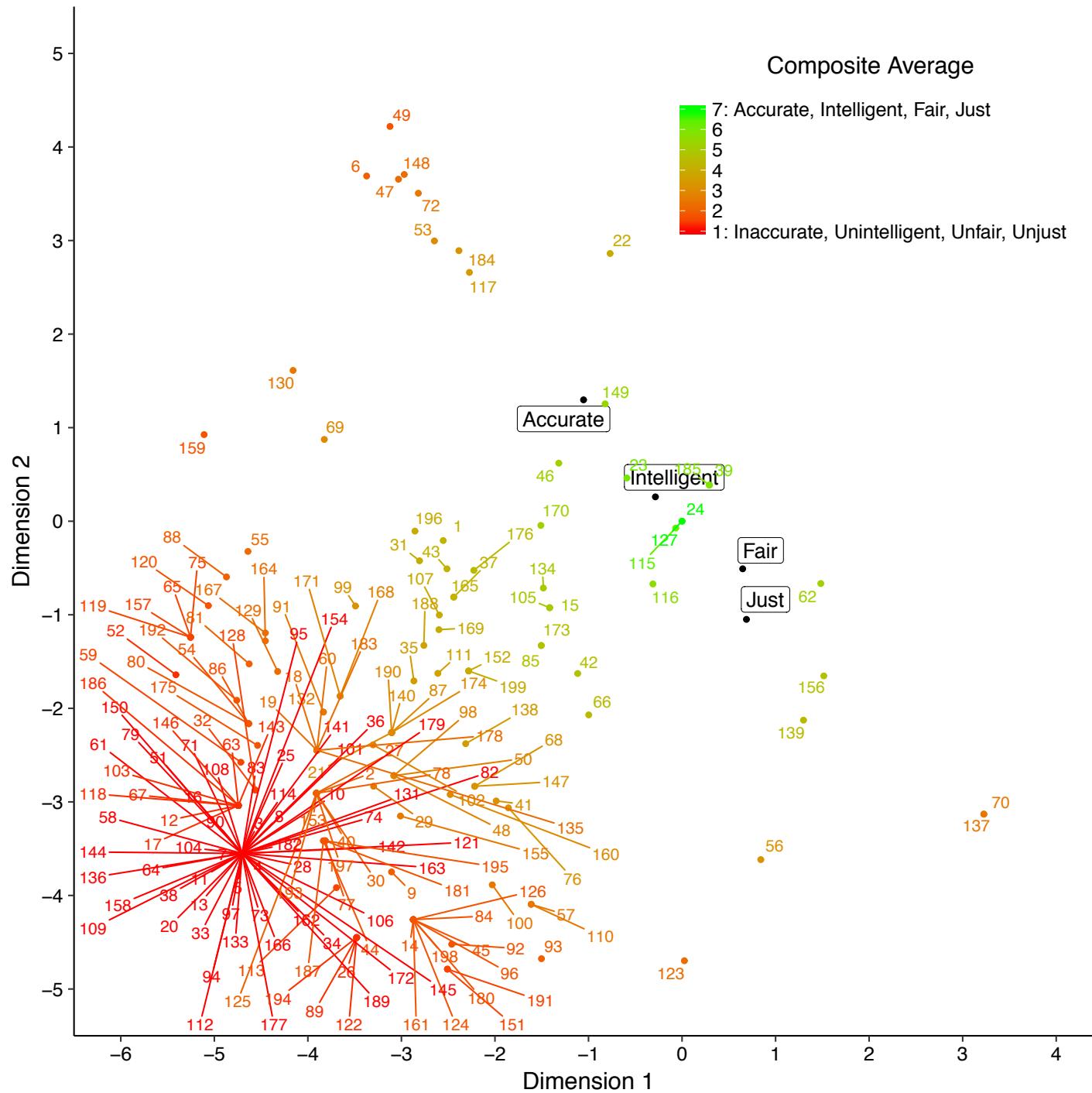










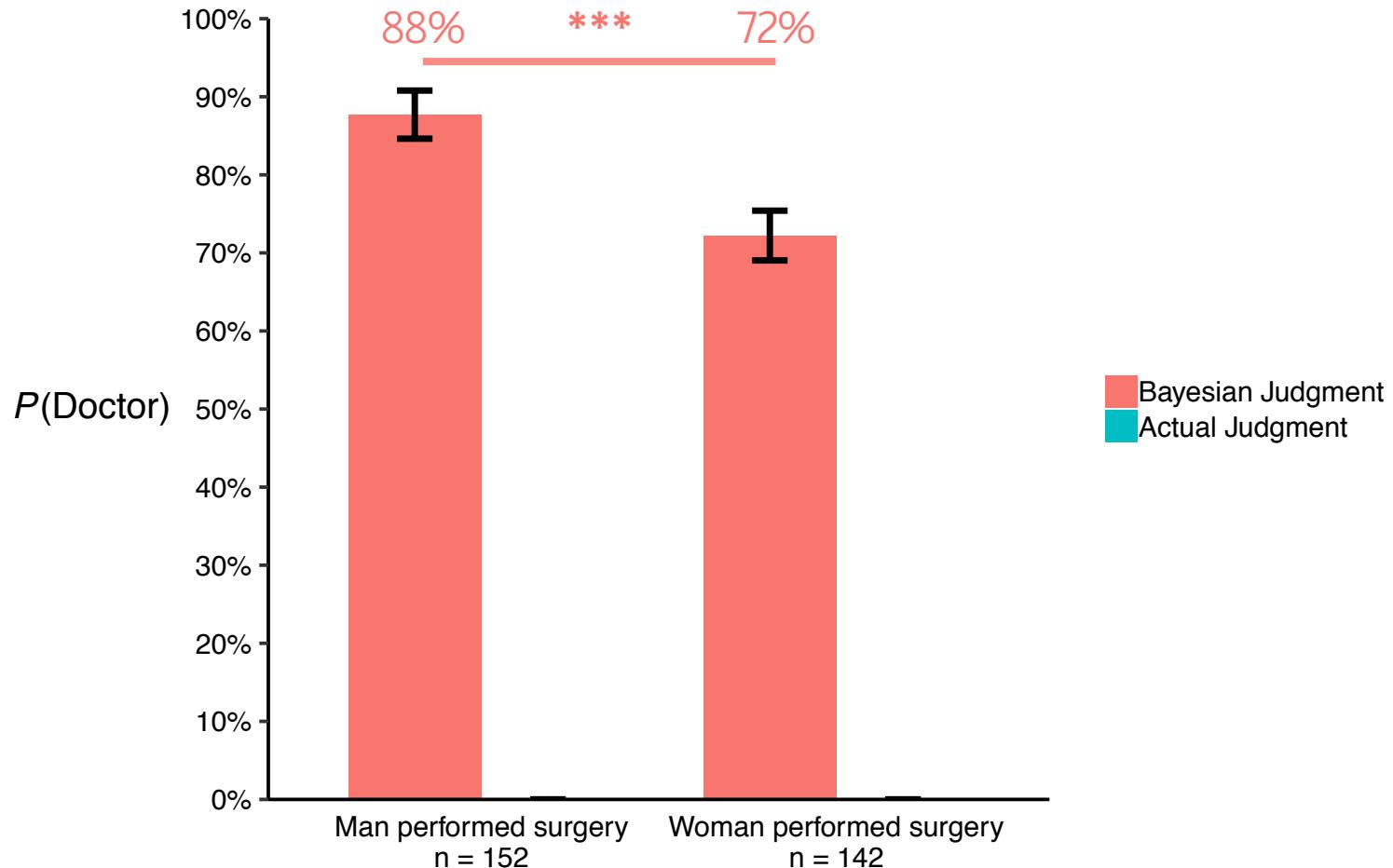




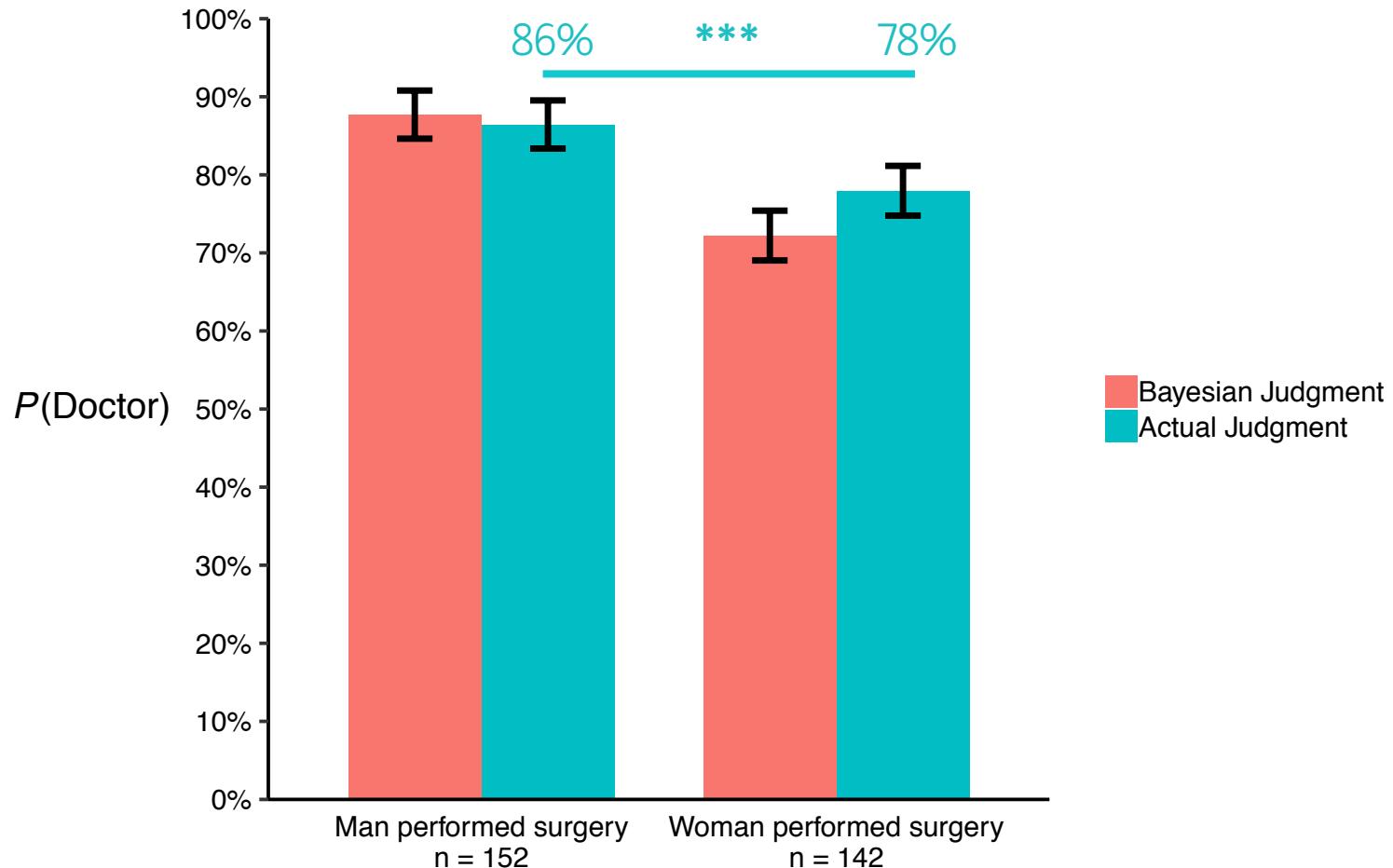
What judgments do people make when  
Bayesian principles and egalitarian values are at stake?

How likely is a man vs. a woman to be a doctor given that each  
performed surgery?

# What judgments do people make when Bayesian principles and egalitarian values are at stake?



# What judgments do people make when Bayesian principles and egalitarian values are at stake?



What judgments do people make when  
Bayesian principles and egalitarian values are at stake?

1. People undermine their commitment to egalitarian values by making Bayesian judgments
2. Formalizing "egalitarian values"
3. Google Images as a proxy for the social environment

# What judgments do people make when Bayesian principles and egalitarian values are at stake?

1. People undermine their commitment to egalitarian values by making Bayesian judgments
2. Formalizing "egalitarian values"
3. Google Images as a proxy for the social environment

How do you go from something like this...



to something like this...

010111 11100111 00001110 111111  
010101 11101100 11010000 110000  
000111 00001000 11011011 001111  
001111 00010000 11011100 000111  
000111 00000110 11001010 001111  
00010 01000000 11101010 001111  
00011 11111111 11001010 001111

# What is unfair in the social domain?

# What is unfair in the social domain?

When outcomes depend on social group, *ceteris paribus*

# What is unfair in the social domain?

When outcomes depend on social group, *ceteris paribus*

Policing Correll et al. (2007), *JPSP*

Medicine Hoffman, Trawalter, et al. (2016), *PNAS*

Hiring Moss-Racusin et al. (2012), *PNAS*

# What is unfair in the social domain?

When outcomes depend on social group, *ceteris paribus*

Policing Correll et al. (2007), *JPSP*

Medicine Hoffman, Trawalter, et al. (2016), *PNAS*

Hiring Moss-Racusin et al. (2012), *PNAS*

$$P(\text{Shoot} \mid \text{Unarmed, White}) < P(\text{Shoot} \mid \text{Unarmed, Black})$$

# What is unfair in the social domain?

When outcomes depend on social group, *ceteris paribus*

Policing Correll et al. (2007), JPSP

Medicine Hoffman, Trawalter, et al. (2016), PNAS

Hiring Moss-Racusin et al. (2012), PNAS

