

# Exploring Just Enough?

## How Implicit Search Cost Can Limit Diversity

Xuechunzi Bai

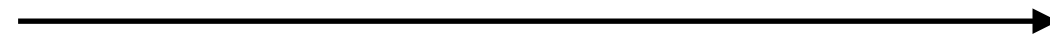
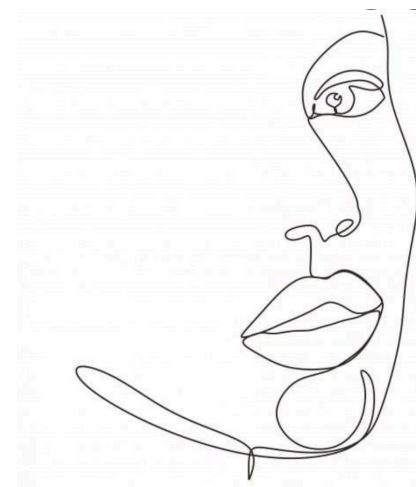
Princeton University

# The Psychology of How We **Make Sense of** the Social World

# The Psychology of How We **Make Sense of** the Social World

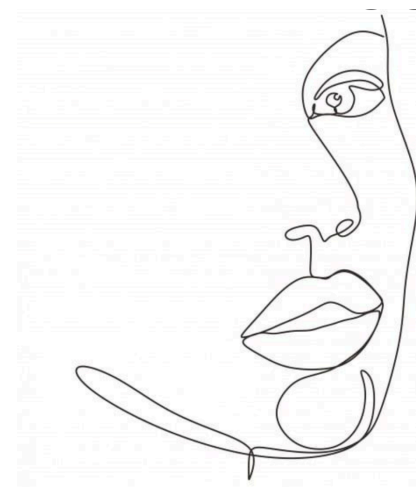
# The Psychology of How We **Make Sense of** the Social World

How do individual minds adaptively but selectively construct the social world?



# The Psychology of How We **Make Sense of** the Social World

How do individual minds adaptively but selectively construct the social world?

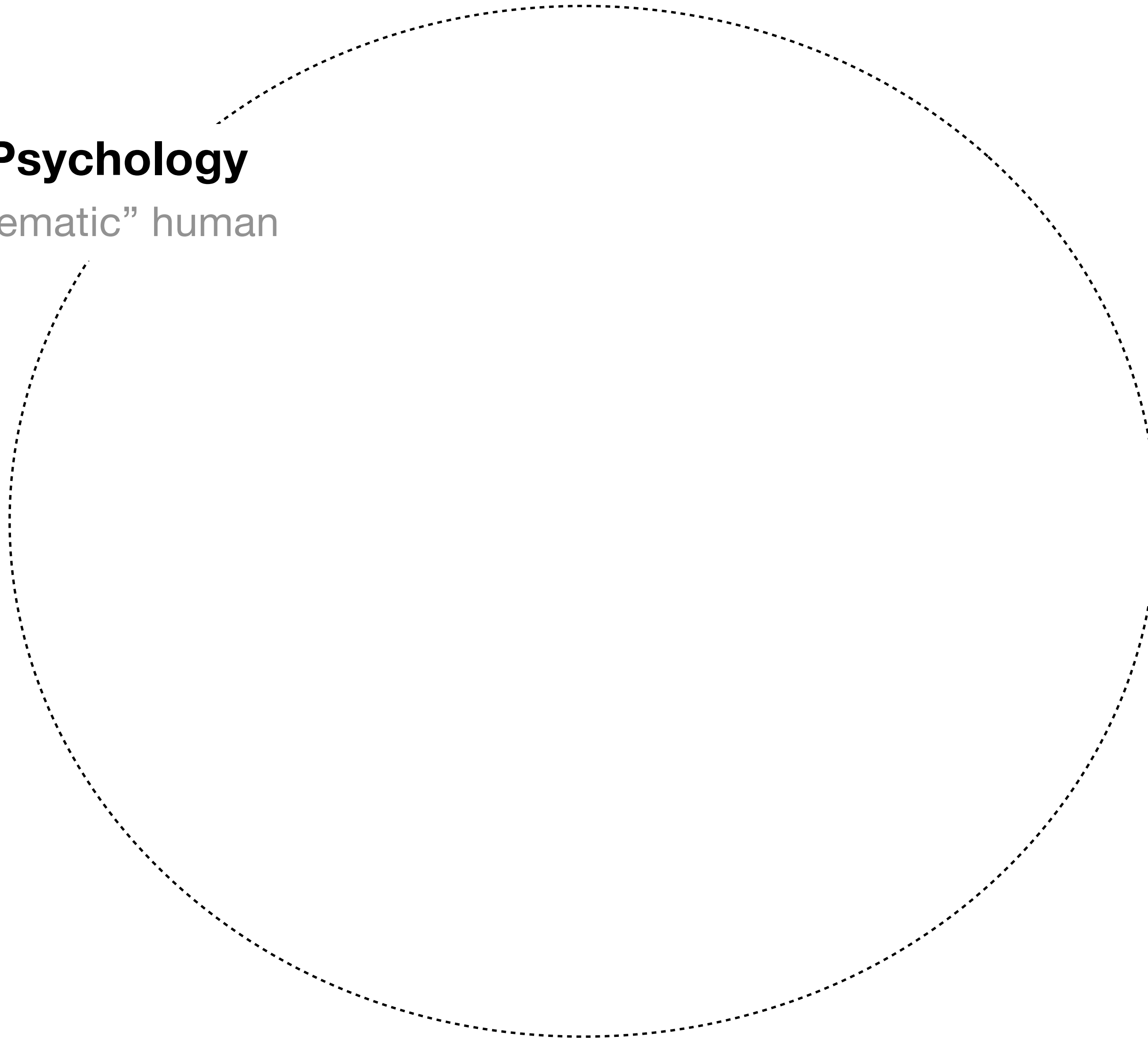


How does the constructed social world mold individual minds?

# The Psychology of How We **Make Sense of** the Social World

## **Social Psychology**

The “problematic” human



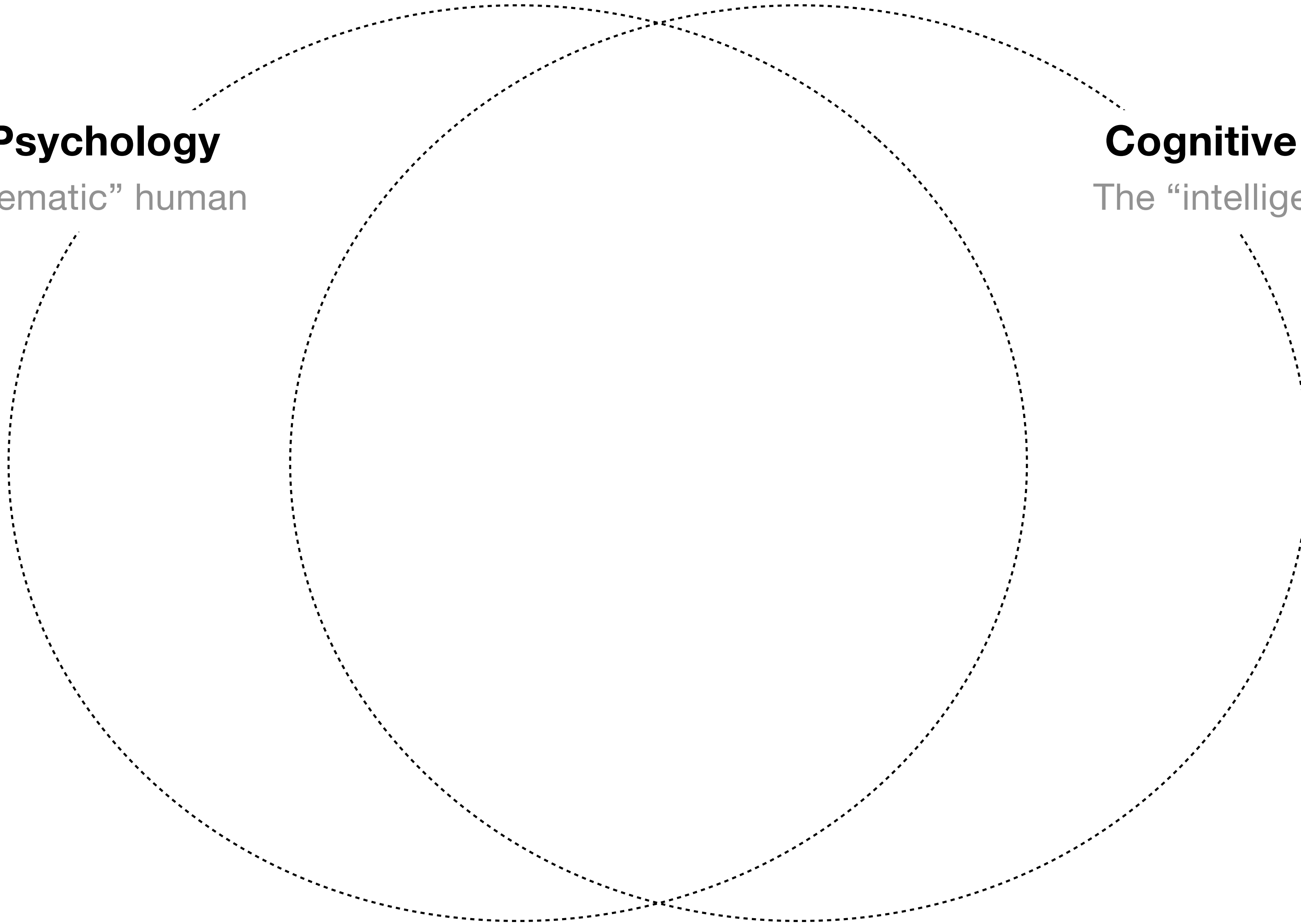
# The Psychology of How We **Make Sense of** the Social World

## **Social Psychology**

The “problematic” human

## **Cognitive Science**

The “intelligent” human



# The Psychology of How We **Make Sense** of the Social World

**Social Psychology**  
The “problematic” human

**Cognitive Science**  
The “intelligent” human

**Where do stereotypes come from?**



# The Psychology of How We **Make Sense of** the Social World

## **Social Psychology**

The “problematic” human

## **Cognitive Science**

The “intelligent” human

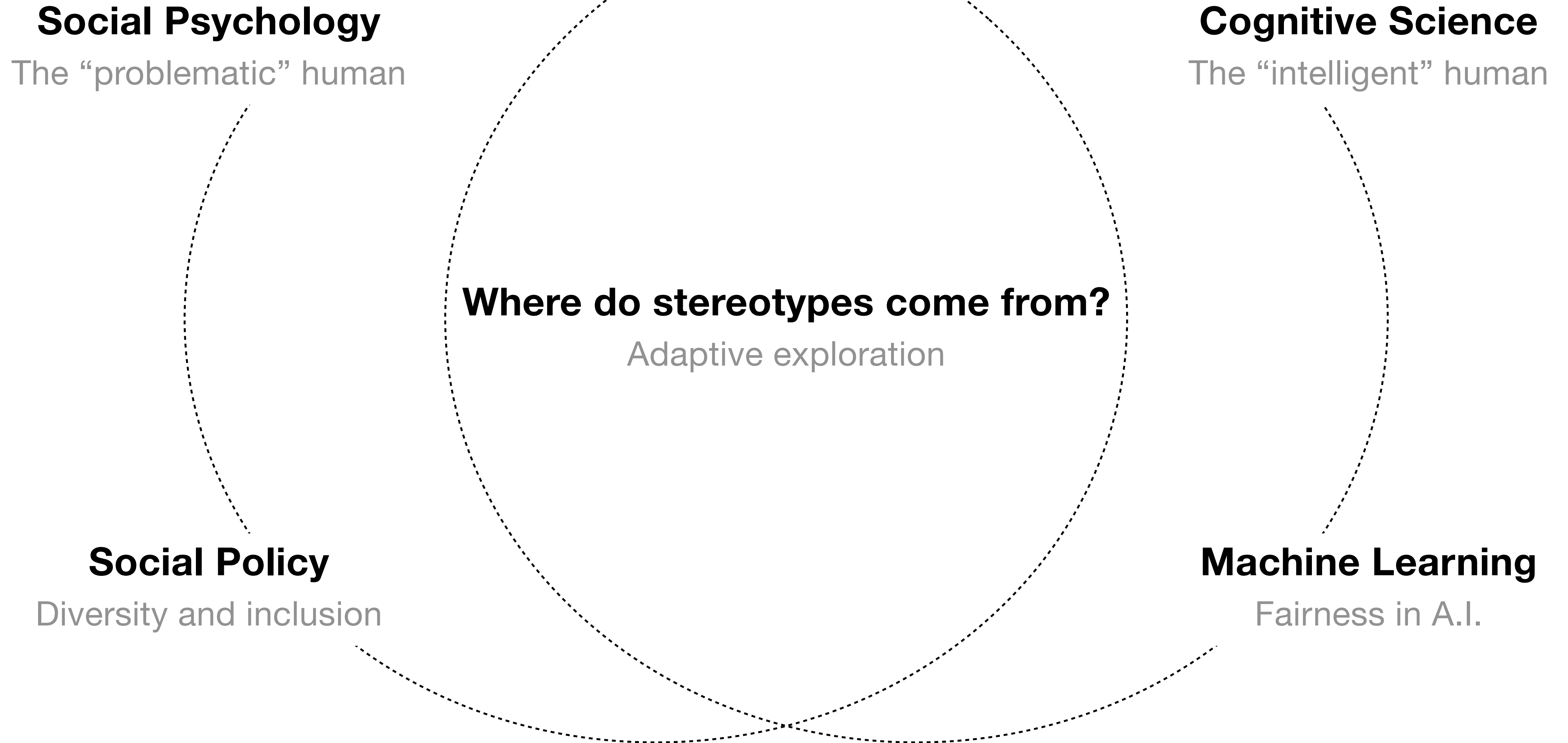
**Where do stereotypes come from?**

Adaptive exploration

# The Psychology of How We **Make Sense of** the Social World



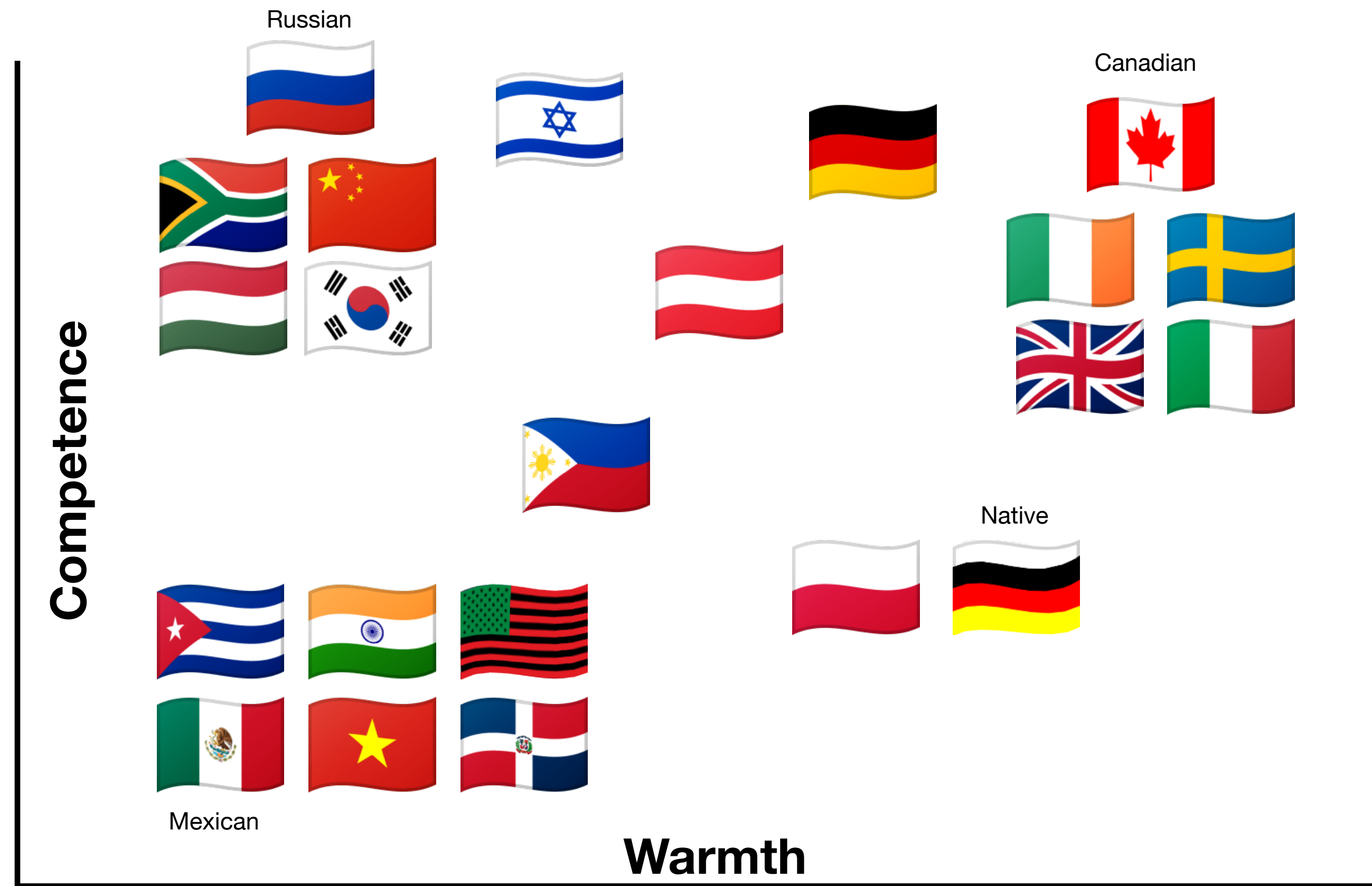
# The Psychology of How We **Make Sense of** the Social World