$$P(\text{Undertreat} \mid \text{In Pain, White}) < P(\text{Undertreat} \mid \text{In Pain, Black})$$

# What is unfair in the social domain?

When outcomes depend on social group, *ceteris paribus*

Policing Correll et al. (2007), *JPSP*

Medicine Hoffman, Trawalter, et al. (2016), *PNAS*

Hiring Moss-Racusin et al. (2012), *PNAS*

$$P(\text{Hire} \mid \text{Qualified, Female}) < P(\text{Hire} \mid \text{Qualified, Male})$$

$$P(\text{Shoot} \mid \text{Unarmed, White}) < P(\text{Shoot} \mid \text{Unarmed, Black})$$
$$P(\text{Untreat} \mid \text{In Pain, White}) < P(\text{Untreat} \mid \text{In Pain, Black})$$
$$P(\text{Hire} \mid \text{Qualified, Female}) < P(\text{Hire} \mid \text{Qualified, Male})$$

## Unequal false positive rates, FPRs

$P(\text{Shoot} \mid \text{Unarmed, White}) < P(\text{Shoot} \mid \text{Unarmed, Black})$

$P(\text{Untreat} \mid \text{In Pain, White}) < P(\text{Untreat} \mid \text{In Pain, Black})$

$P(\text{Hire} \mid \text{Qualified, Female}) < P(\text{Hire} \mid \text{Qualified, Male})$

Unequal false positive rates, FPRs

$P(\text{Shoot} \mid \text{Unarmed, White}) < P(\text{Shoot} \mid \text{Unarmed, Black})$

Unequal false negative rates, FNRs

$P(\text{Undertreat} \mid \text{In Pain, White}) < P(\text{Undertreat} \mid \text{In Pain, Black})$

$P(\text{Hire} \mid \text{Qualified, Female}) < P(\text{Hire} \mid \text{Qualified, Male})$

### Unequal false positive rates, FPRs

$P(\text{Shoot} \mid \text{Unarmed, White}) < P(\text{Shoot} \mid \text{Unarmed, Black})$

### Unequal false negative rates, FNRs

$P(\text{Undertreat} \mid \text{In Pain, White}) < P(\text{Undertreat} \mid \text{In Pain, Black})$

### Unequal positive predictive values, PPVs

$P(\text{Hire} \mid \text{Qualified, Female}) < P(\text{Hire} \mid \text{Qualified, Male})$

# Not egalitarian

Unequal false positive rates, FPRs

$$P(\text{Shoot} \mid \text{Unarmed, White}) < P(\text{Shoot} \mid \text{Unarmed, Black})$$

Unequal false negative rates, FNRs

$$P(\text{Undertreat} \mid \text{In Pain, White}) < P(\text{Undertreat} \mid \text{In Pain, Black})$$

Unequal positive predictive values, PPVs

$$P(\text{Hire} \mid \text{Qualified, Female}) < P(\text{Hire} \mid \text{Qualified, Male})$$

# Egalitarian

Equal false positive rates, FPRs

$$P(\text{Shoot} \mid \text{Unarmed, White}) = P(\text{Shoot} \mid \text{Unarmed, Black})$$

Equal false negative rates, FNRs

$$P(\text{Undertreat} \mid \text{In Pain, White}) = P(\text{Undertreat} \mid \text{In Pain, Black})$$

Equal positive predictive values, PPVs

$$P(\text{Hire} \mid \text{Qualified, Female}) = P(\text{Hire} \mid \text{Qualified, Male})$$

# Why these three ways of thinking about egalitarian values?

$$\text{FPR}_{\text{Group A}} = \text{FPR}_{\text{Group B}}$$

$$\text{FNR}_{\text{Group A}} = \text{FNR}_{\text{Group B}}$$

$$\text{PPV}_{\text{Group A}} = \text{PPV}_{\text{Group B}}$$

People care about them

# Why these three ways of thinking about egalitarian values?

$$\text{FPR}_{\text{Group A}} = \text{FPR}_{\text{Group B}}$$

$$\text{FNR}_{\text{Group A}} = \text{FNR}_{\text{Group B}}$$

$$\text{PPV}_{\text{Group A}} = \text{PPV}_{\text{Group B}}$$

People care about them; Mathematically precise

# Why these three ways of thinking about egalitarian values?

$$\text{FPR}_{\text{Group A}} = \text{FPR}_{\text{Group B}}$$

$$\text{FNR}_{\text{Group A}} = \text{FNR}_{\text{Group B}}$$

$$\text{PPV}_{\text{Group A}} = \text{PPV}_{\text{Group B}}$$

People care about them; Mathematically precise; Sinister tradeoff

# Why these three ways of thinking about egalitarian values?

$$\text{FPR}_{\text{Group A}} = \text{FPR}_{\text{Group B}}$$

$$\text{FNR}_{\text{Group A}} = \text{FNR}_{\text{Group B}}$$

$$\text{PPV}_{\text{Group A}} = \text{PPV}_{\text{Group B}}$$

When base rates between Group A and Group B differ, all three definitions cannot be simultaneously met. At least one must be given up.

Kleinberg, Mullainathan, & Raghavan (2016)

# The sinister tradeoff

$$P(\text{cancer} \mid \text{pos}) = \frac{P(\text{pos} \mid \text{cancer}) \times P(\text{cancer})}{P(\text{pos} \mid \text{cancer}) \times P(\text{cancer}) + P(\text{pos} \mid \text{no cancer}) \times P(\text{no cancer})}$$

# The sinister tradeoff

$$P(\text{cancer} \mid \text{pos}) = \frac{P(\text{pos} \mid \text{cancer}) \times P(\text{cancer})}{P(\text{pos} \mid \text{cancer}) \times P(\text{cancer}) + P(\text{pos} \mid \text{no cancer}) \times P(\text{no cancer})}$$

$$P(\text{cancer} \mid \text{pos}) = \frac{[1 - P(\text{neg} \mid \text{cancer})] \times P(\text{cancer})}{[1 - P(\text{neg} \mid \text{cancer})] \times P(\text{cancer}) + P(\text{pos} \mid \text{no cancer}) \times [1 - P(\text{cancer})]}$$

# The sinister tradeoff

$$P(\text{cancer} \mid \text{pos}) = \frac{P(\text{pos} \mid \text{cancer}) \times P(\text{cancer})}{P(\text{pos} \mid \text{cancer}) \times P(\text{cancer}) + P(\text{pos} \mid \text{no cancer}) \times P(\text{no cancer})}$$

$$P(\text{cancer} \mid \text{pos}) = \frac{[1 - P(\text{neg} \mid \text{cancer})] \times P(\text{cancer})}{[1 - P(\text{neg} \mid \text{cancer})] \times P(\text{cancer}) + P(\text{pos} \mid \text{no cancer}) \times [1 - P(\text{cancer})]}$$

$$\text{PPV} = \frac{[1 - \text{FNR}] \times \text{BR}}{[1 - \text{FNR}] \times \text{BR} + \text{FPR} \times [1 - \text{BR}]}$$

# The sinister tradeoff

$$P(\text{cancer} \mid \text{pos}) = \frac{P(\text{pos} \mid \text{cancer}) \times P(\text{cancer})}{P(\text{pos} \mid \text{cancer}) \times P(\text{cancer}) + P(\text{pos} \mid \text{no cancer}) \times P(\text{no cancer})}$$

$$P(\text{cancer} \mid \text{pos}) = \frac{[1 - P(\text{neg} \mid \text{cancer})] \times P(\text{cancer})}{[1 - P(\text{neg} \mid \text{cancer})] \times P(\text{cancer}) + P(\text{pos} \mid \text{no cancer}) \times [1 - P(\text{cancer})]}$$

$$\text{PPV} = \frac{[1 - \text{FNR}] \times \text{BR}}{[1 - \text{FNR}] \times \text{BR} + \text{FPR} \times [1 - \text{BR}]}$$

$$\text{FPR} = \frac{\text{BR}}{[1 - \text{BR}]} \times \frac{[1 - \text{PPV}]}{\text{PPV}} \times [1 - \text{FNR}]$$

# The sinister tradeoff

$$FPR_w = \frac{BR_w}{1-BR_w} \times \frac{1-PPV_w}{PPV_w} \times [1-FNR_w]$$

$$FPR_m = \frac{BR_m}{1-BR_m} \times \frac{1-PPV_m}{PPV_m} \times [1-FNR_m]$$

# The sinister tradeoff

$$FPR_w = \frac{BR_w}{1-BR_w} \times \frac{1-PPV_w}{PPV_w} \times [1-FNR_w]$$

$$FPR_m = \frac{BR_m}{1-BR_m} \times \frac{1-PPV_m}{PPV_m} \times [1-FNR_m]$$

$$\frac{FPR_w}{FPR_m} = \frac{\frac{BR_w}{1-BR_w} \times \frac{1-PPV_w}{PPV_w} \times [1-FNR_w]}{\frac{BR_m}{1-BR_m} \times \frac{1-PPV_m}{PPV_m} \times [1-FNR_m]}$$

# The sinister tradeoff

$$FPR_w = \frac{BR_w}{1-BR_w} \times \frac{1-PPV_w}{PPV_w} \times [1-FNR_w]$$

$$FPR_m = \frac{BR_m}{1-BR_m} \times \frac{1-PPV_m}{PPV_m} \times [1-FNR_m]$$

Meet two definitions of fairness

$$PPV_w = PPV_m$$

$$FNR_w = FNR_m$$

$$\frac{FPR_w}{FPR_m} = \frac{\frac{BR_w}{1-BR_w} \times \frac{1-PPV_w}{PPV_w} \times [1-FNR_w]}{\frac{BR_m}{1-BR_m} \times \frac{1-PPV_m}{PPV_m} \times [1-FNR_m]}$$

# The sinister tradeoff

$$FPR_w = \frac{BR_w}{1-BR_w} \times \frac{1-PPV_w}{PPV_w} \times [1-FNR_w]$$

$$FPR_m = \frac{BR_m}{1-BR_m} \times \frac{1-PPV_m}{PPV_m} \times [1-FNR_m]$$

Meet two definitions of fairness

$$PPV_w = PPV_m$$

$$FNR_w = FNR_m$$

$$\frac{FPR_w}{FPR_m} = \frac{\cancel{\frac{BR_w}{1-BR_w}} \times \cancel{\frac{1-PPV_w}{PPV_w}} \times \cancel{[1-FNR_w]}}{\cancel{\frac{BR_m}{1-BR_m}} \times \cancel{\frac{1-PPV_m}{PPV_m}} \times \cancel{[1-FNR_m]}}$$

# The sinister tradeoff

$$FPR_w = \frac{BR_w}{1-BR_w} \times \frac{1-PPV_w}{PPV_w} \times [1-FNR_w]$$

$$FPR_m = \frac{BR_m}{1-BR_m} \times \frac{1-PPV_m}{PPV_m} \times [1-FNR_m]$$

Meet two definitions of fairness

$$PPV_w = PPV_m$$

$$FNR_w = FNR_m$$

$$\frac{FPR_w}{FPR_m} = \frac{\cancel{\frac{BR_w}{1-BR_w}} \times \cancel{\frac{1-PPV_w}{PPV_w}} \times \cancel{[1-FNR_w]}}{\cancel{\frac{BR_m}{1-BR_m}} \times \cancel{\frac{1-PPV_m}{PPV_m}} \times \cancel{[1-FNR_m]}}$$

$$\frac{FPR_w}{FPR_m} = \frac{\cancel{\frac{BR_w}{1-BR_w}}}{\frac{\cancel{BR_m}}{\cancel{1-BR_m}}}$$