# Where do stereotypes come from?

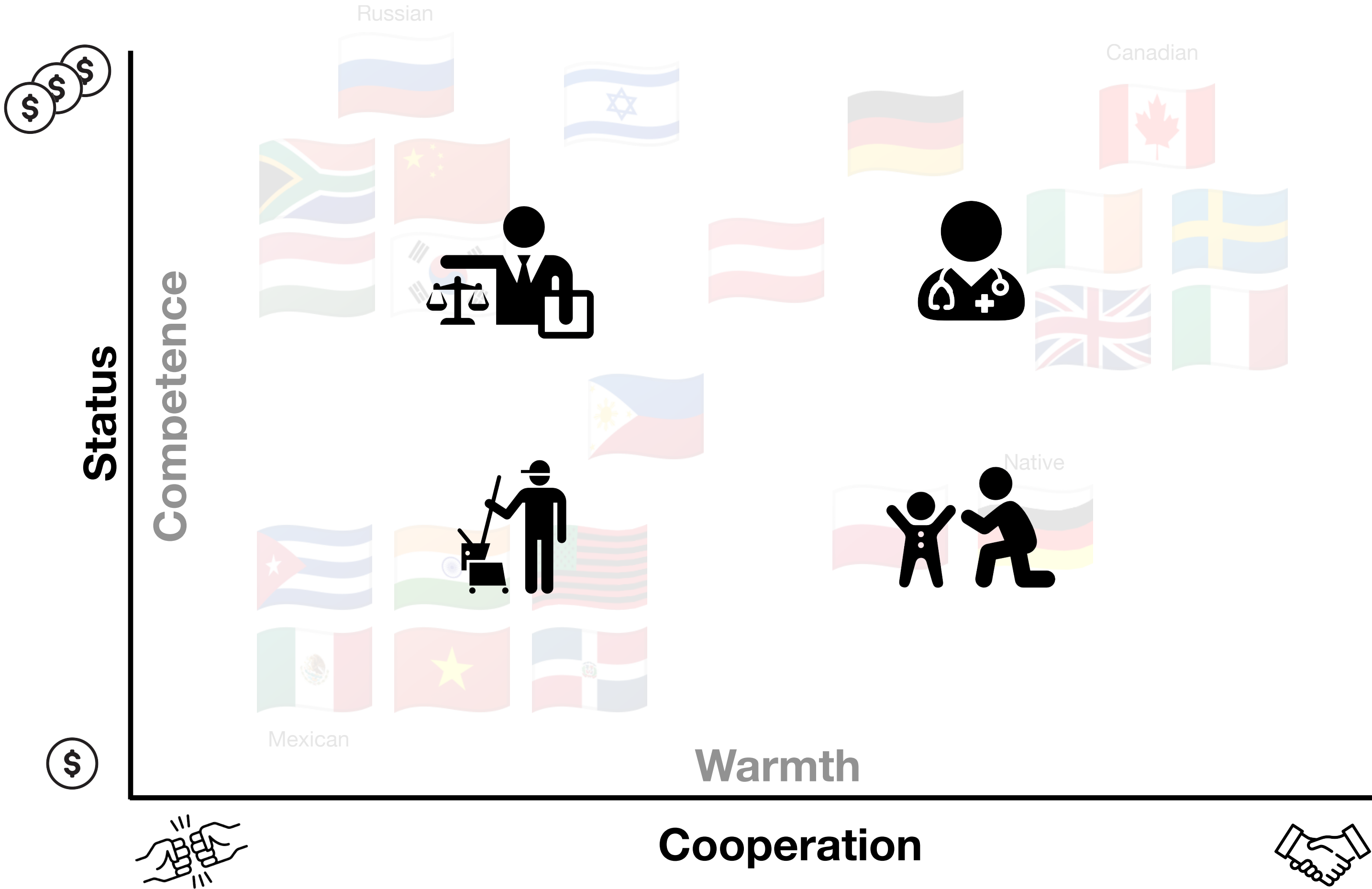


# Where do stereotypes come from?





# Where do stereotypes come from?



**Where do stereotypes come from?**

**Why do social groups end up  
with their current positions?**



# Where do stereotypes come from?

Background | Intuition | Formalism | Simulation | Experiments | Implications

## **Background: Psychological Mechanisms**

## Existing Explanations:

### Motivational biases

- Identity
- Dominance

Social identity theory (Tajfel & Turner, 1979)  
In-group favoritism (Brewer, 1999)  
Social dominance theory (Sidanius & Pratto, 1999)  
System justification theory (Jost & Banaji, 1994)

## Existing Explanations:

### Motivational biases

- Identity
- Dominance

### Cognitive biases

- Limited memory
- Selective attention

Social identity theory (Tajfel & Turner, 1979)  
In-group favoritism (Brewer, 1999)  
Social dominance theory (Sidanius & Pratto, 1999)  
System justification theory (Jost & Banaji, 1994)

Error-prone heuristics (Tversky & Kahneman, 1974)  
Cognitive miser (Fiske & Taylor, 1984)  
Illusory correlation (Hamilton & Gifford, 1976)  
Attention in stereotype formation (Sherman et al., 2009)

## Existing Explanations:

### Motivational biases

- Identity
- Dominance

### Cognitive biases

- Limited memory
- Selective attention

### Sample biases

- Unequal group size

Social identity theory (Tajfel & Turner, 1979)  
In-group favoritism (Brewer, 1999)  
Social dominance theory (Sidanius & Pratto, 1999)  
System justification theory (Jost & Banaji, 1994)

Error-prone heuristics (Tversky & Kahneman, 1974)  
Cognitive miser (Fiske & Taylor, 1984)  
Illusory correlation (Hamilton & Gifford, 1976)  
Attention in stereotype formation (Sherman et al., 2009)

Beware of samples (Fiedler, 2000)  
Hot-stove effect (Denrell, 2005)

## Existing Explanations:

### Motivational biases

- Identity
- Dominance

Social identity theory (Tajfel & Turner, 1979)  
In-group favoritism (Brewer, 1999)  
Social dominance theory (Sidanius & Pratto, 1999)  
System justification theory (Jost & Banaji, 1994)

### Cognitive biases

- Limited memory
- Selective attention

Error-prone heuristics (Tversky & Kahneman, 1974)  
Cognitive miser (Fiske & Taylor, 1984)  
Illusory correlation (Hamilton & Gifford, 1976)  
Attention in stereotype formation (Sherman et al., 2009)

### Sample biases

- Unequal group size

Beware of samples (Fiedler, 2000)  
Hot-stove effect (Denrell, 2005)

### Group differences

- Gender

Biosocial constructionism (Wood & Eagly, 2012)  
Stereotype accuracy (Jussim, 2017)

## Existing Explanations:

### Motivational biases

- Identity
- Dominance

Social identity theory (Tajfel & Turner, 1979)  
In-group favoritism (Brewer, 1999)  
Social dominance theory (Sidanius & Pratto, 1999)  
System justification theory (Jost & Banaji, 1994)

### Cognitive biases

- Limited memory
- Selective attention

Error-prone heuristics (Tversky & Kahneman, 1974)  
Cognitive miser (Fiske & Taylor, 1984)  
Illusory correlation (Hamilton & Gifford, 1976)  
Attention in stereotype formation (Sherman et al., 2009)

### Sample biases

- Unequal group size

Beware of samples (Fiedler, 2000)  
Hot-stove effect (Denrell, 2005)

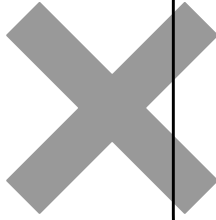
### Group differences

- Gender

Biosocial constructionism (Wood & Eagly, 2012)  
Stereotype accuracy (Jussim, 2017)

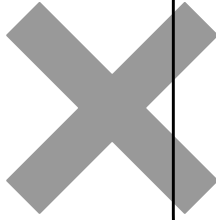
## **Our Proposal: Adaptive Exploration**



 Globally Accurate

 Morally Right

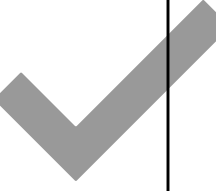
## **Our Proposal: Adaptive Exploration**

 Globally Accurate

 Morally Right

## **Our Proposal: Adaptive Exploration**

 Adaptive to self-interested decision-makers

 Detrimental to collective society

**This talk:  
One individual**

**Future: Consensus  
across individuals**

## **Our Proposal: Adaptive Exploration**

 **Globally Accurate**

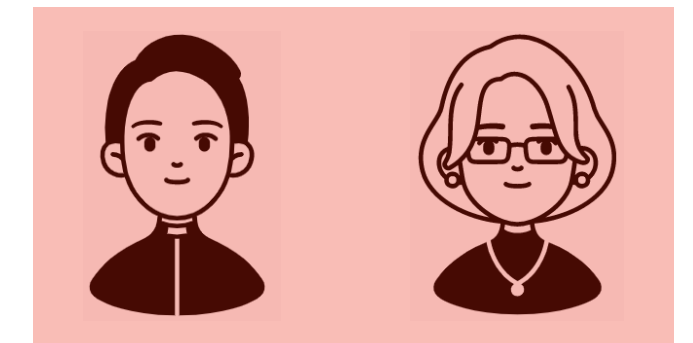
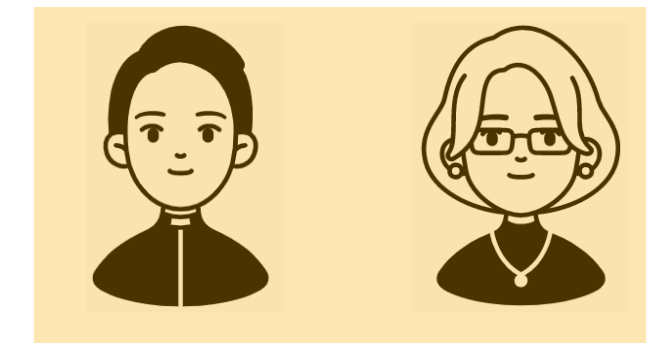
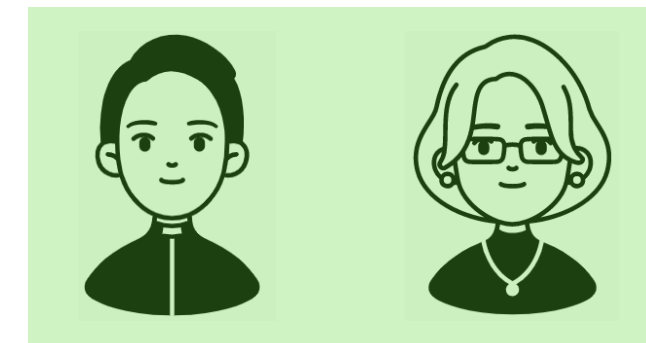
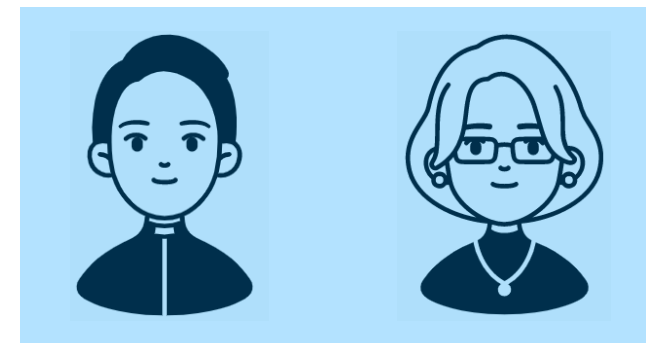
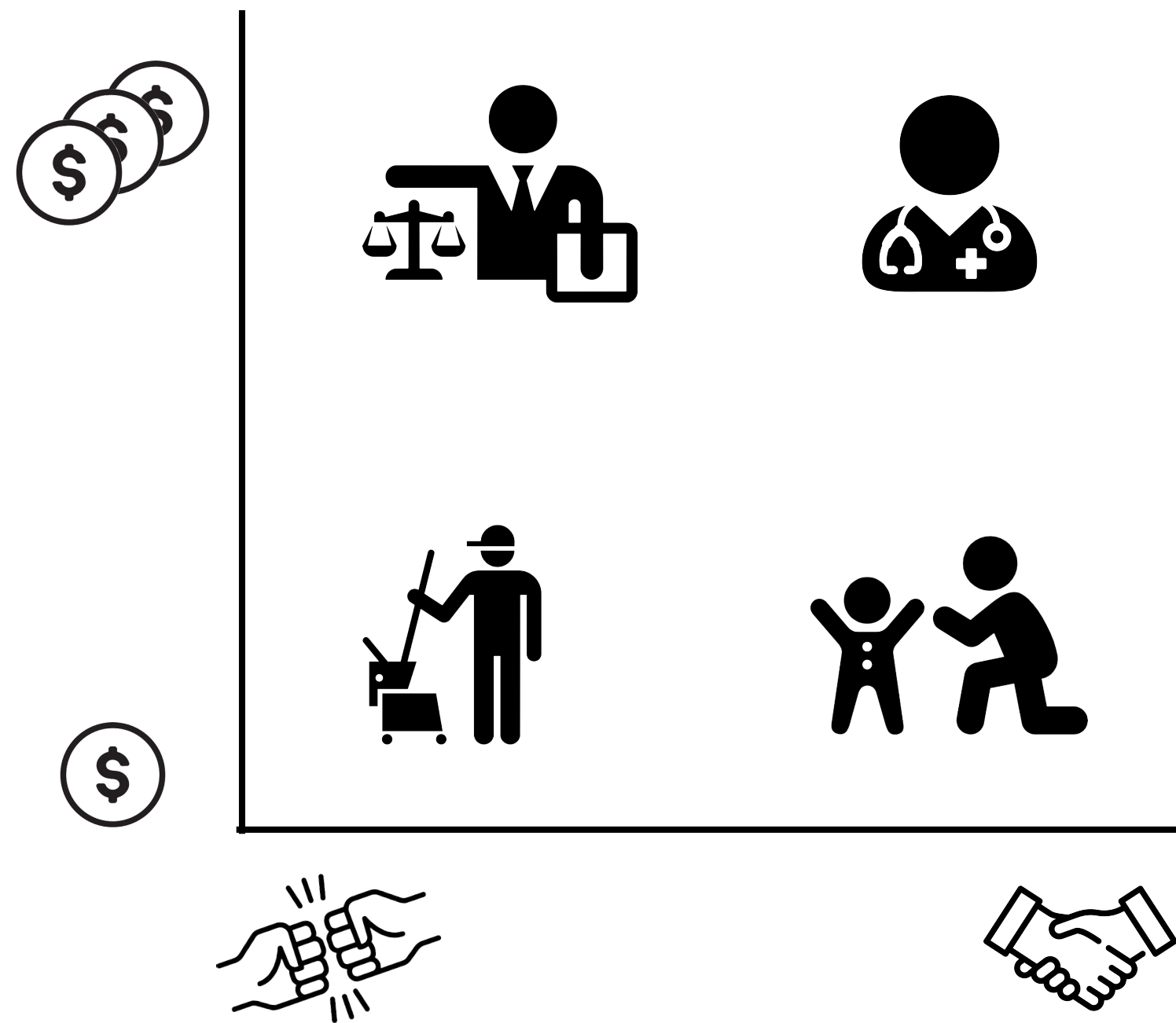
 **Morally Right**

 **Adaptive to self-  
interested  
decision-makers**

 **Detrimental to  
collective society**

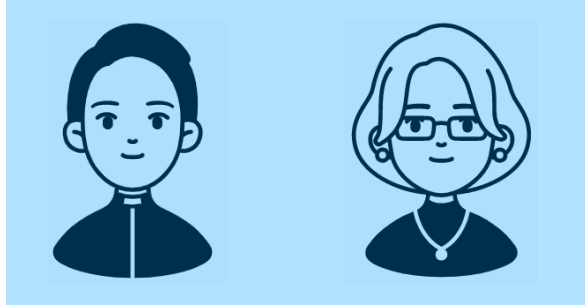

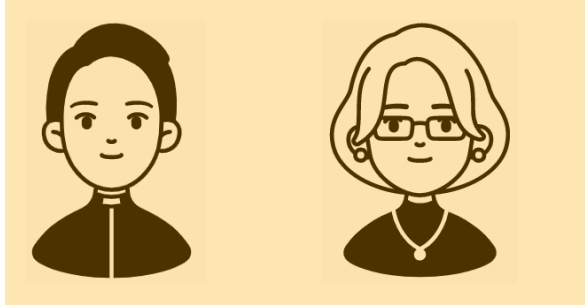
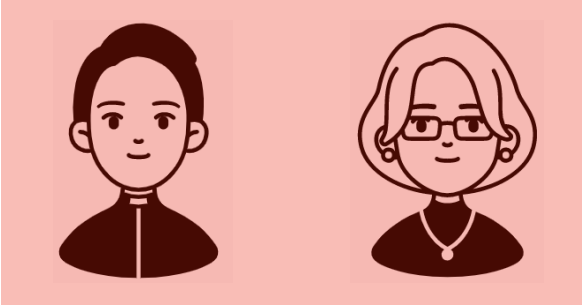