# The sinister tradeoff

$$FPR_w = \frac{BR_w}{1-BR_w} \times \frac{1-PPV_w}{PPV_w} \times [1-FNR_w]$$

$$FPR_m = \frac{BR_m}{1-BR_m} \times \frac{1-PPV_m}{PPV_m} \times [1-FNR_m]$$

Meet two definitions of fairness

$$PPV_w = PPV_m$$

$$FNR_w = FNR_m$$

$$\frac{FPR_w}{FPR_m} = \frac{\cancel{\frac{BR_w}{1-BR_w}} \times \cancel{\frac{1-PPV_w}{PPV_w}} \times \cancel{[1-FNR_w]}}{\cancel{\frac{BR_m}{1-BR_m}} \times \cancel{\frac{1-PPV_m}{PPV_m}} \times \cancel{[1-FNR_m]}}$$

$$\frac{FPR_w}{FPR_m} = \frac{\frac{BR_w}{1-BR_w}}{\frac{BR_m}{1-BR_m}}$$

$$FPR_w = \frac{BR_w}{BR_m} \times \frac{[1-BR_m]}{[1-BR_w]} \times FPR_m$$

# The sinister tradeoff

$$FPR_w = \frac{BR_w}{1-BR_w} \times \frac{1-PPV_w}{PPV_w} \times [1-FNR_w]$$

$$FPR_m = \frac{BR_m}{1-BR_m} \times \frac{1-PPV_m}{PPV_m} \times [1-FNR_m]$$

Meet two definitions of fairness

$$PPV_w = PPV_m$$

$$FNR_w = FNR_m$$

$$\frac{FPR_w}{FPR_m} = \frac{\cancel{\frac{BR_w}{1-BR_w}} \times \cancel{\frac{1-PPV_w}{PPV_w}} \times \cancel{[1-FNR_w]}}{\cancel{\frac{BR_m}{1-BR_m}} \times \cancel{\frac{1-PPV_m}{PPV_m}} \times \cancel{[1-FNR_m]}}$$

$$\frac{FPR_w}{FPR_m} = \frac{\cancel{\frac{BR_w}{1-BR_w}}}{\cancel{\frac{BR_m}{1-BR_m}}}$$

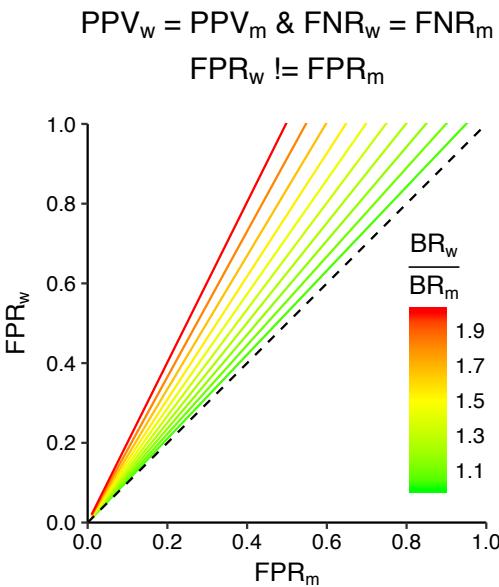
$$FPR_w = \frac{BR_w}{BR_m} \times \frac{[1-BR_m]}{[1-BR_w]} \times FPR_m$$

Cannot meet third definition of fairness

$$FPR_w \neq FPR_m$$

*Breast cancer is more common among women than men*

# The sinister tradeoff



$$FPR_w = \frac{BR_w}{1-BR_w} \times \frac{1-PPV_w}{PPV_w} \times [1-FNR_w]$$

$$FPR_m = \frac{BR_m}{1-BR_m} \times \frac{1-PPV_m}{PPV_m} \times [1-FNR_m]$$

$$\frac{FPR_w}{FPR_m} = \frac{\frac{BR_w}{1-BR_w} \times \frac{1-PPV_w}{PPV_w} \times [1-FNR_w]}{\frac{BR_m}{1-BR_m} \times \frac{1-PPV_m}{PPV_m} \times [1-FNR_m]}$$

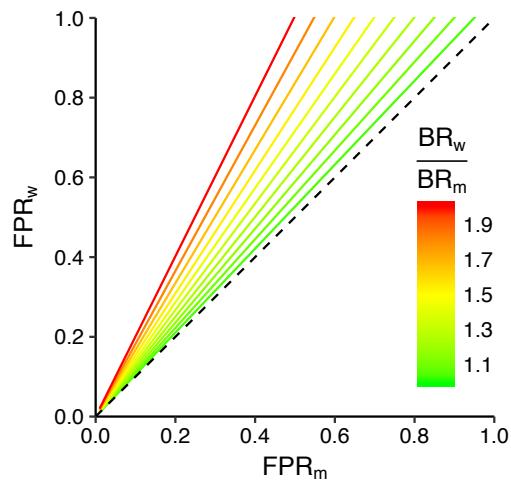
$$\frac{FPR_w}{FPR_m} = \frac{\frac{BR_w}{1-BR_w}}{\frac{BR_m}{1-BR_m}}$$

$$FPR_w = \frac{BR_w}{BR_m} \times \frac{[1-BR_m]}{[1-BR_w]} \times FPR_m$$

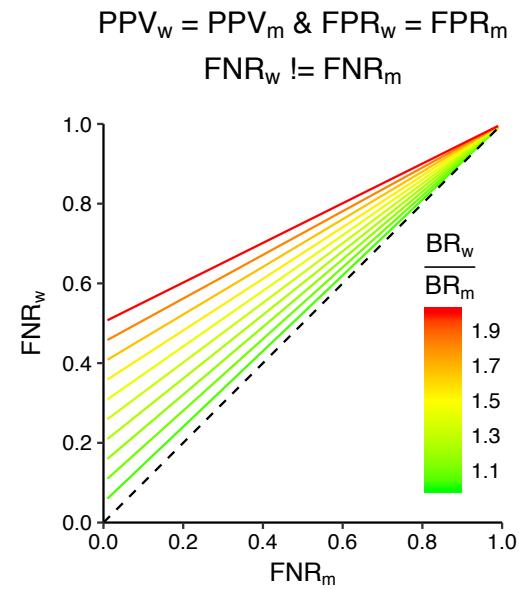
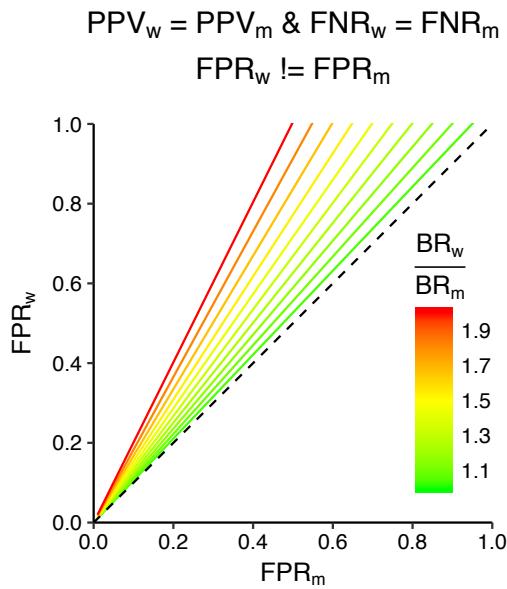
# The sinister tradeoff

$$\text{PPV}_w = \text{PPV}_m \text{ & } \text{FNR}_w = \text{FNR}_m$$

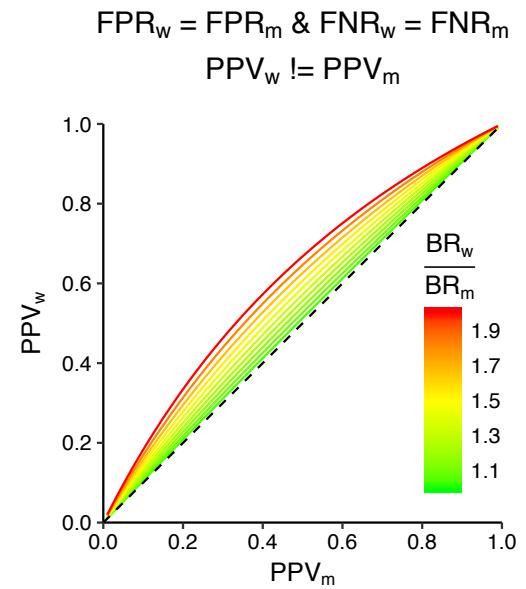
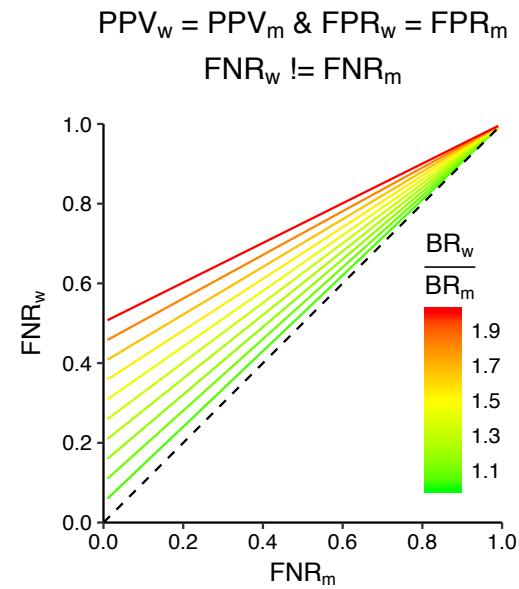
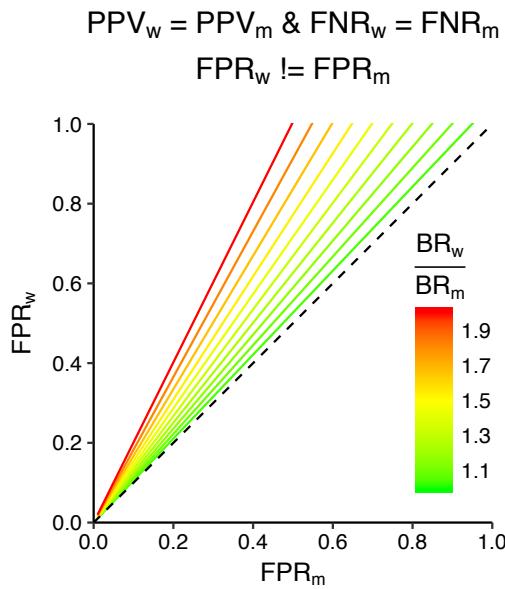
$$\text{FPR}_w \neq \text{FPR}_m$$



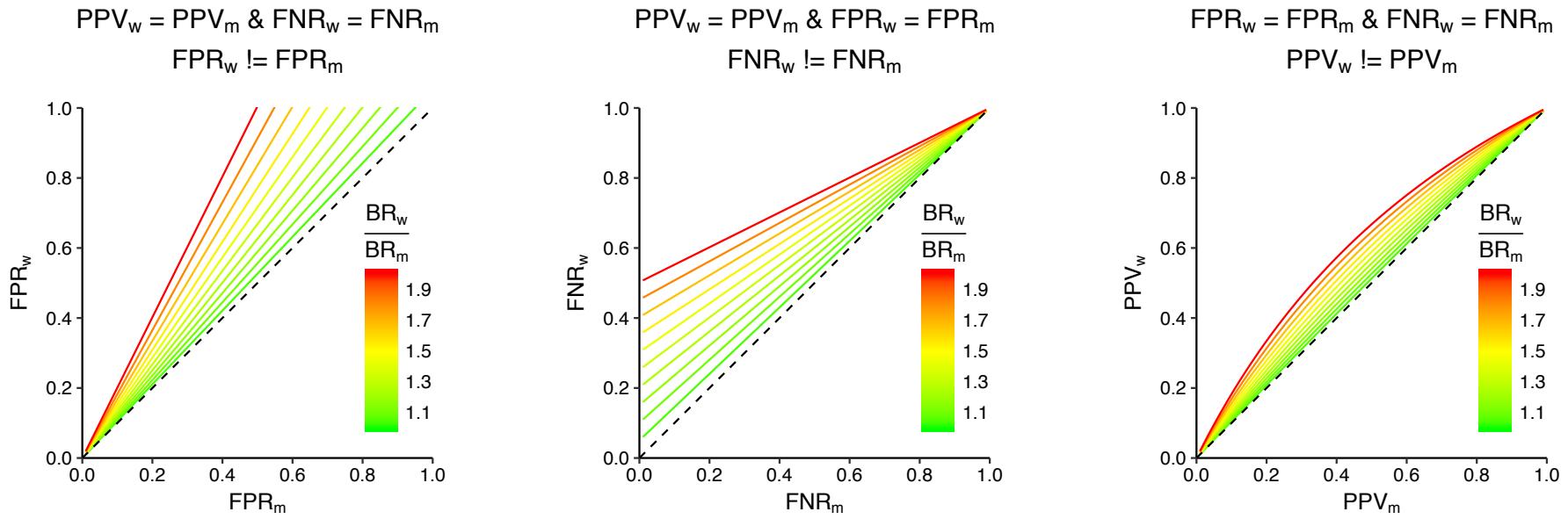
# The sinister tradeoff



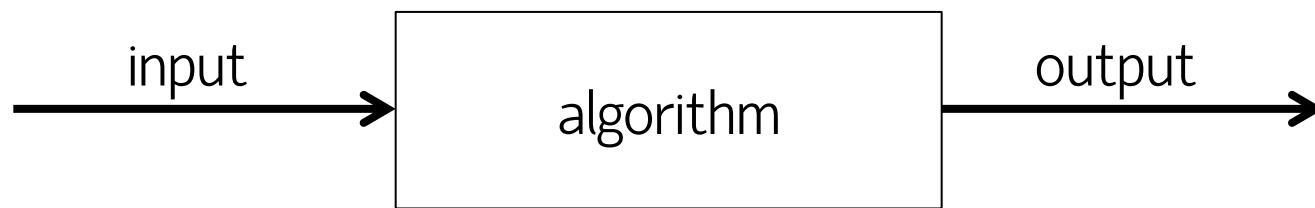
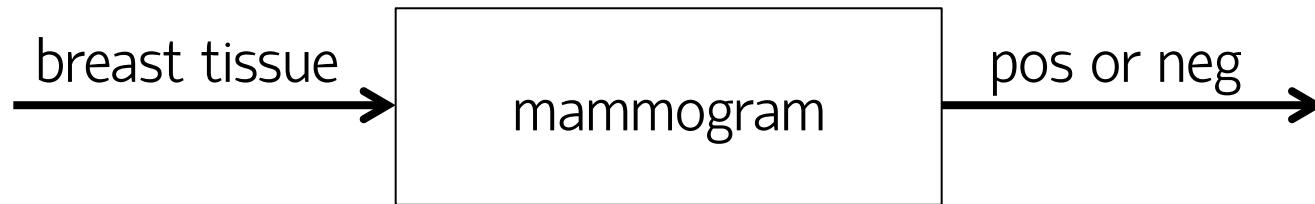
# The sinister tradeoff



# The sinister tradeoff



Of the three "worlds" depicted above, which is preferred?  
Which notion of fairness is valued more? Which is valued less?