## Intuition: Hiring

Background | **Intuition** | Formalism | Simulation | Experiments | Implications



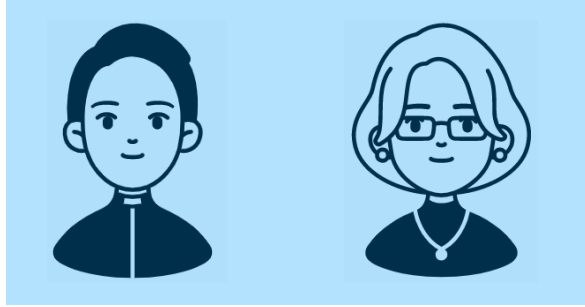

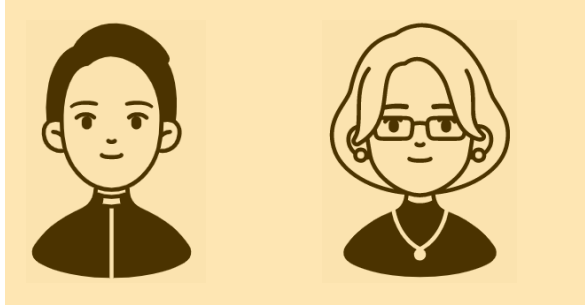
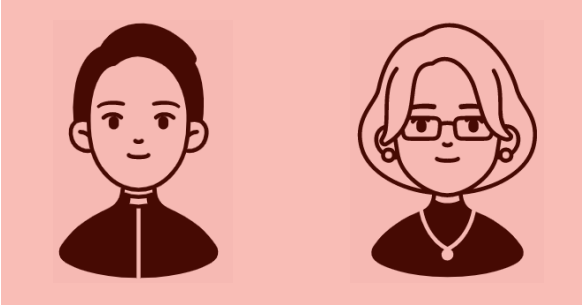




# Background | **Intuition** | Formalism | Simulation | Experiments | Implications



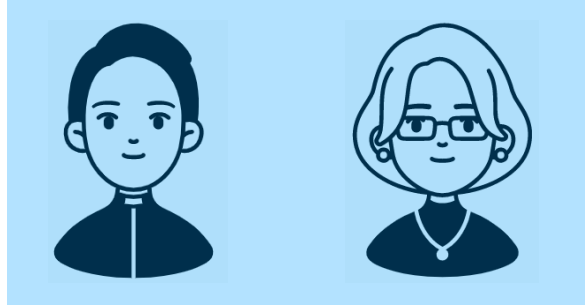

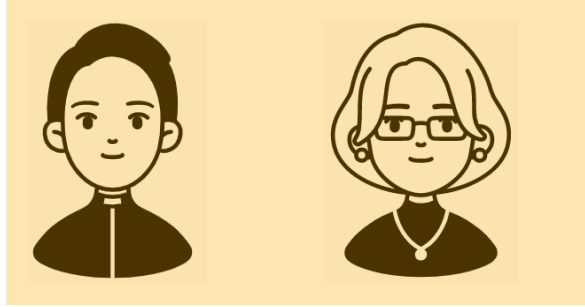
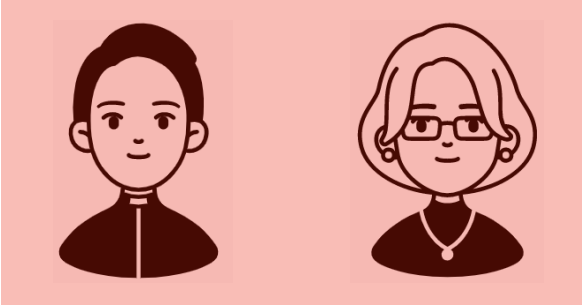







# Background | **Intuition** | Formalism | Simulation | Experiments | Implications



# Background | **Intuition** | Formalism | Simulation | Experiments | Implications

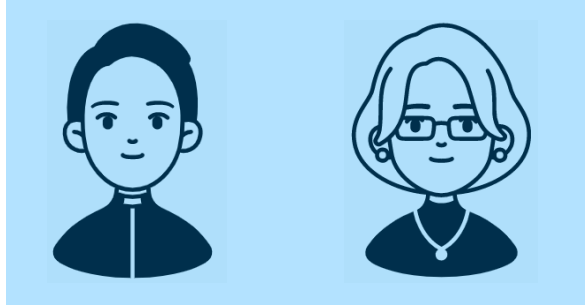

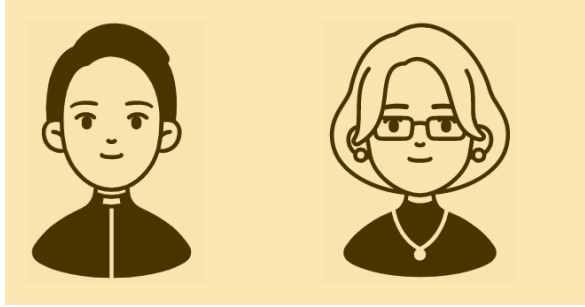
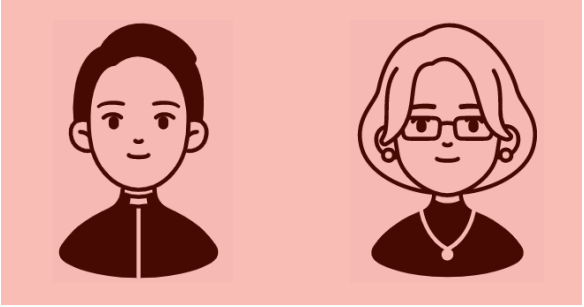














Background | **Intuition** | Formalism | Simulation | Experiments | Implications



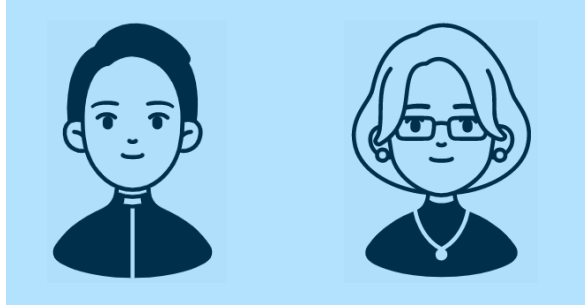

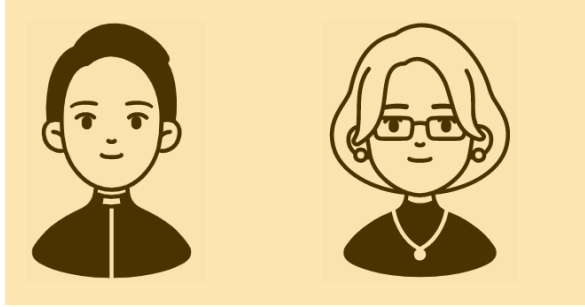
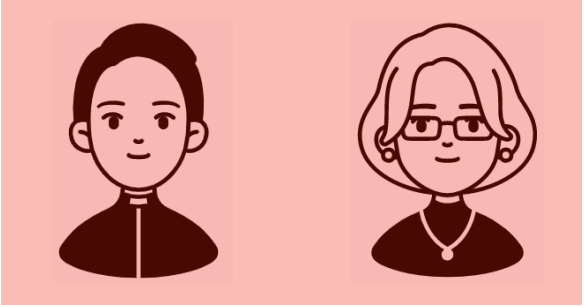






				
  				
  				

# Background | **Intuition** | Formalism | Simulation | Experiments | Implications




Background | **Intuition** | Formalism | Simulation | Experiments | Implications



				
  	✓	?	?	?
  	✗	?	?	?

# Background | **Intuition** | Formalism | Simulation | Experiments | Implications



		✓	?	?	?
		✗	?	?	?

# Background | Intuition | Formalism | Simulation | Experiments | Implications



	✓	?	?	?
	✗	?	?	?

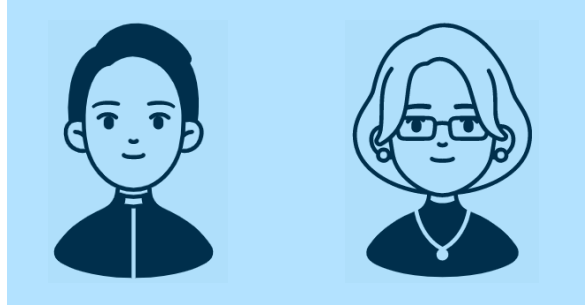

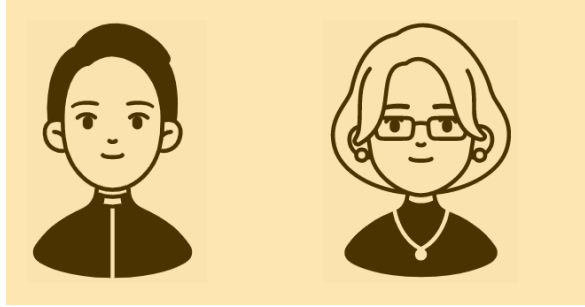
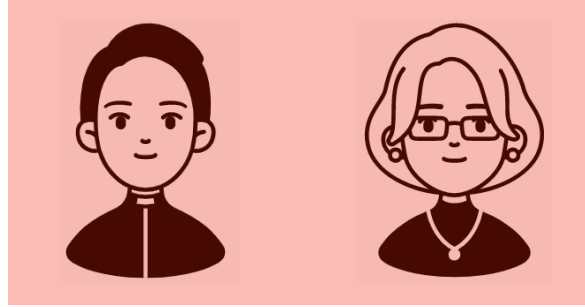












Background | **Intuition** | Formalism | Simulation | Experiments | Implications



	✓	?	?	?
	✗	?	?	?
	?	?	✓	?

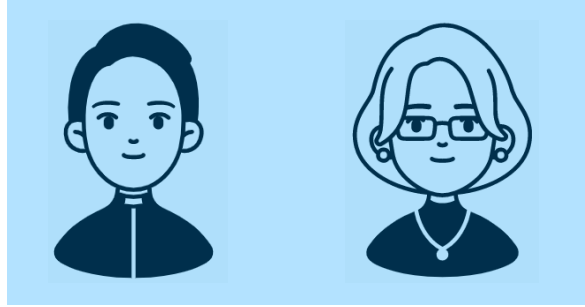

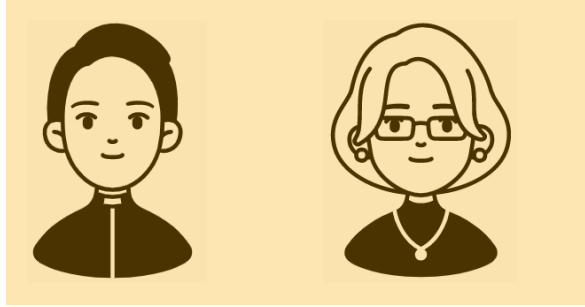
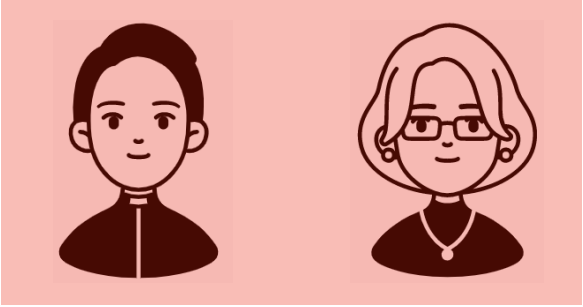

























# Background | **Intuition** | Formalism | Simulation | Experiments | Implications



					
  		✓	?	?	?
  		✗	?	?	?
  		?	?	✓	?
  					

Background | **Intuition** | Formalism | Simulation | Experiments | Implications





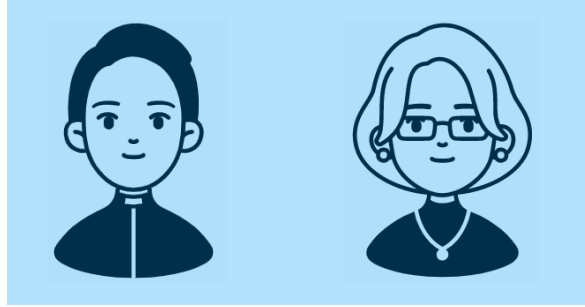

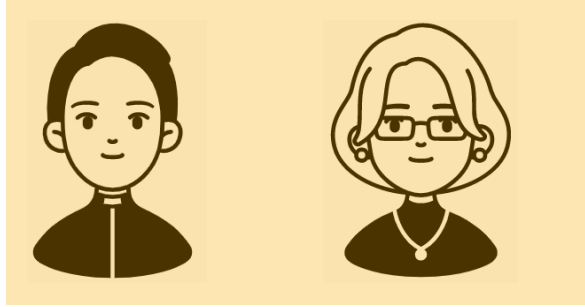
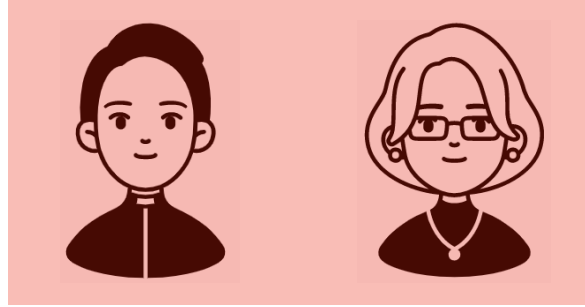




























Background | **Intuition** | Formalism | Simulation | Experiments | Implications



		?	?	?	
		?	?	?	
	?	?		?	
		?	?	?	

Background | **Intuition** | Formalism | Simulation | Experiments | Implications

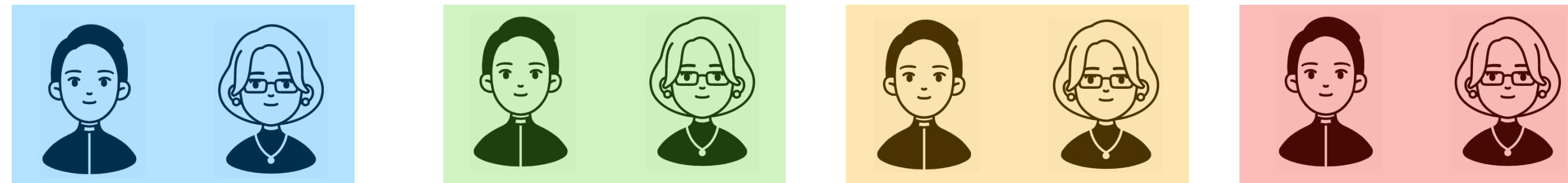


## **Formalism: Contextual Multi-Armed Bandit**

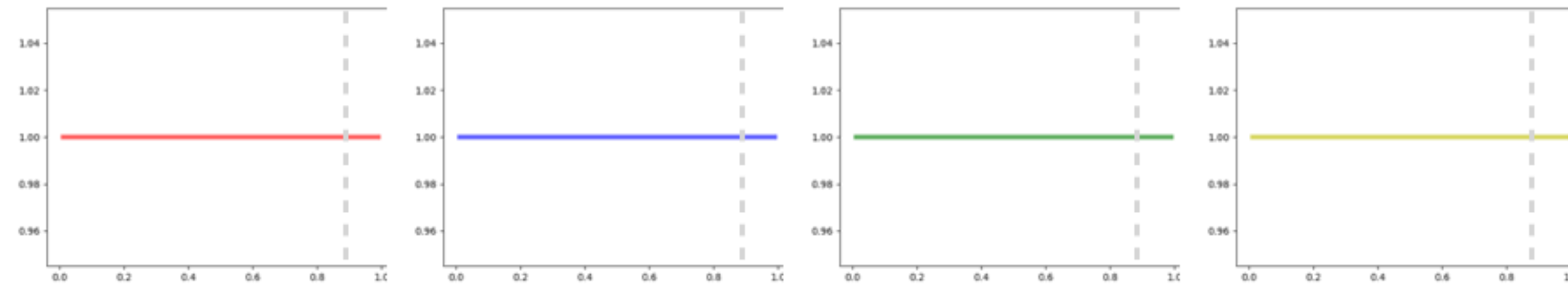
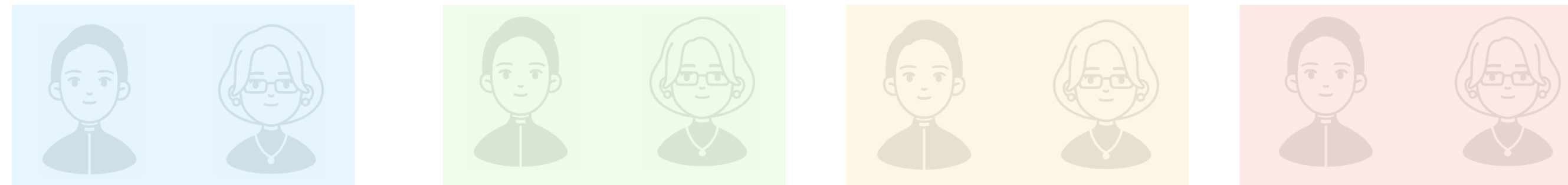
## **Formalism: Contextual Multi-Armed Bandit**

## Challenge One: Exploration vs. Exploitation



Explore v. Exploit dilemma in Reinforcement Learning (Sutton & Barto, 2018)