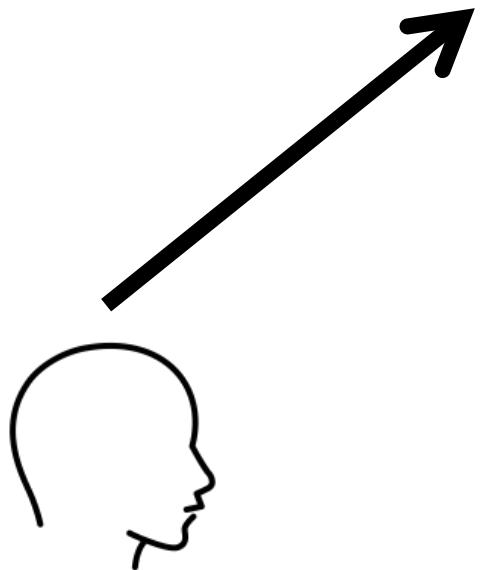
Bail eligibility  
Bank loan  
Whether a post is fake news

# What judgments do people make when Bayesian principles and egalitarian values are at stake?

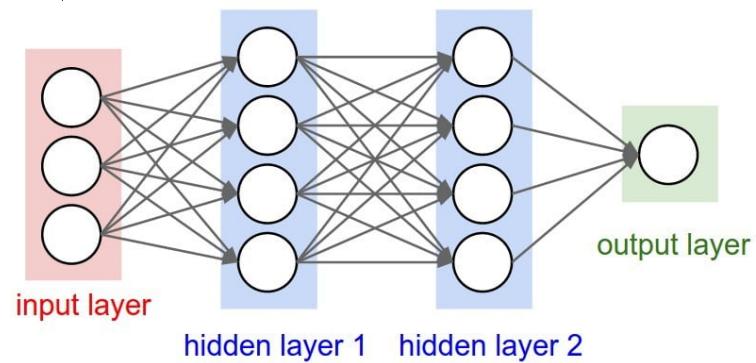
1. People undermine their commitment to egalitarian values by making Bayesian judgments
2. Formalizing "egalitarian values"
3. Google Images as a proxy for the social environment

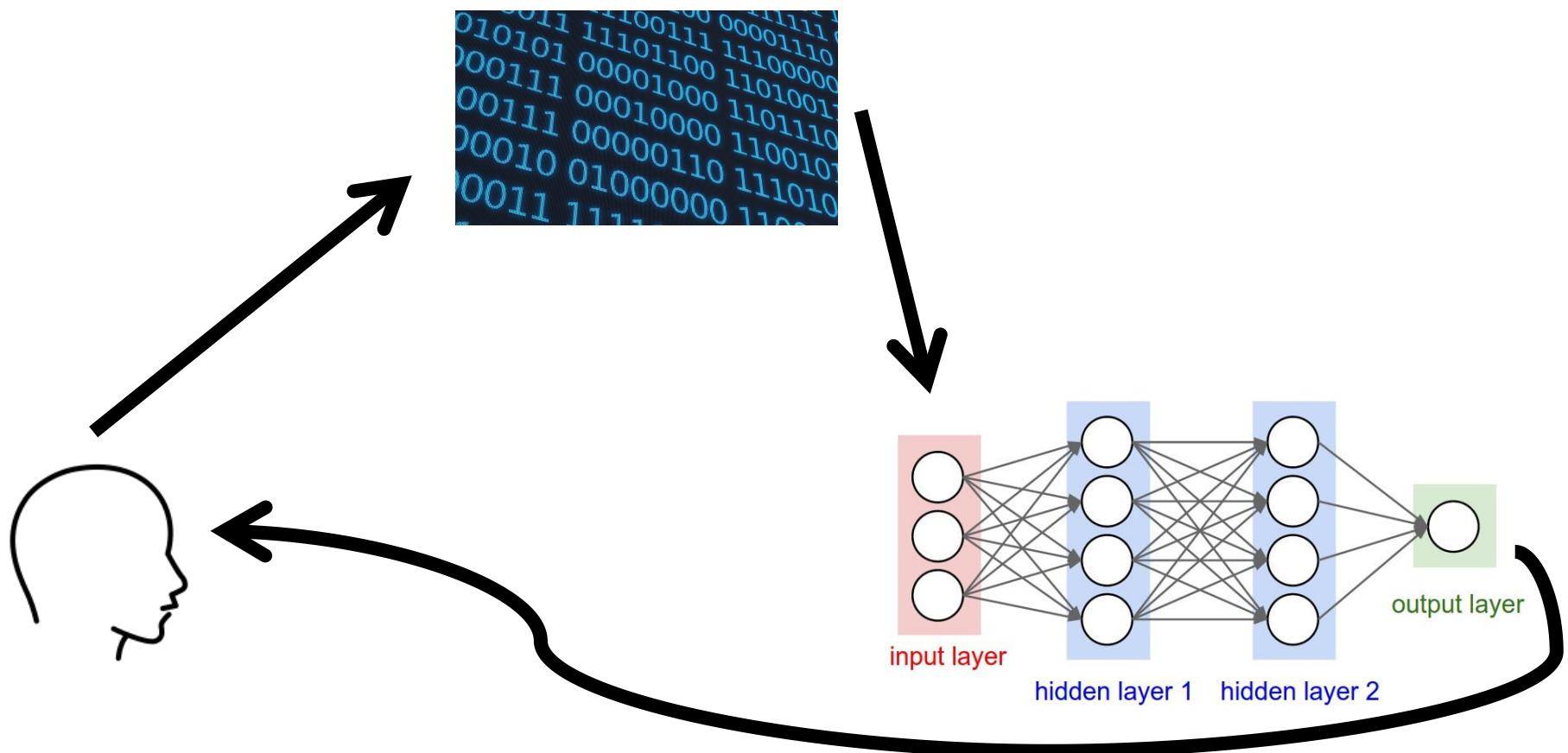
What judgments do people make when  
Bayesian principles and egalitarian values are at stake?

1. People undermine their commitment to egalitarian values by making Bayesian judgments
2. Formalizing "egalitarian values"
3. Google Images as a proxy for the social environment



0111 1100111 00001110  
010101 11101100 11100000  
000111 00001000 1101001  
00111 00010000 1101110  
00010 00000110 1100101  
0011 1111 1100







doctor

All

Maps

Images

News

Videos

More





doctor

All

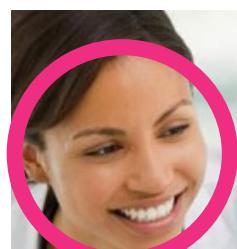
Maps

Images

News

Videos

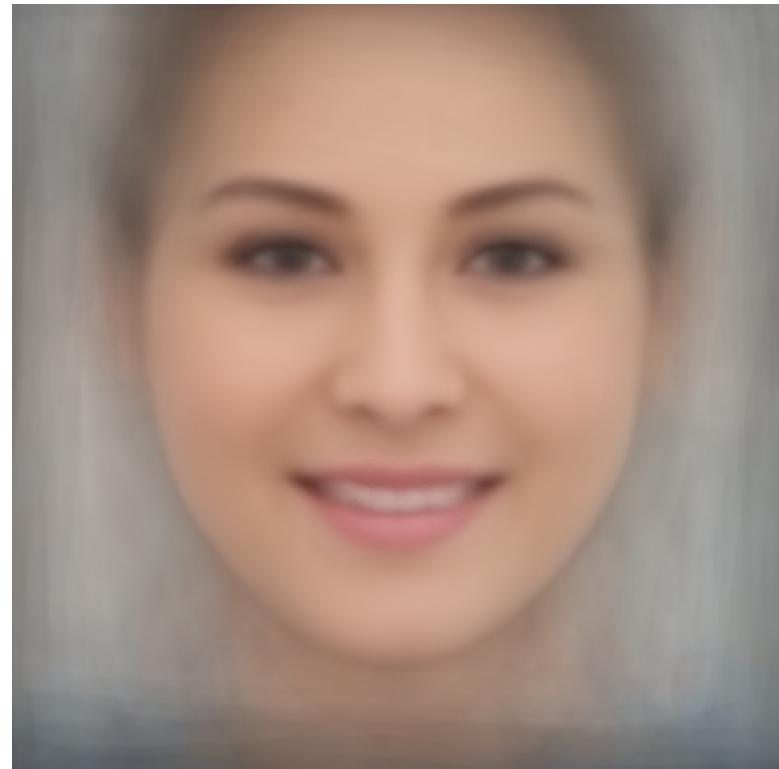
More



"doctor"