## Solution One: Thompson Sampling



90%

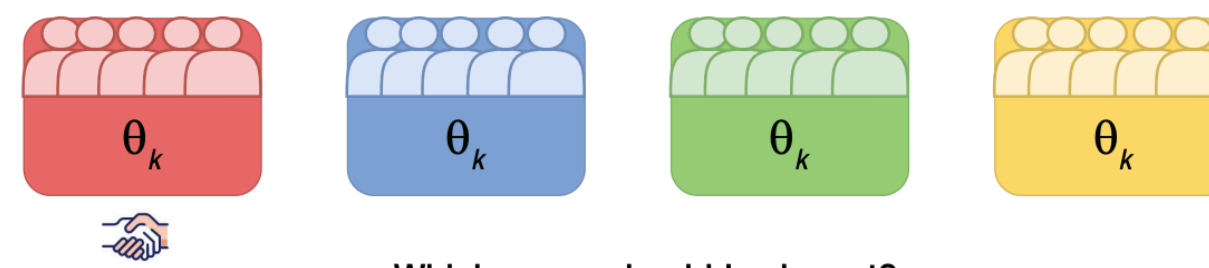
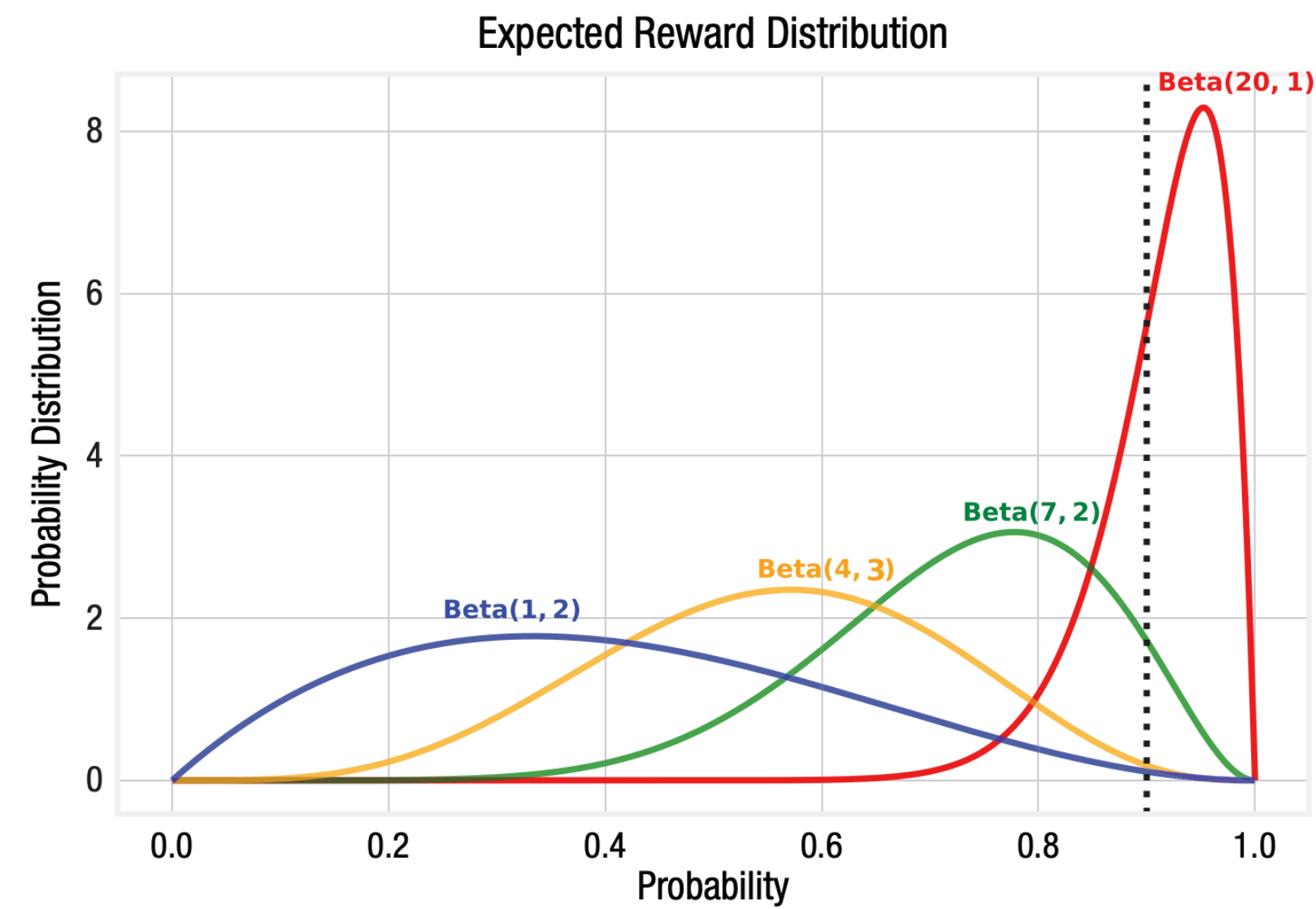
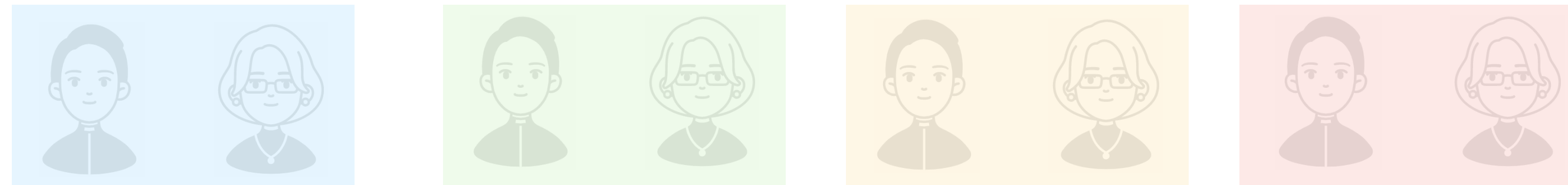
90%

90%

90%

Thompson sampling (Thompson, 1933; Agrawal & Goyal, 2012)

# Social Multi-Armed Bandit

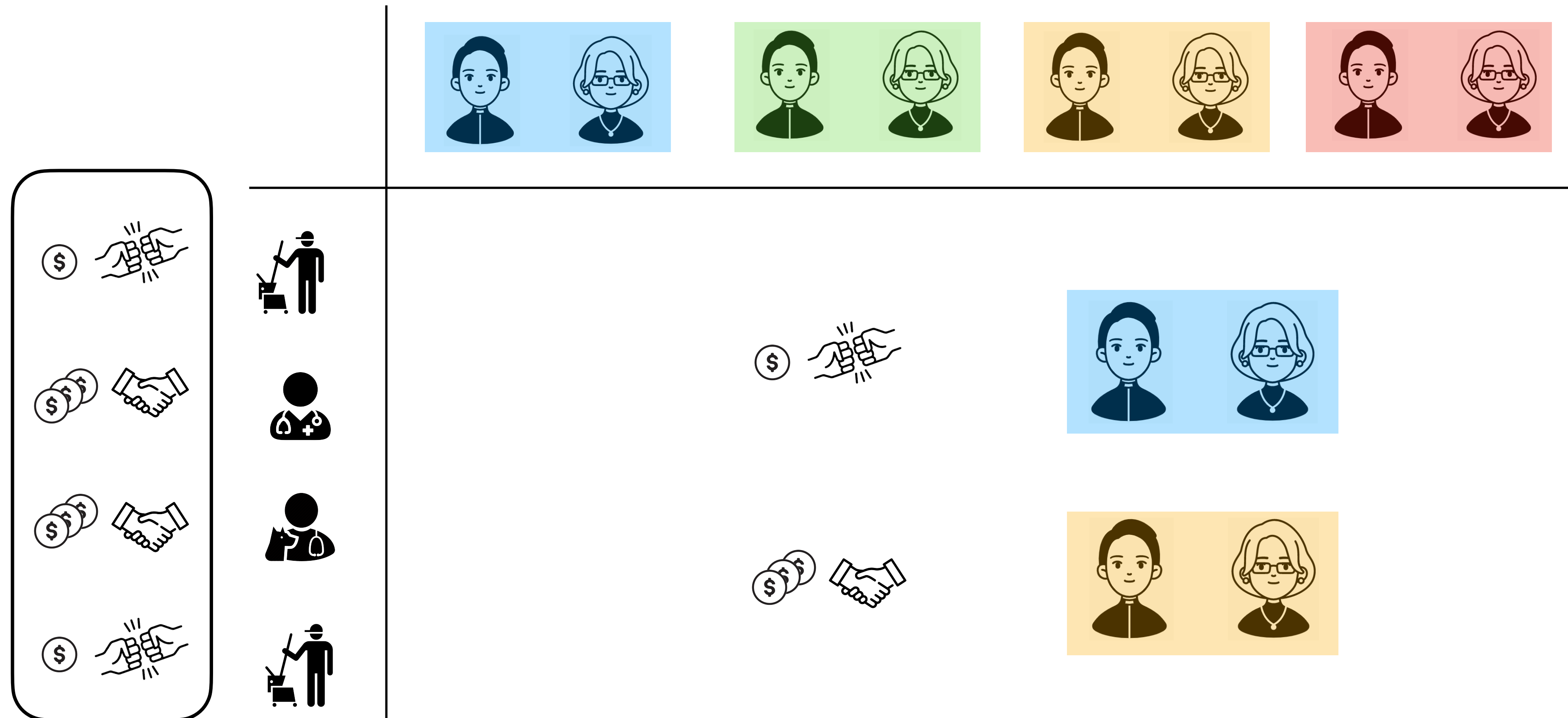


Which group should I ask next?

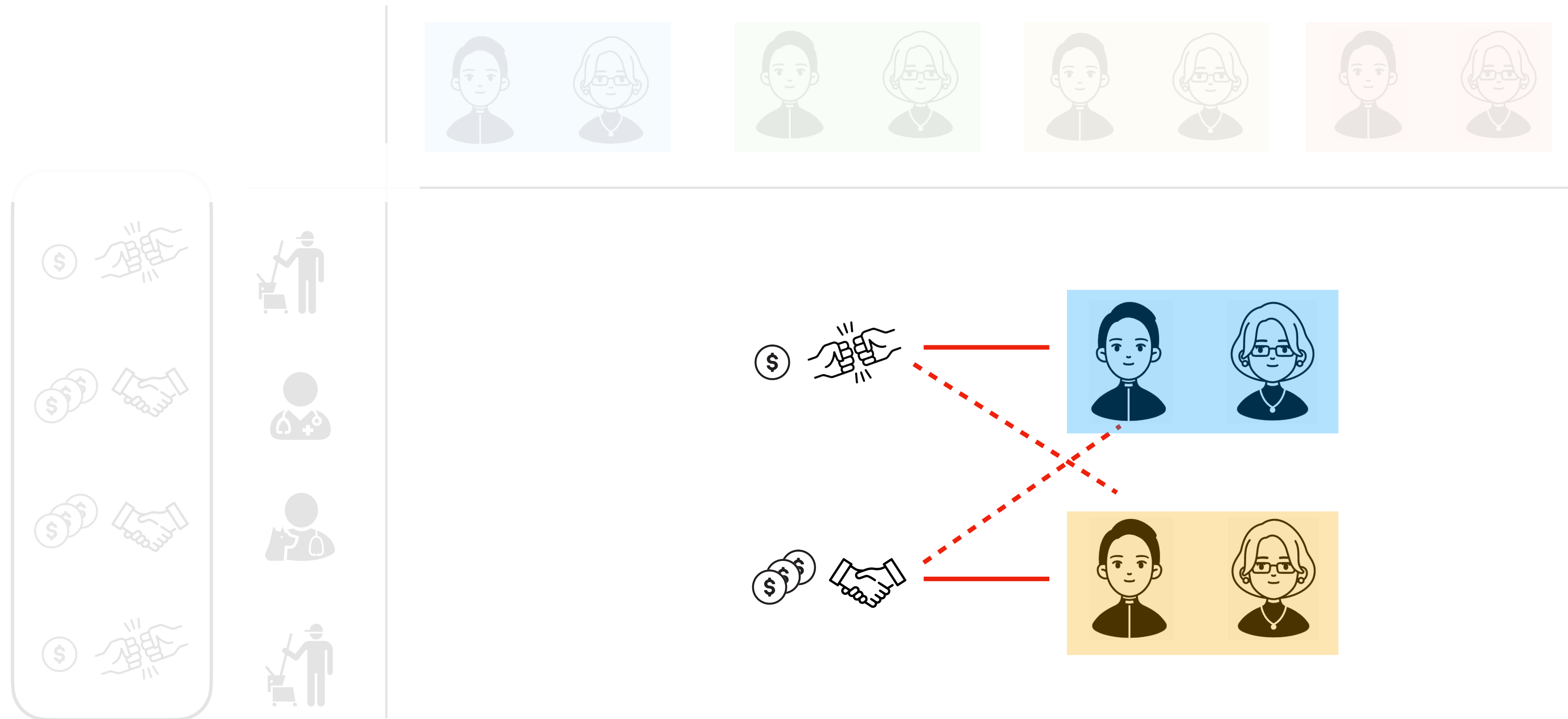
## Formalism: **Contextual Multi-Armed Bandit**



## Challenge Two: Multiple Dimensions



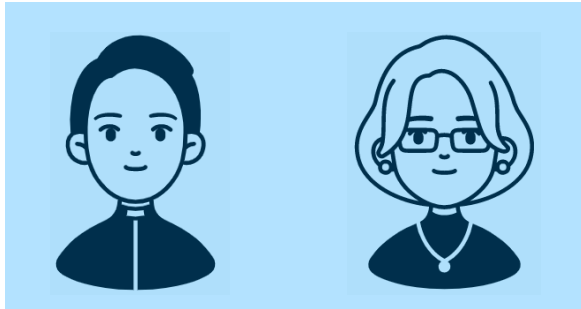

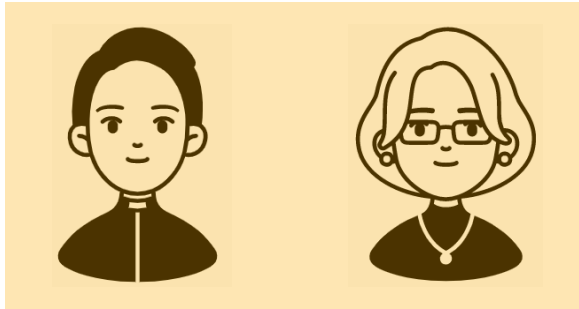
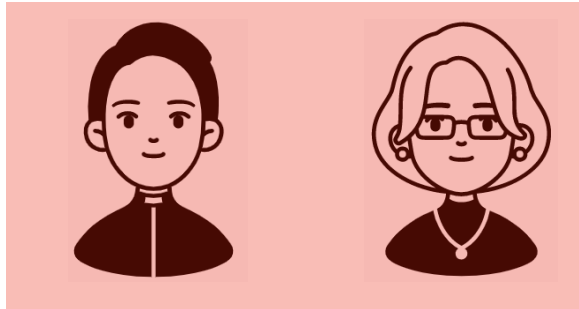
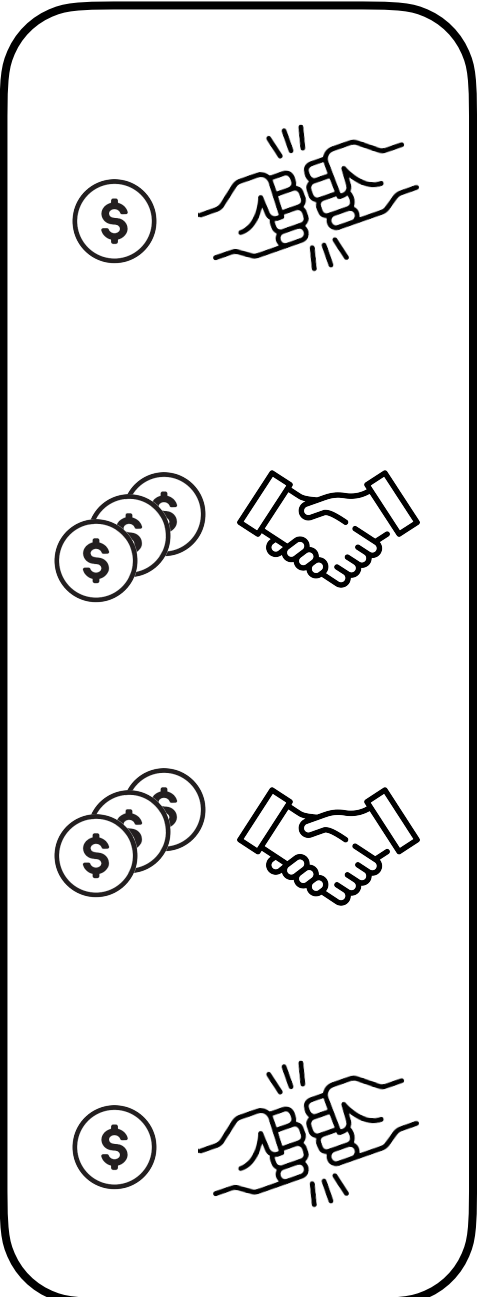