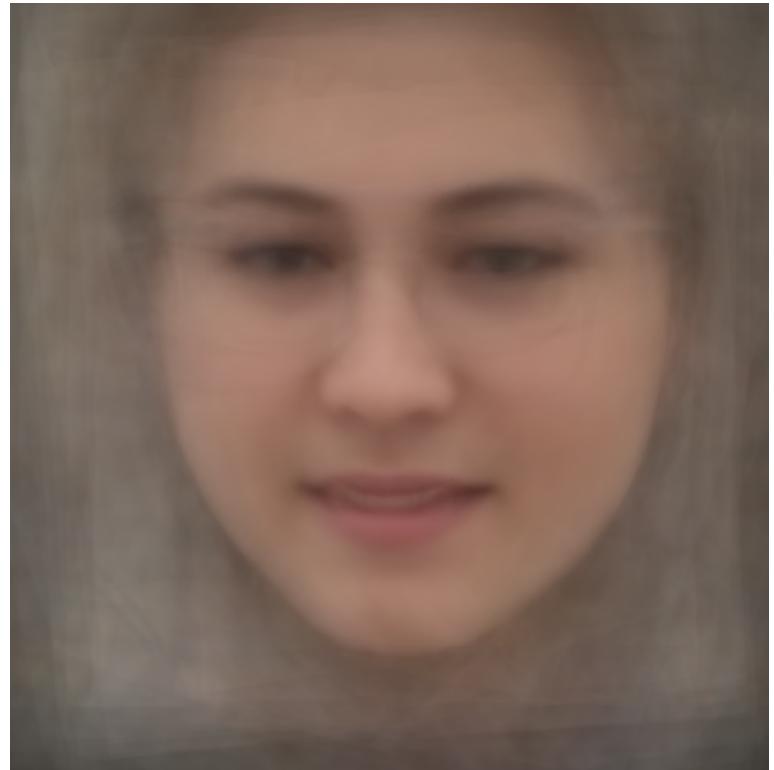
"nurse"



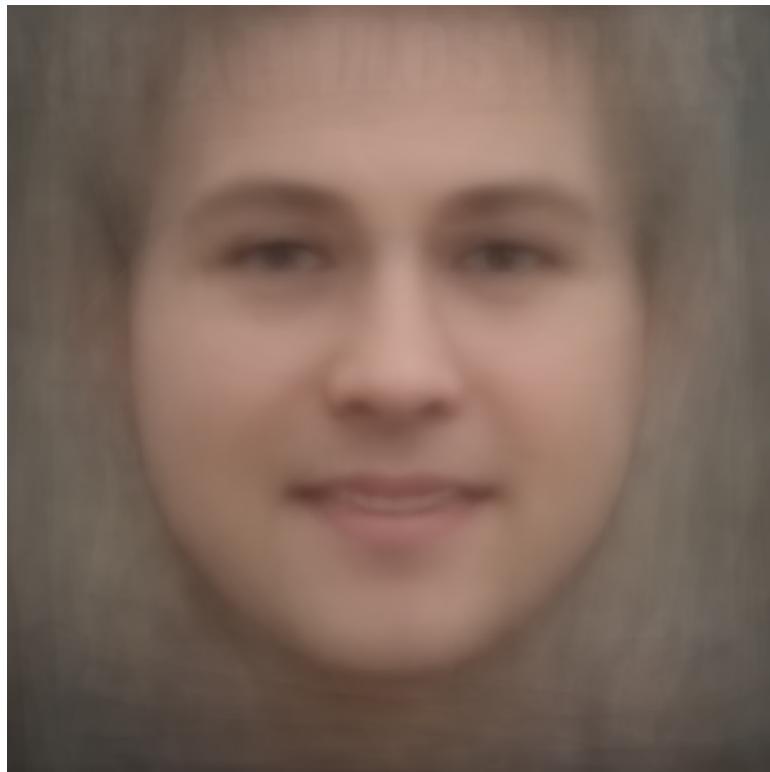
"computer science  
student"



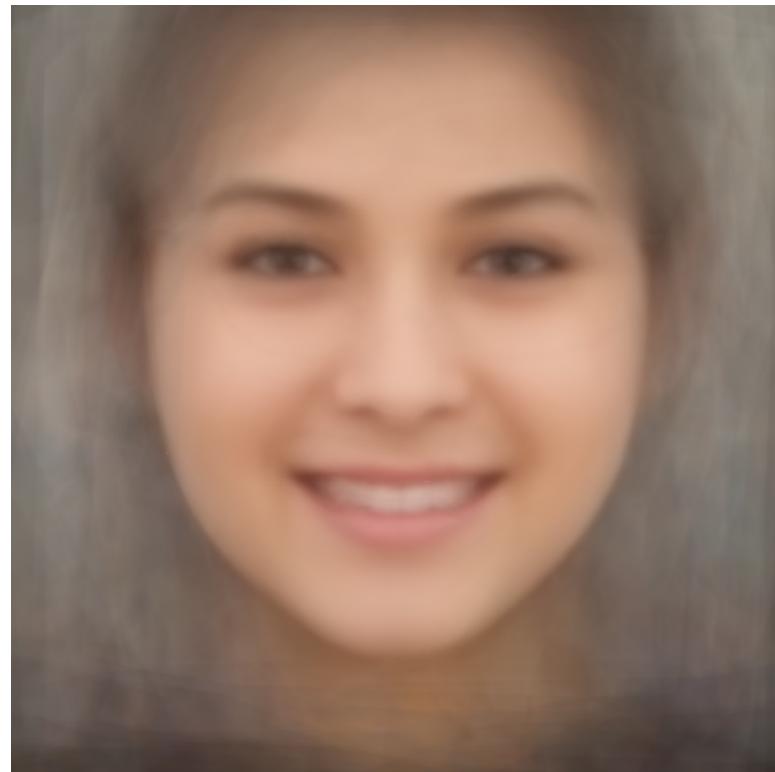
"biology  
student"



"philosophy  
student"



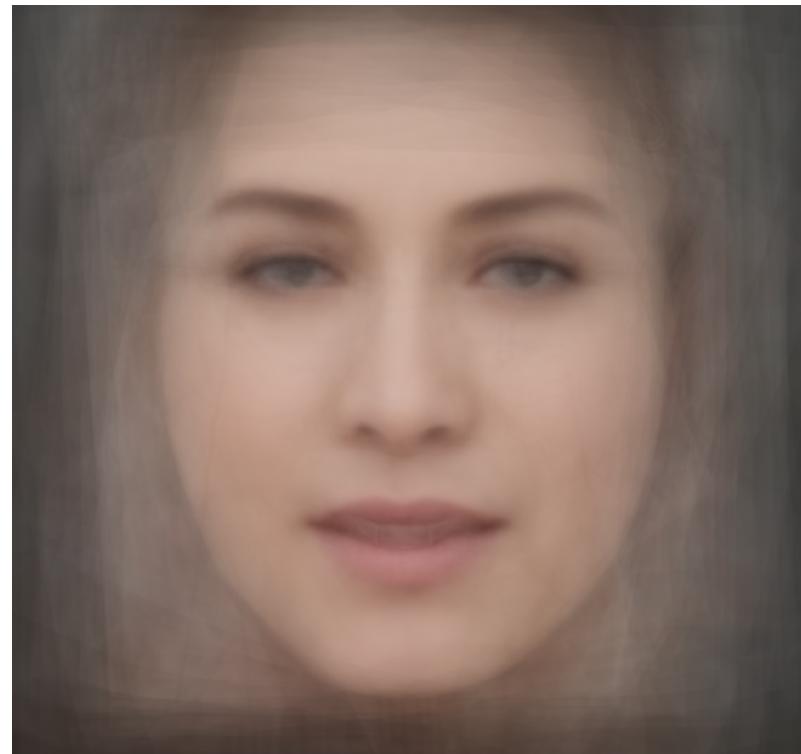
"education  
student"



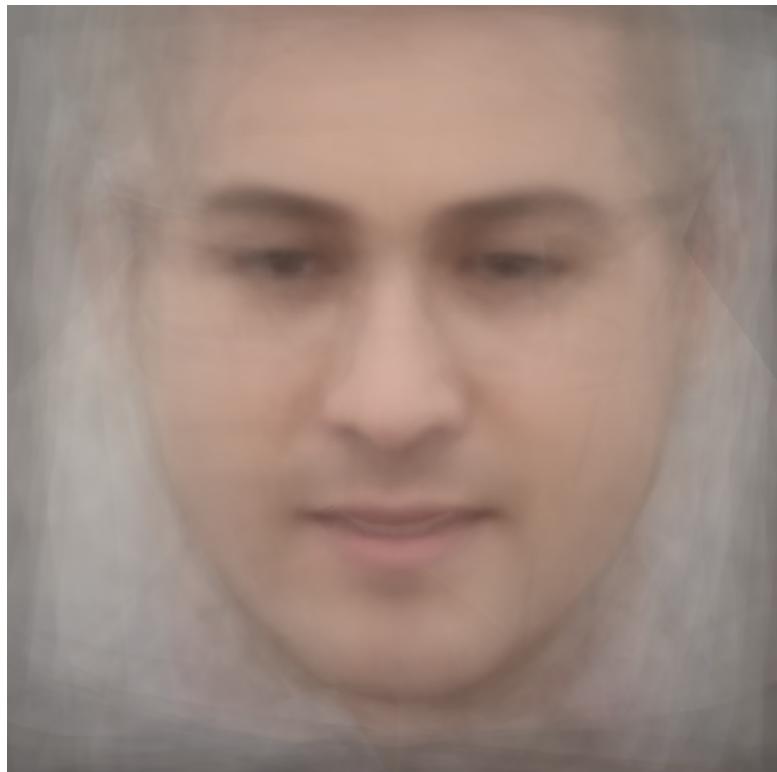
"intelligent  
person"