## Solution Two: Bayesian Logistic Regression



## **Formalism: Contextual Multi-Armed Bandit**

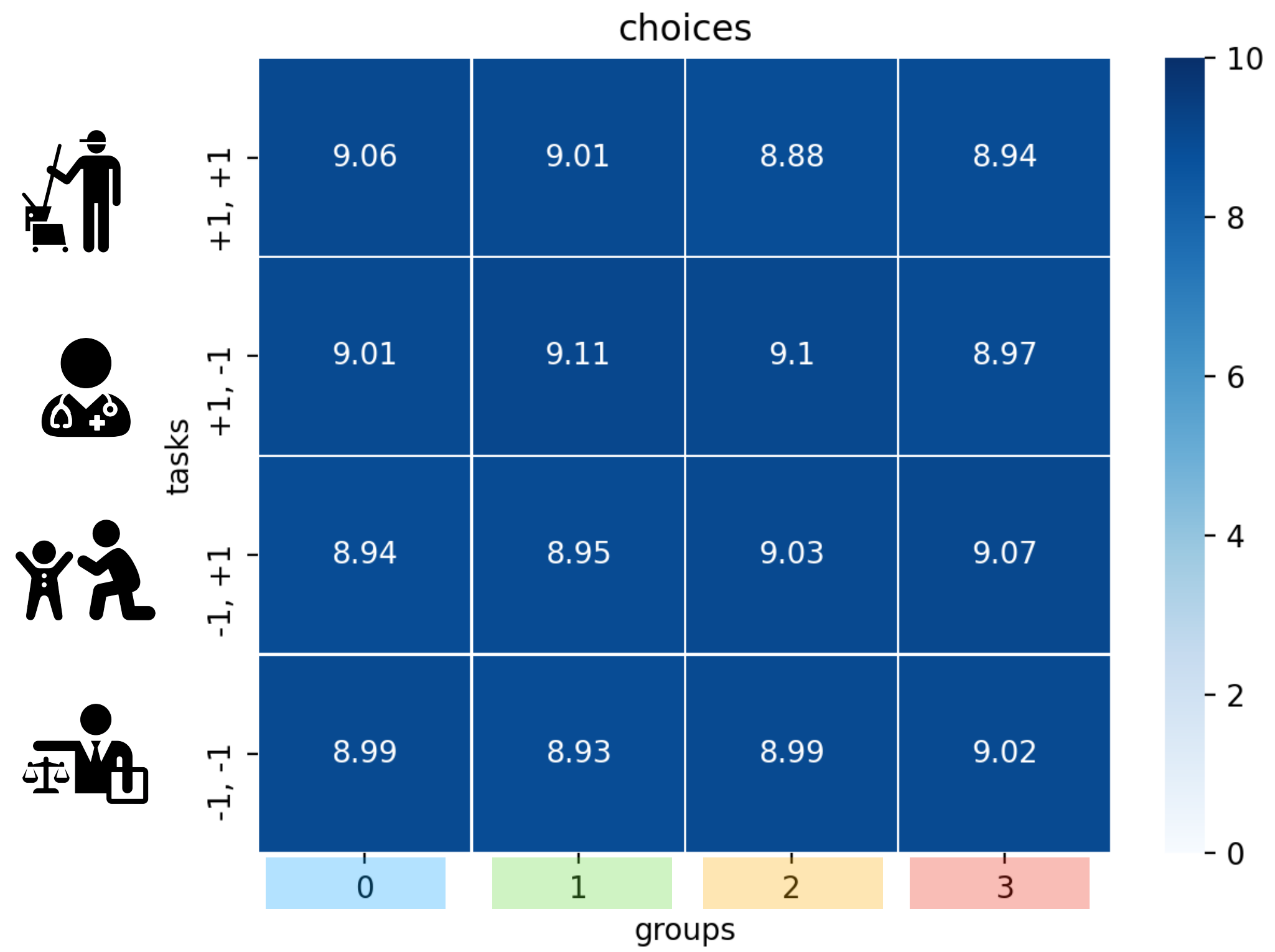
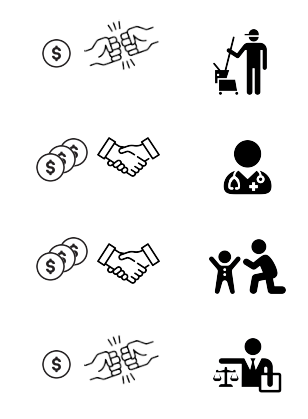
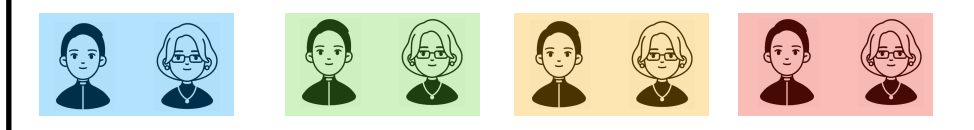
## **Simulation: Hiring as Contextual Bandit**

**Simulation:**

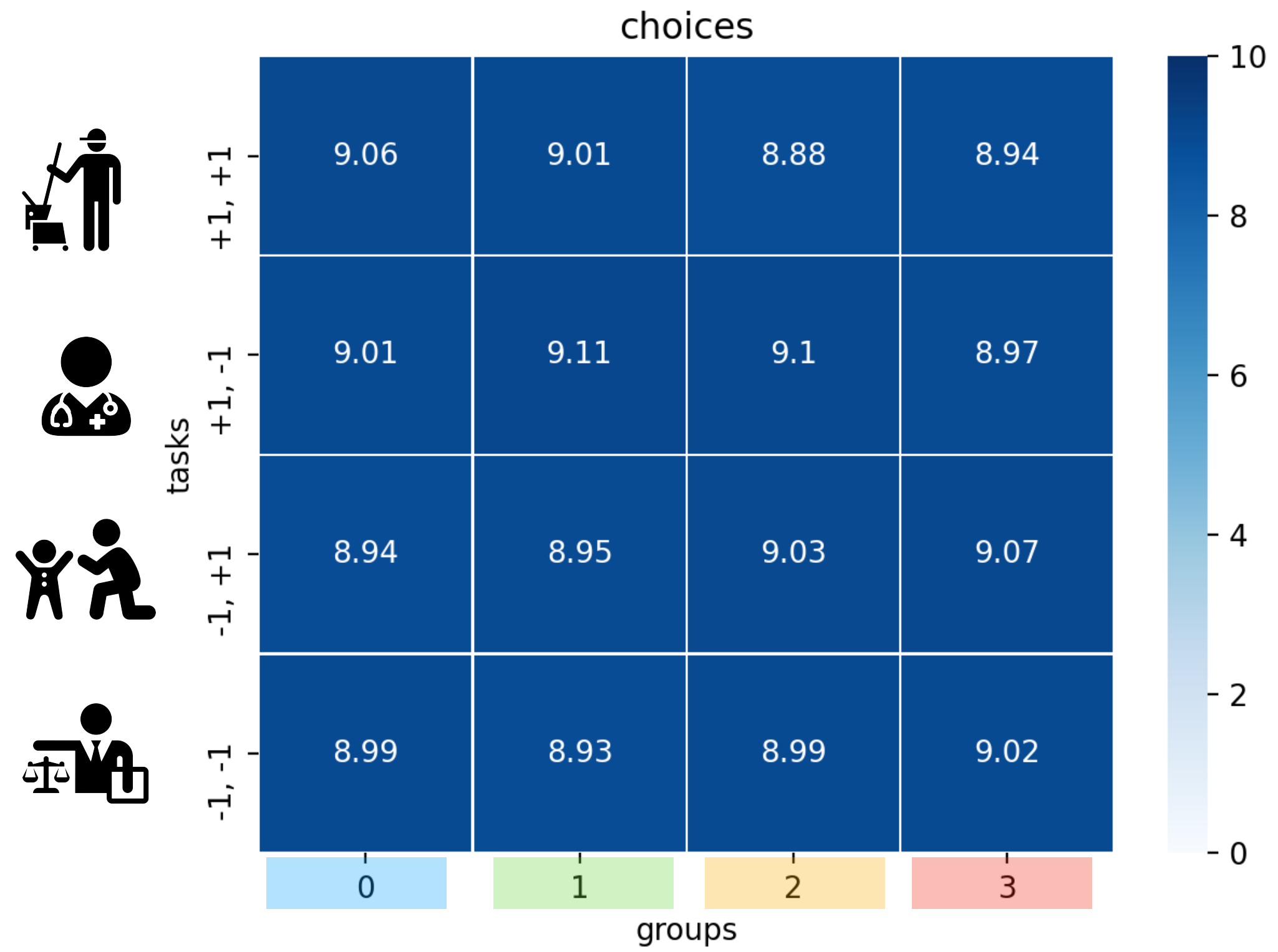
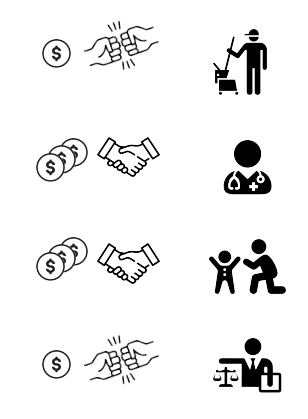
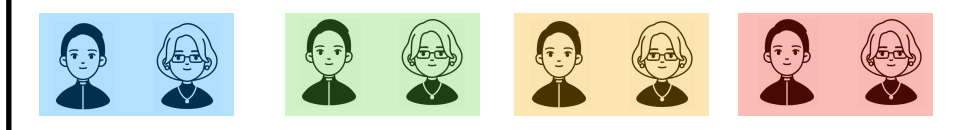
					
		0.9	0.9	0.9	0.9
		0.9	0.9	0.9	0.9
		0.9	0.9	0.9	0.9
		0.9	0.9	0.9	0.9

100 sims; 40 recs each;  $\sim N(0.9, 0.001)$

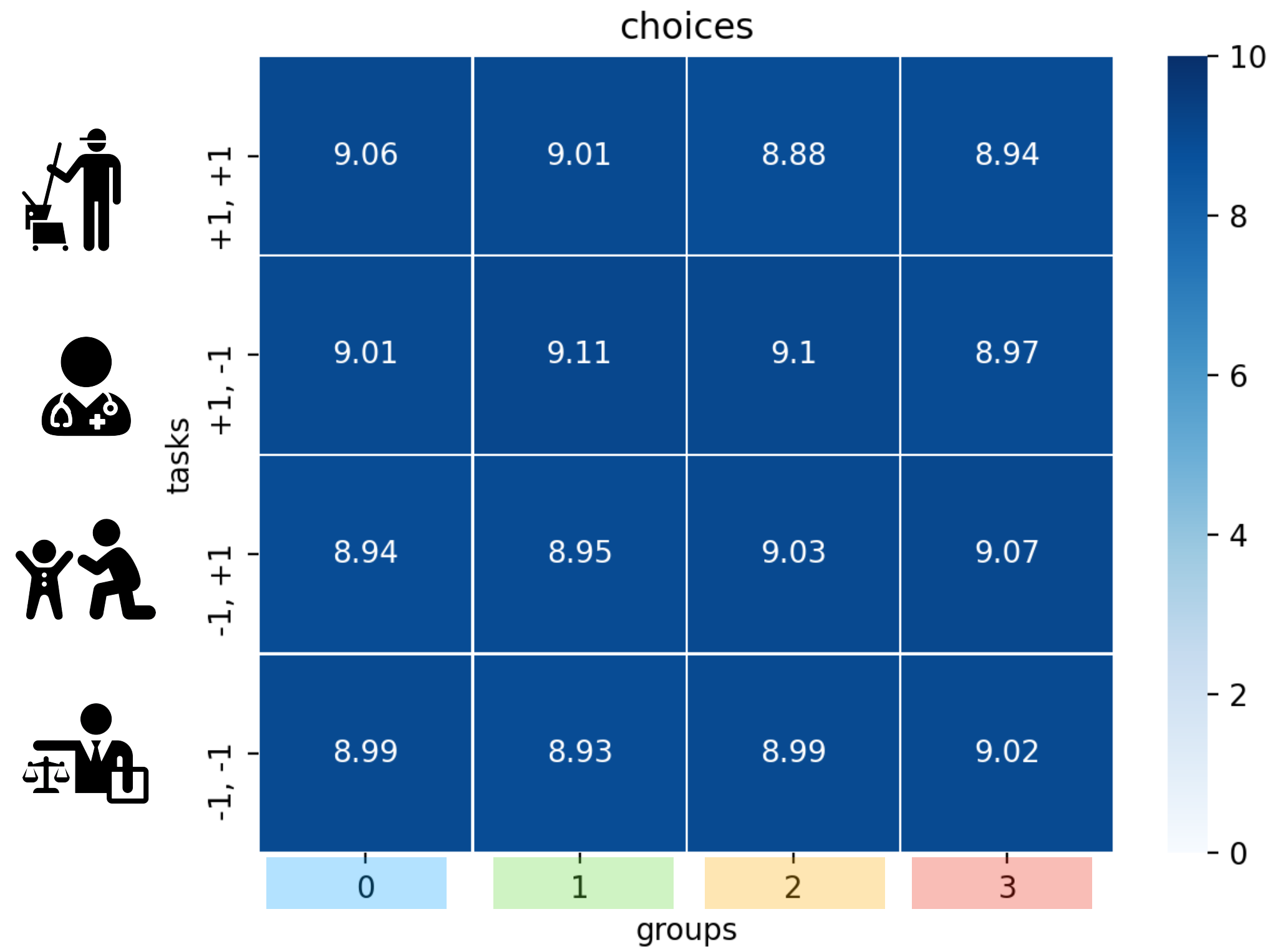
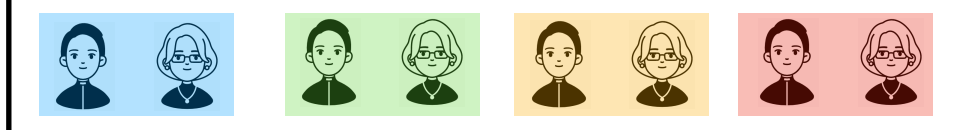
# Simulation: Ground truth



# Simulation: Random decisions

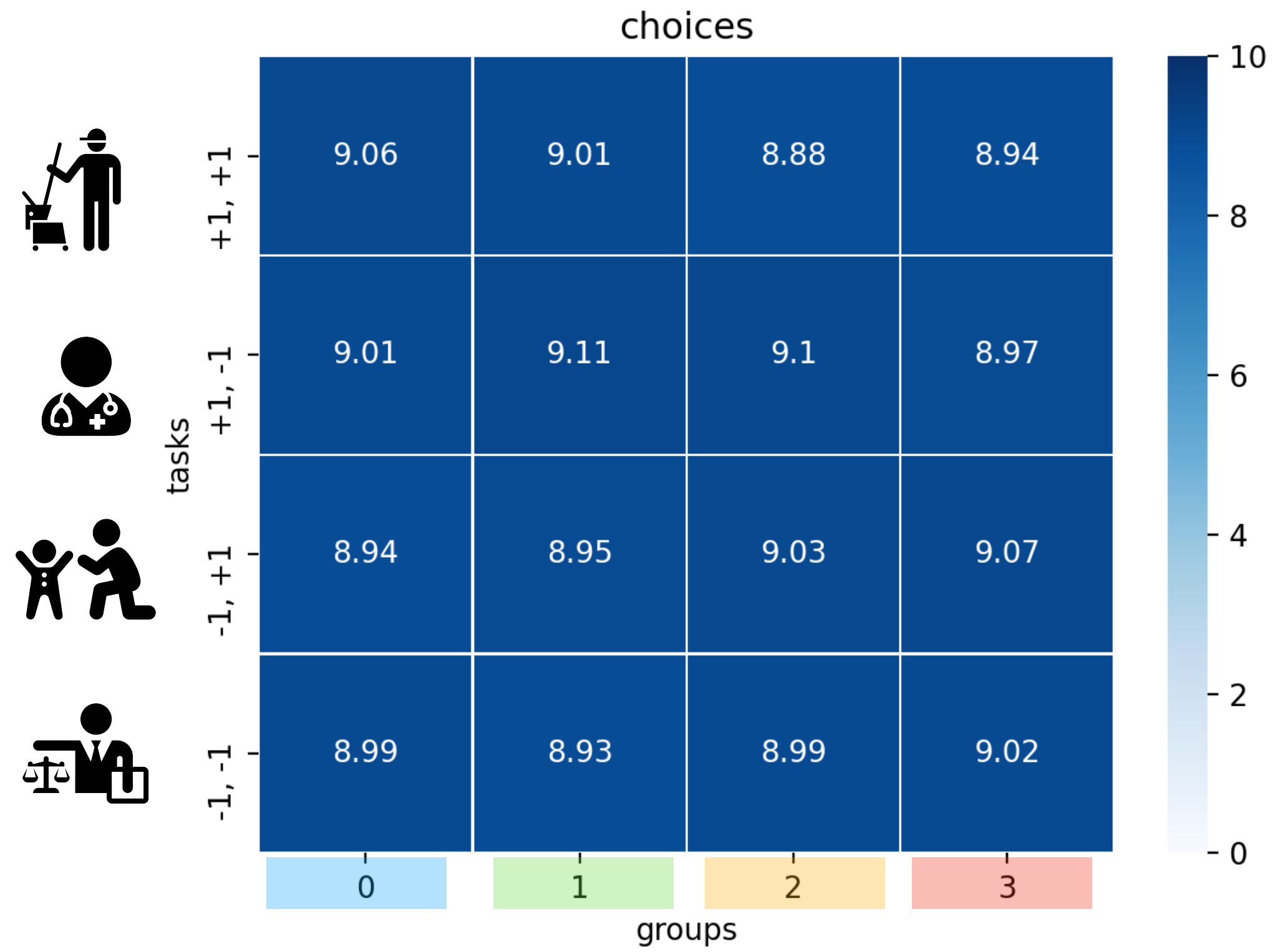
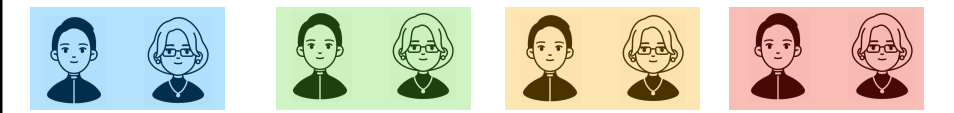


# Simulation: Random decisions

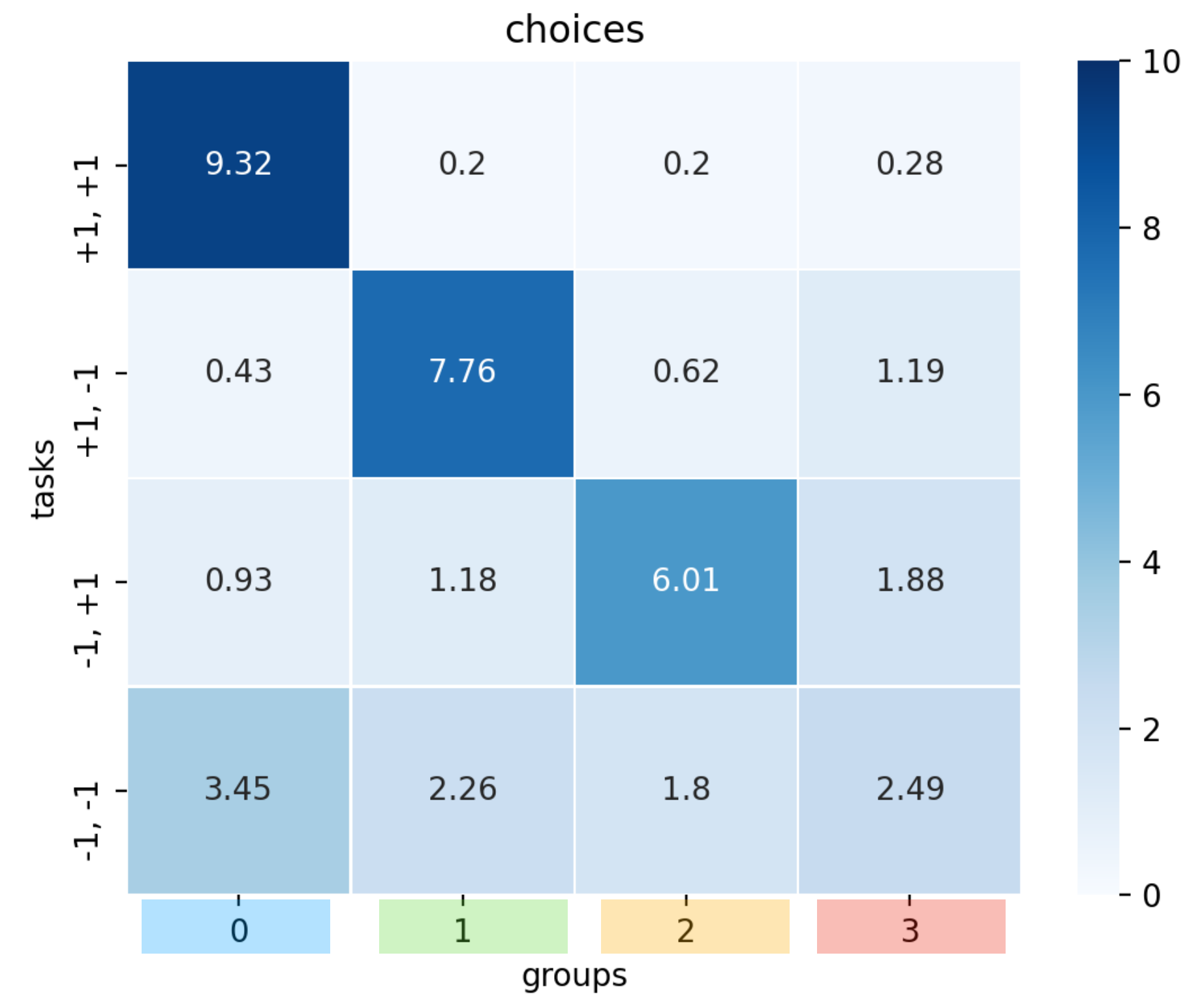
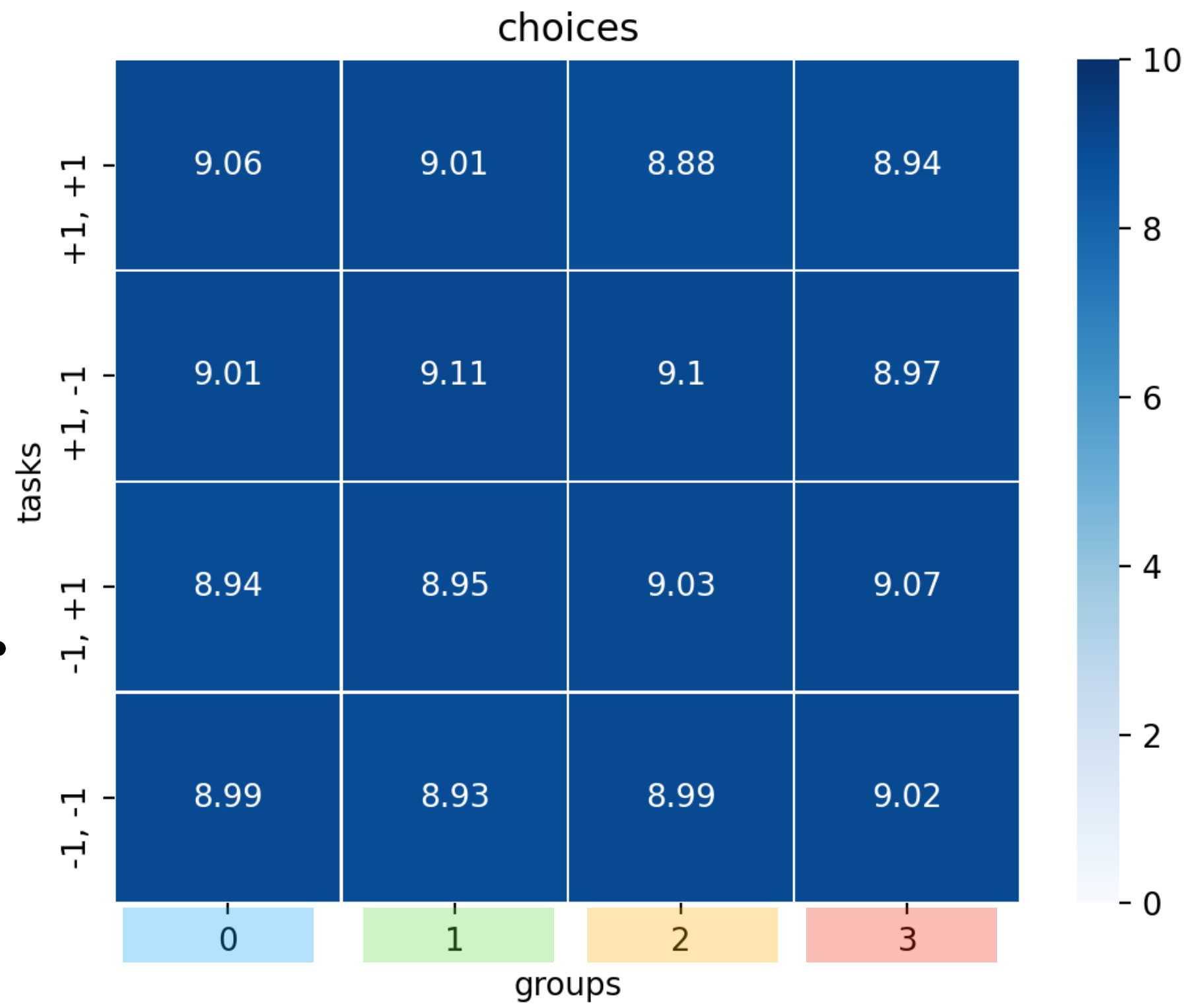
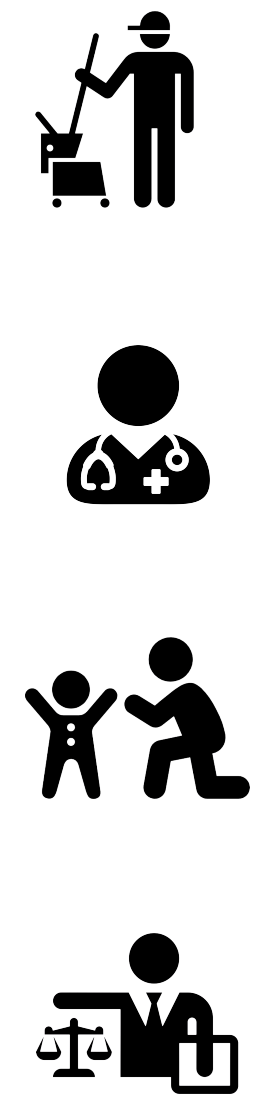
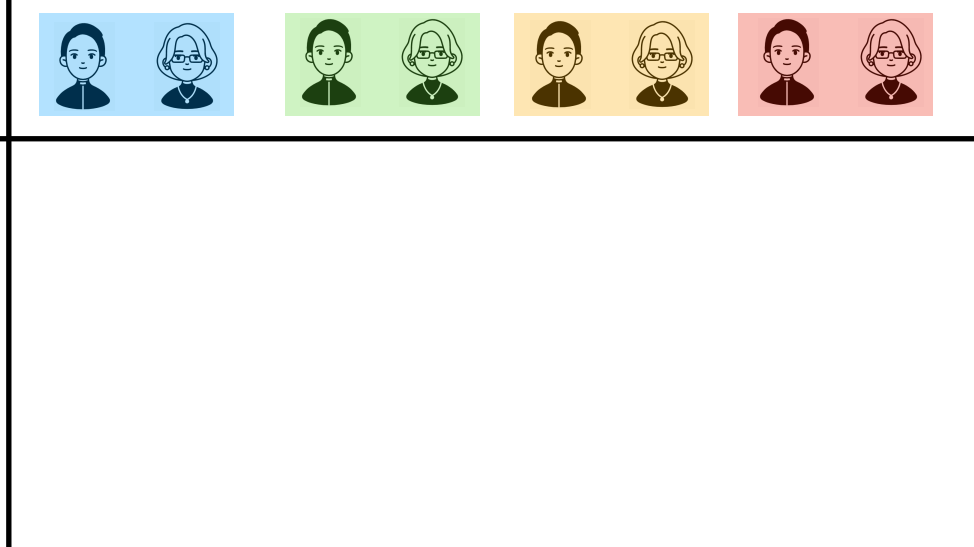




# Simulation: Adaptive decisions



# Simulation: Adaptive decisions



## Interim summary

Groups differ in what they do, from simulated agents making adaptive exploration:

## Interim summary

Groups differ in what they do, from simulated agents making adaptive exploration:

- How: Make new decisions based on past [REDACTED] experiences.
- Why: Utility-maximizing [REDACTED].
- Tradeoff: Early positive experiences discourage [REDACTED] exploration.

## Interim summary

Groups differ in what they do, from simulated agents making adaptive exploration:

- How: Make new decisions based on past **(selective)** experiences.
- Why: Utility-maximizing but **(not belief) maximizing**.
- Tradeoff: Early positive experiences discourage **(exhaustive)** exploration.

## **Experiments: Hiring Toma**

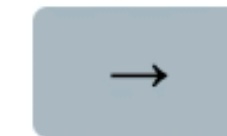
## Experiment: Hiring Consultant for Toma City (total $N=1300$ )



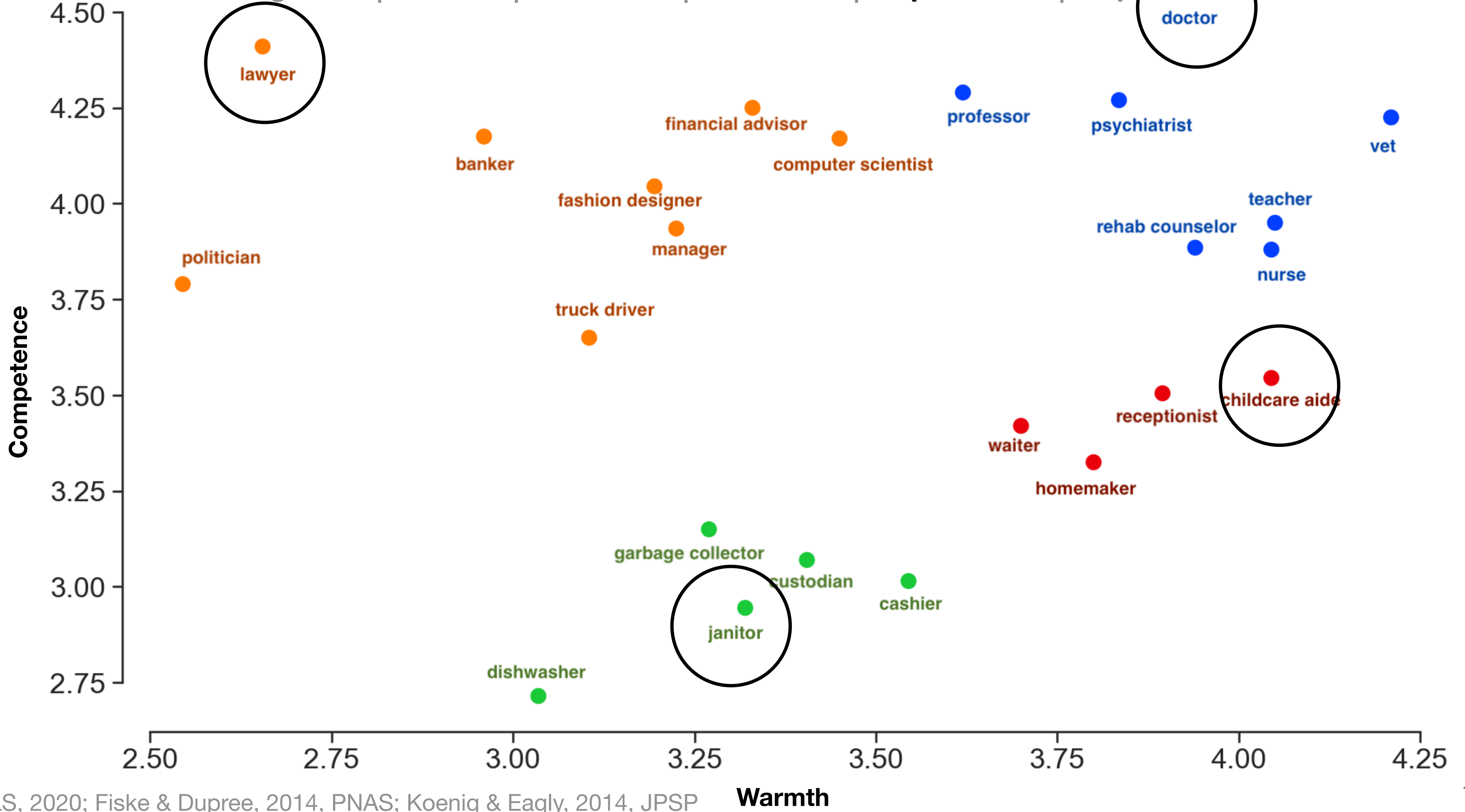
## Experiment: Hiring Consultant for Toma City (total $N=1300$ )

Ready to help the Mayor?

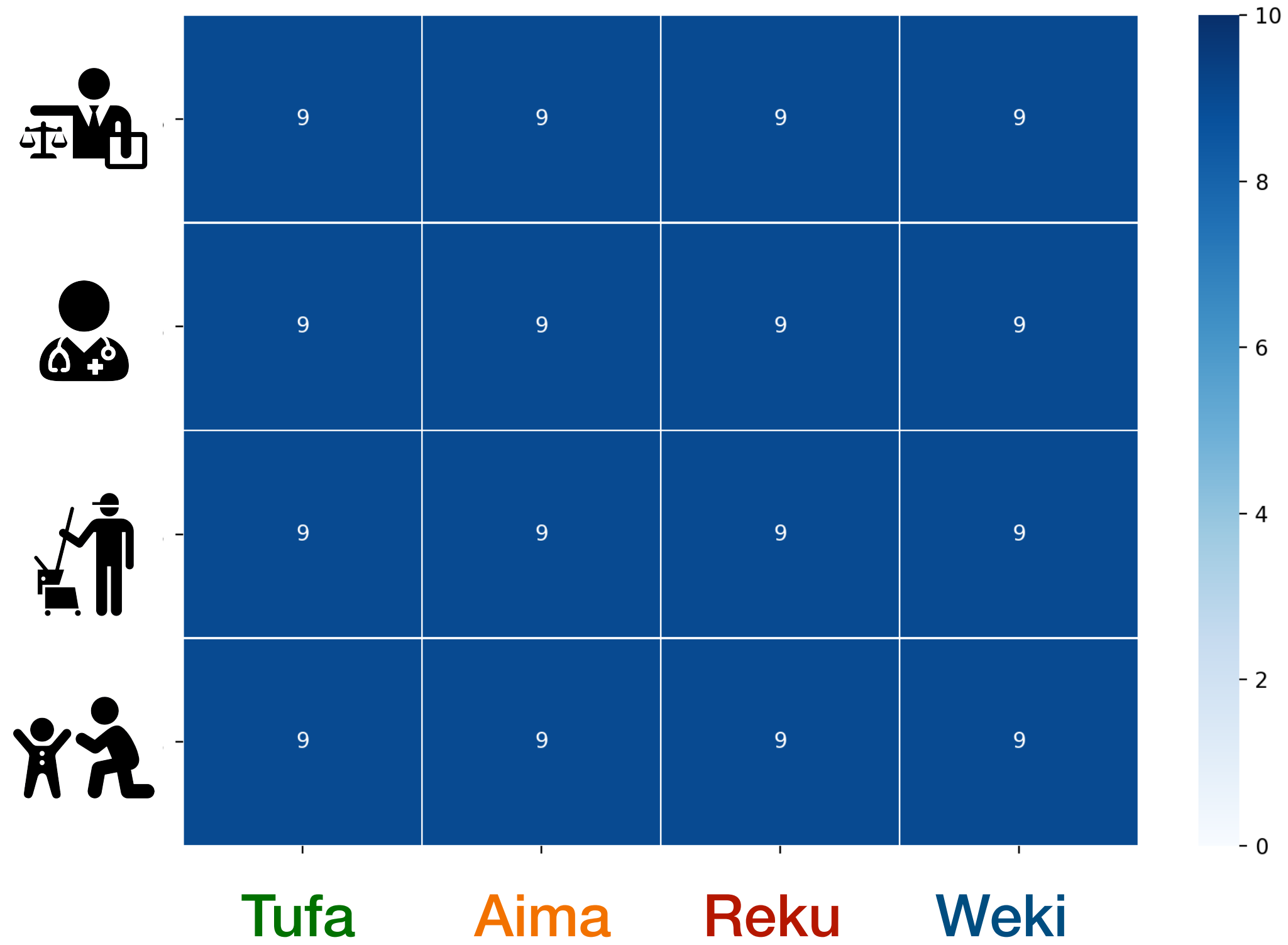
Let's get started!