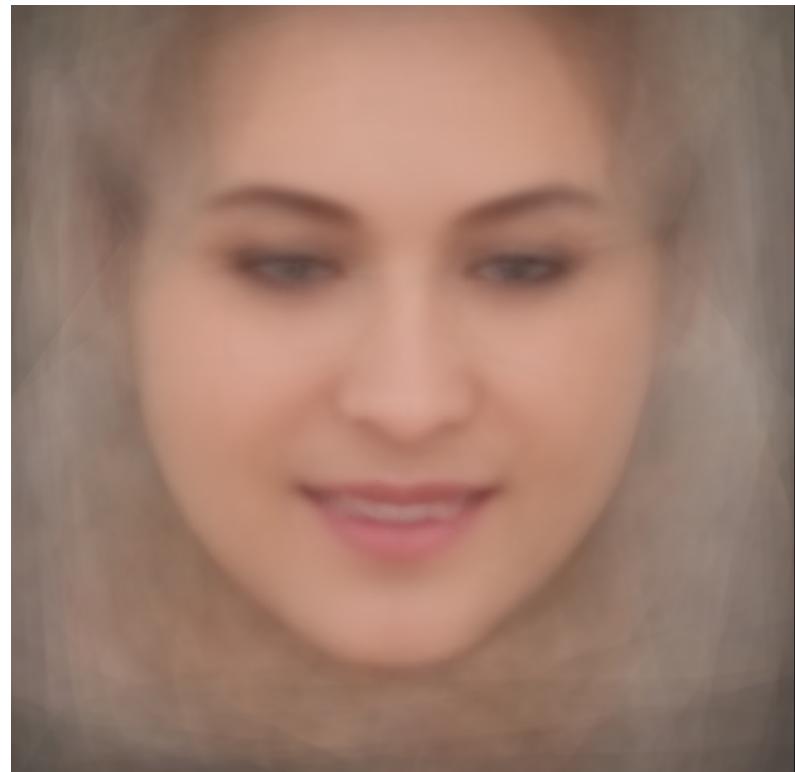
"sensitive  
person"



"person reading a  
newspaper"



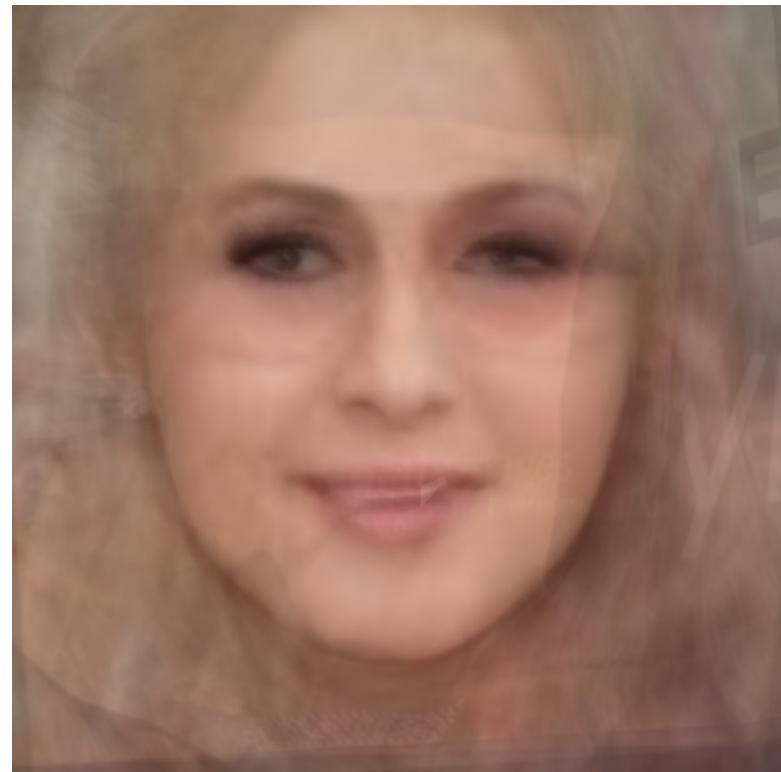
"person reading a  
cookbook"



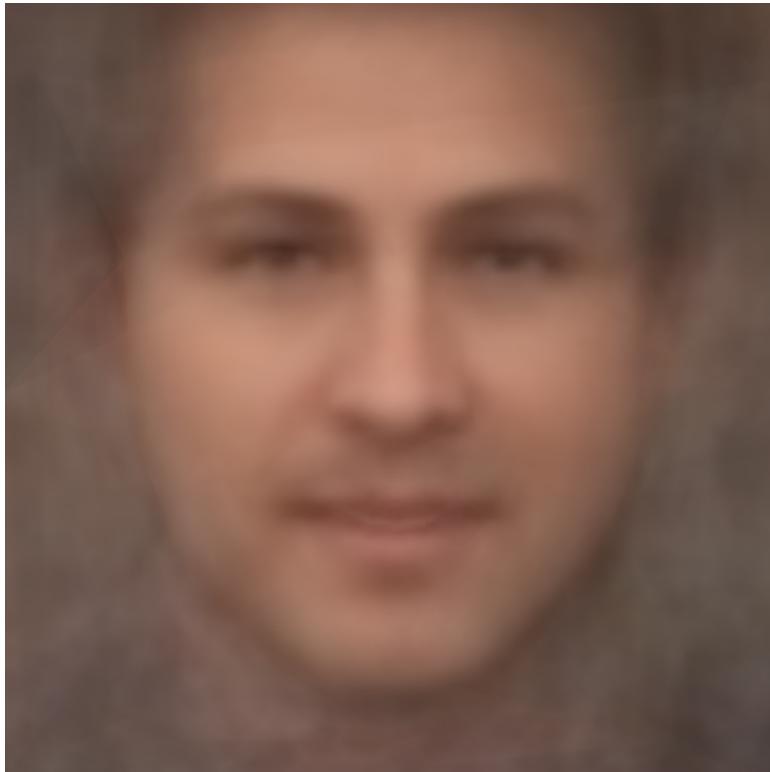
"bulldog  
owner"



"chihuahua  
owner"



"person drinking  
whiskey"



"person drinking  
cosmopolitan"





# Summary

## Undermining of egalitarian values

Cao, Kleiman-Weiner, & Banaji (2019), Psych Sci

From



to

A grid of binary digits (0s and 1s) arranged in a roughly triangular shape, representing digital data or algorithms.

Google  
images

# Closing thoughts

# Closing thoughts

Life is mostly between-subjects.

Make it more within-subjects.

# Closing thoughts

Life is mostly between-subjects.

Make it more within-subjects.

Assess people's preferences before product launch.

# Closing thoughts

Life is mostly between-subjects.

Make it more within-subjects.

Assess people's preferences before product launch.

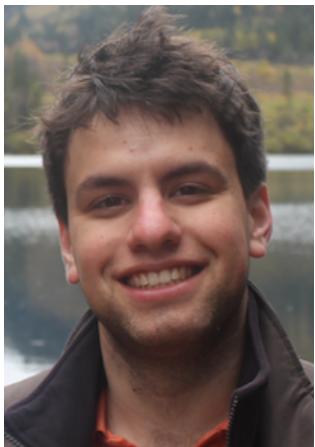
Look for opportunities where incentives are aligned.

# Thank you



Photo Credit: Robert Taylor [www.taylor-photo.co.uk](http://www.taylor-photo.co.uk)

Mahzarin R. Banaji



Max Kleiman-Weiner

Harvard Social Cognition Lab

Jason Mitchell & Jim Sidanius

Research Assistants

Kirsten Morehouse

Juan Lopez Martin



HARVARD  
Mind Brain Behavior



HARVARD Kennedy School  
**MALCOLM WIENER CENTER**  
for Social Policy