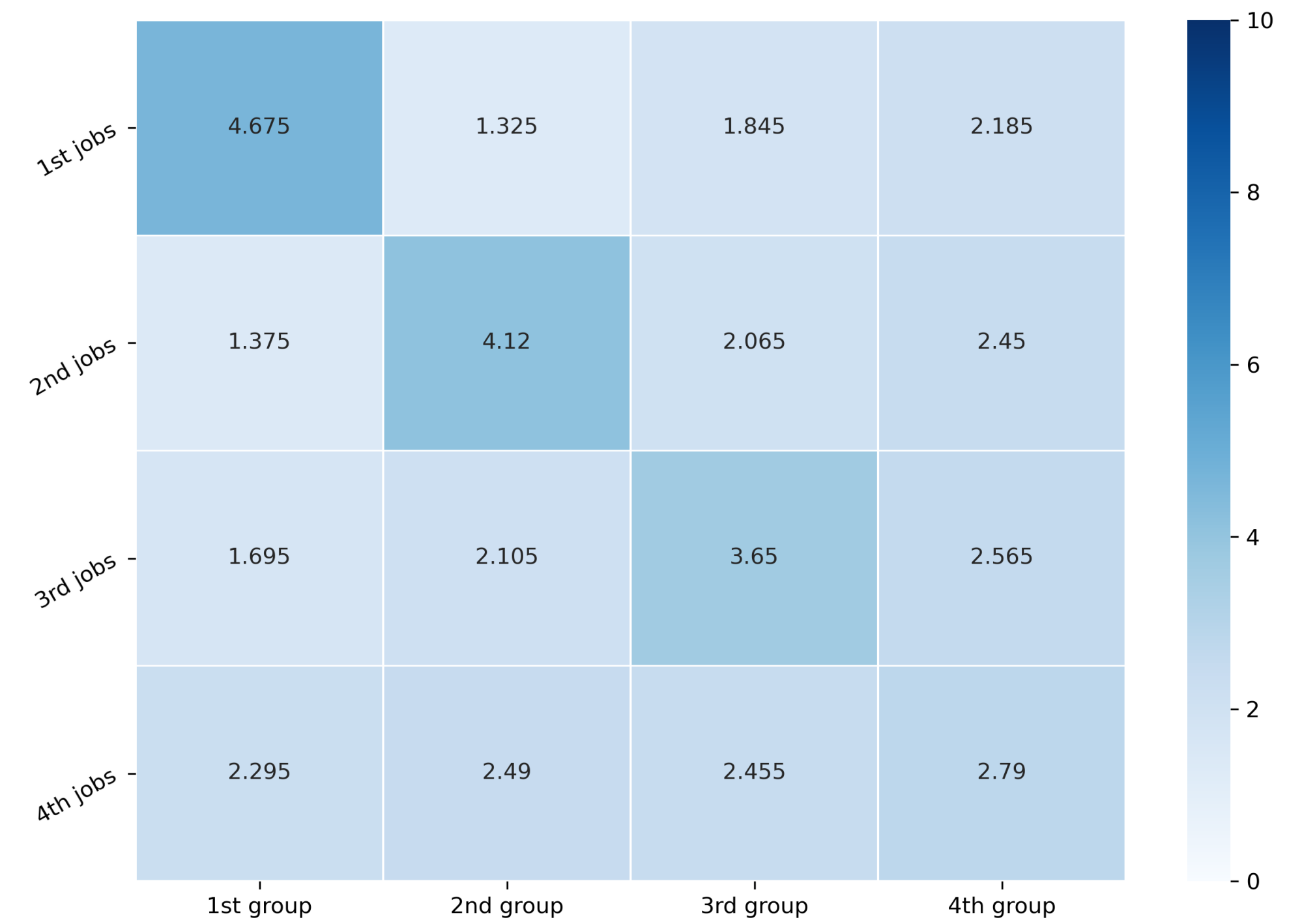
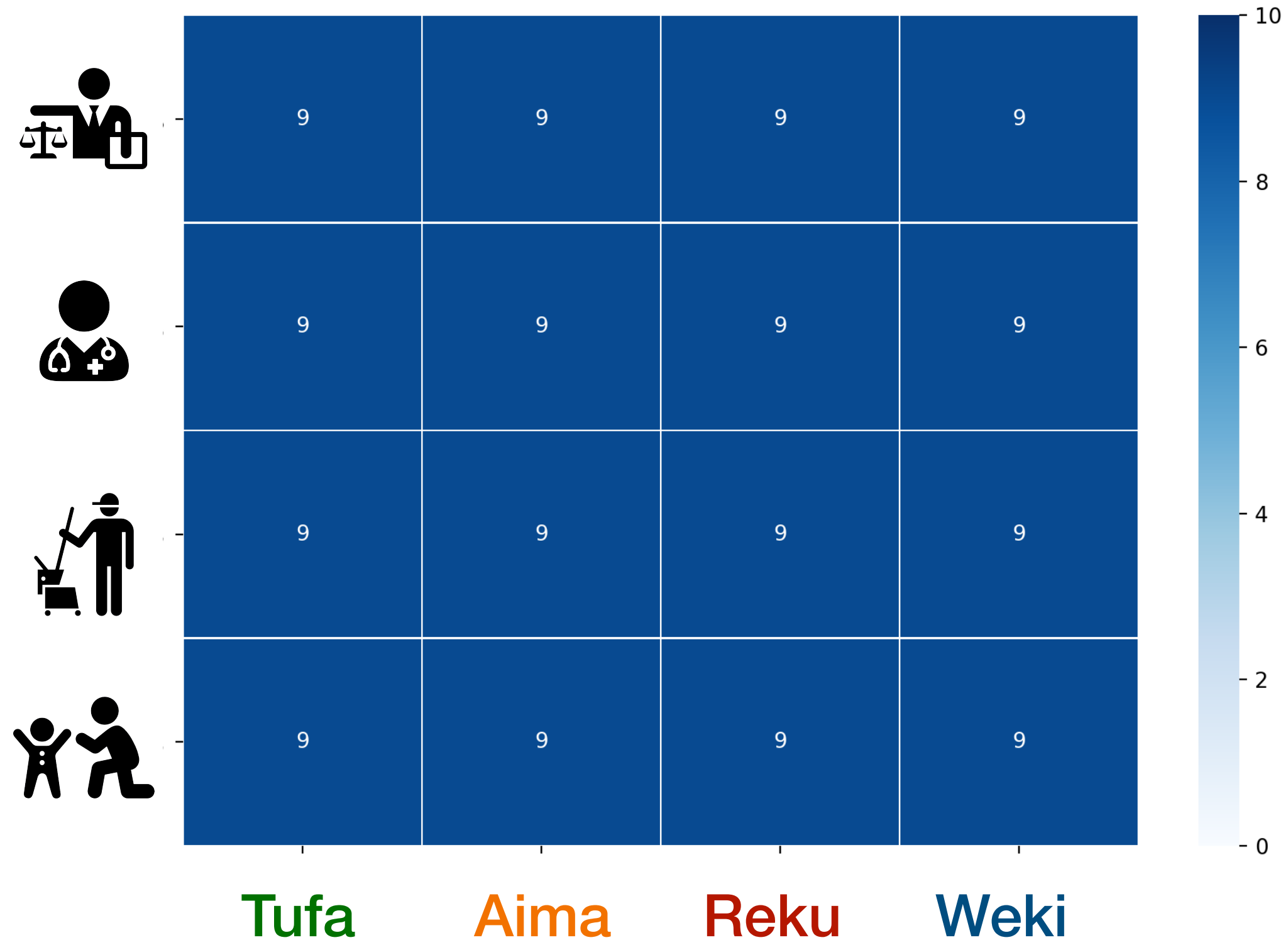




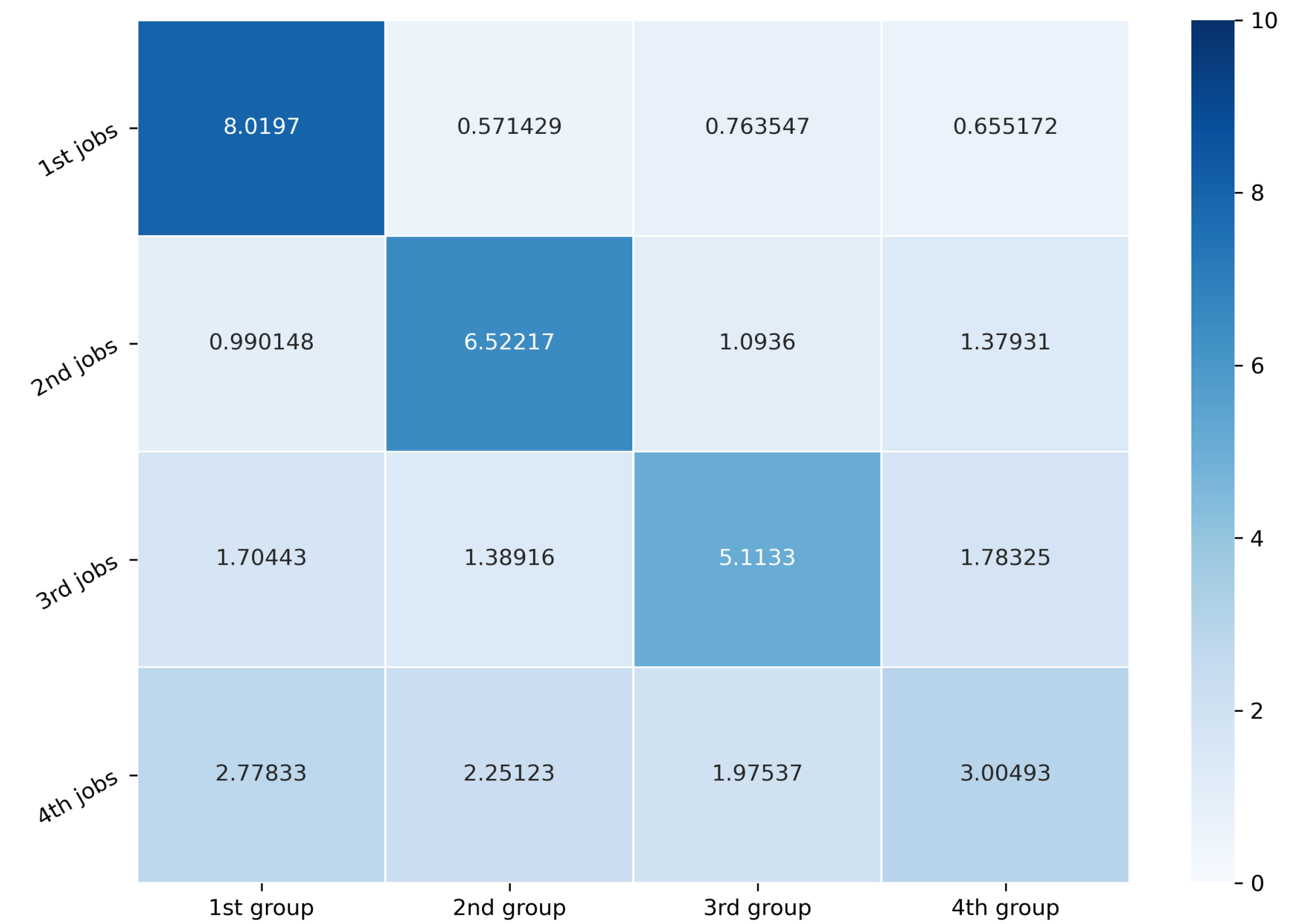
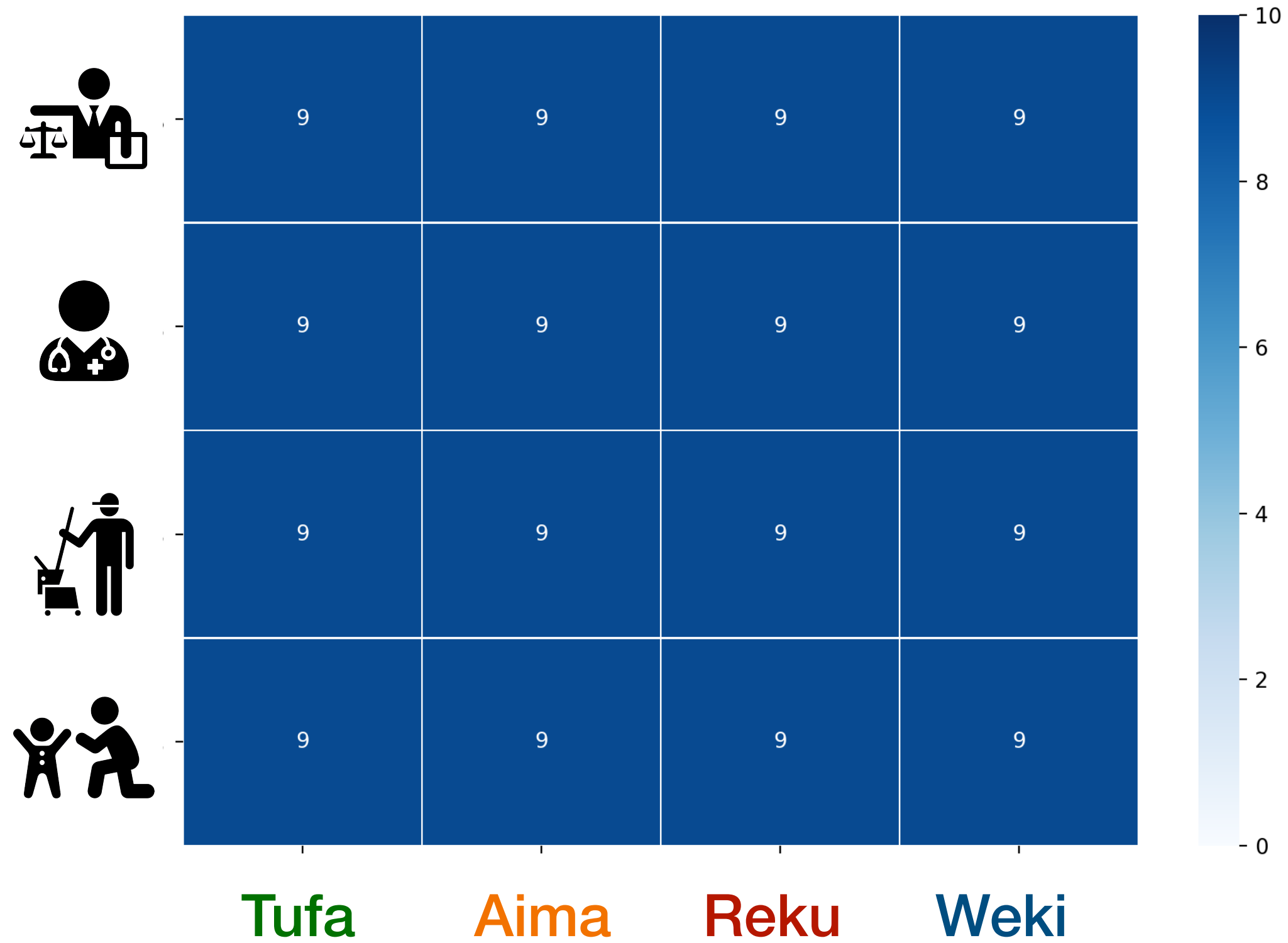
## Experiment: Ground truth



## Experiment: Participant choices in random decisions

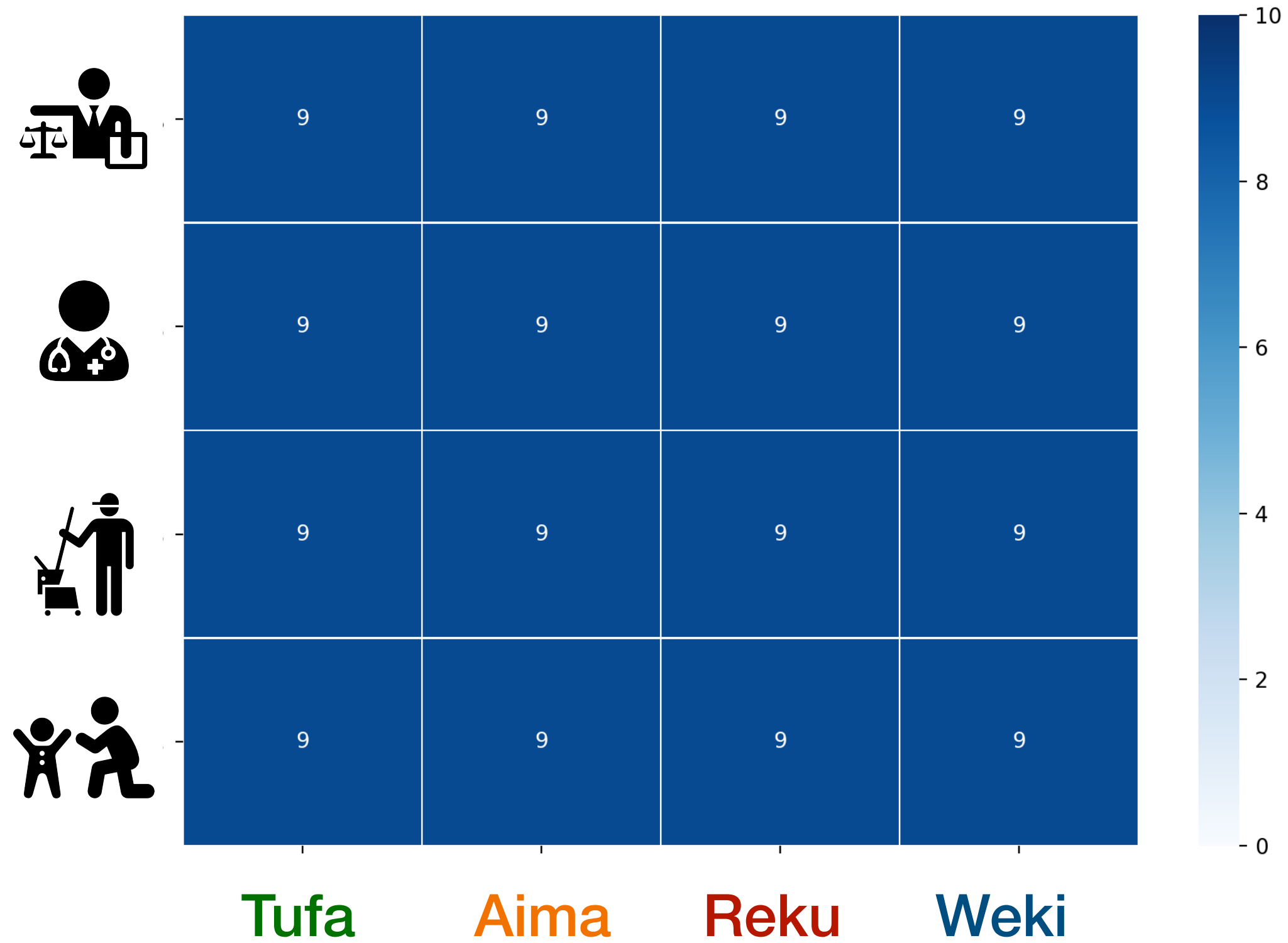


## Experiment: Participant choices in adaptive decisions



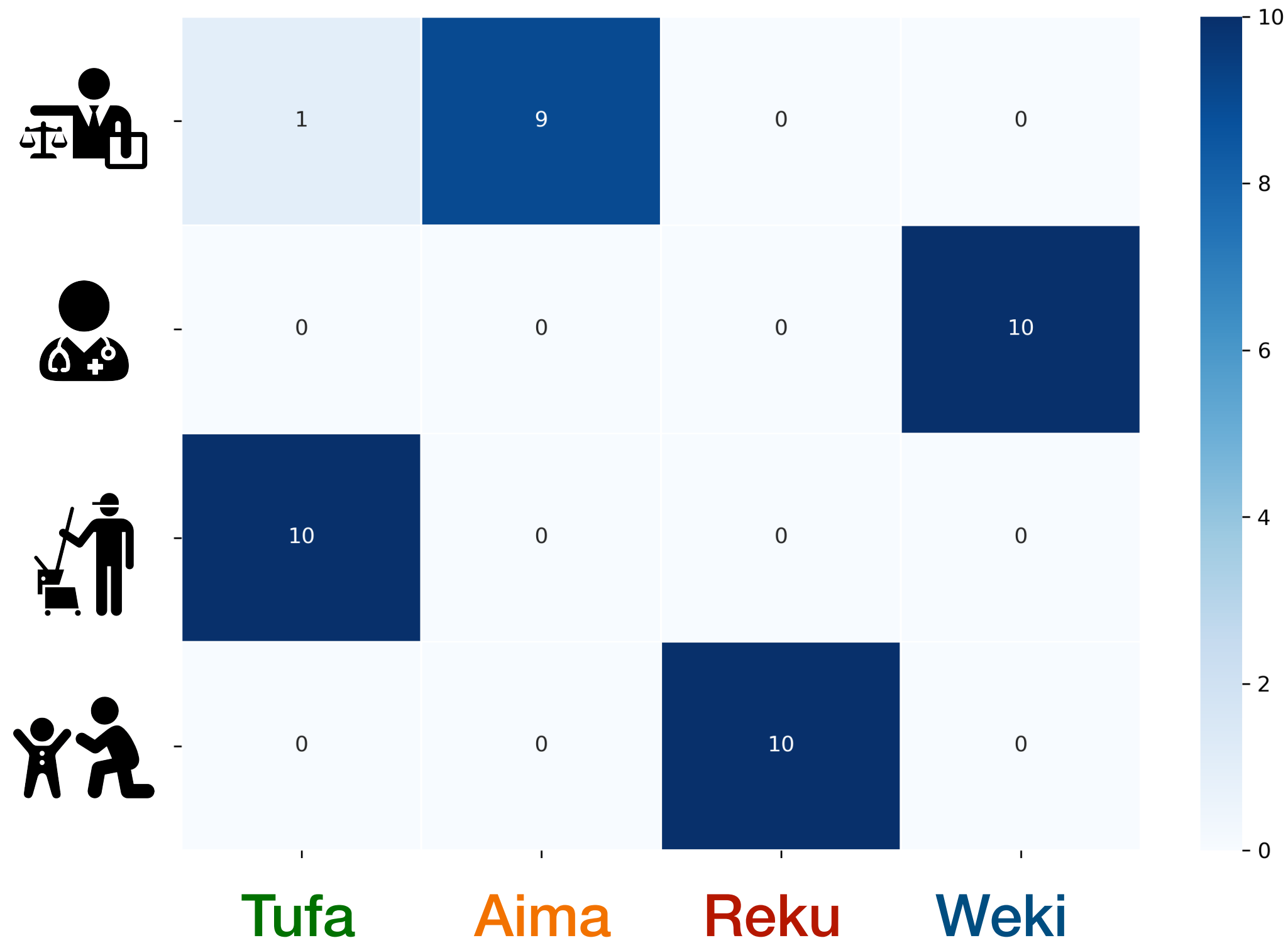
## One concrete example: Subject # 54

True Toma City



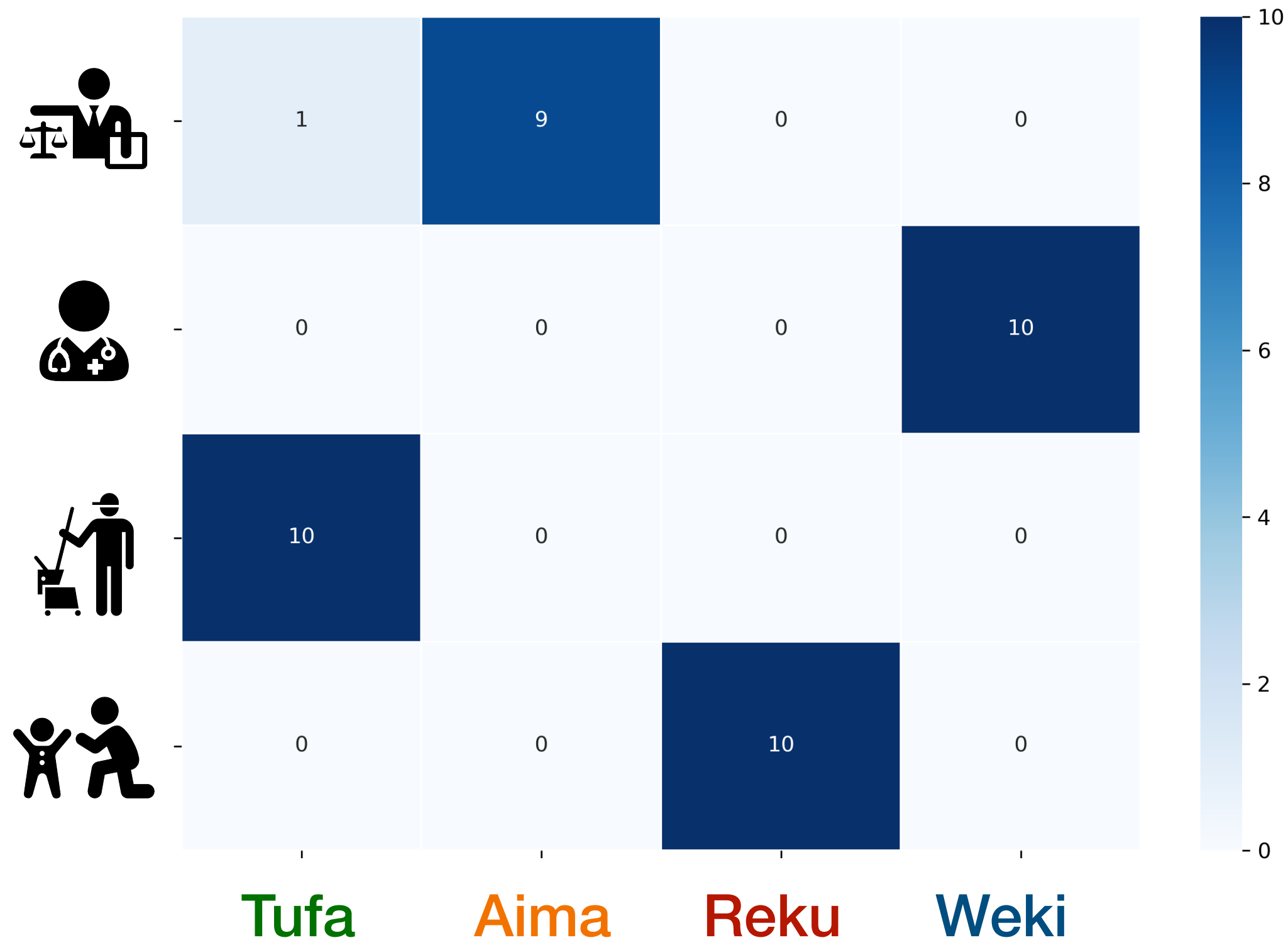
## One concrete example: Subject # 54

Hiring choices

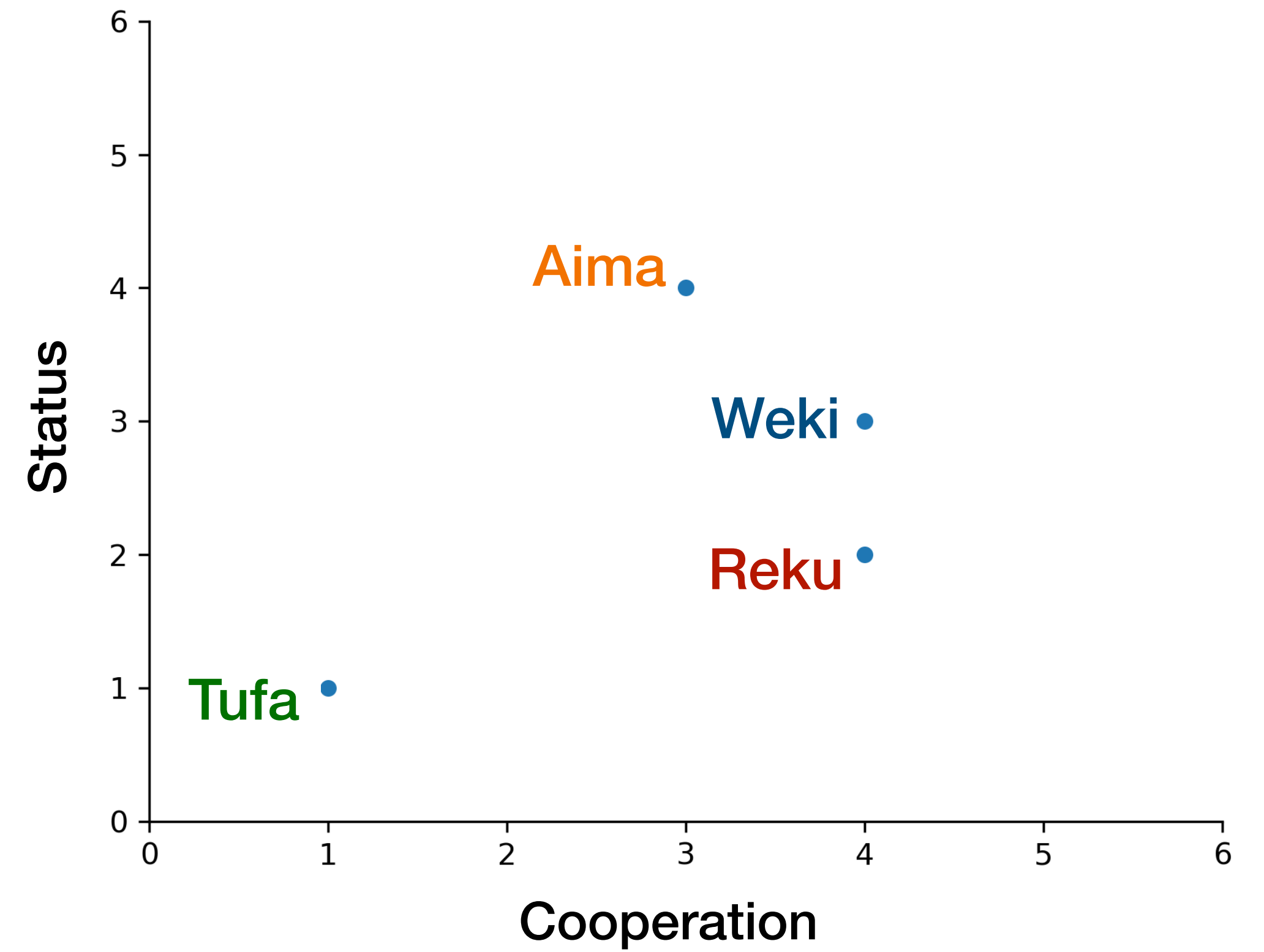


## One concrete example: Subject # 54

### Hiring choices

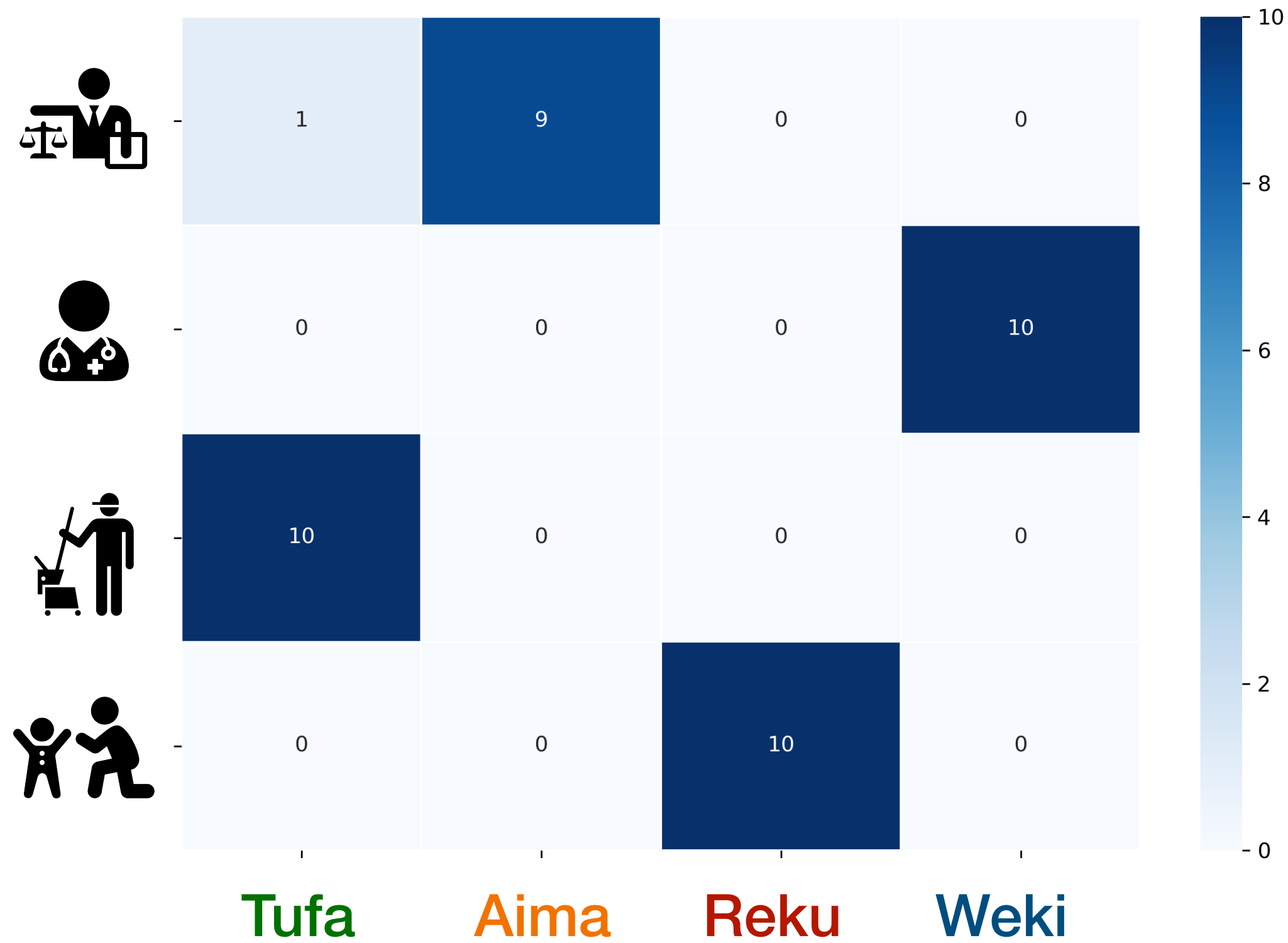


### Social positions

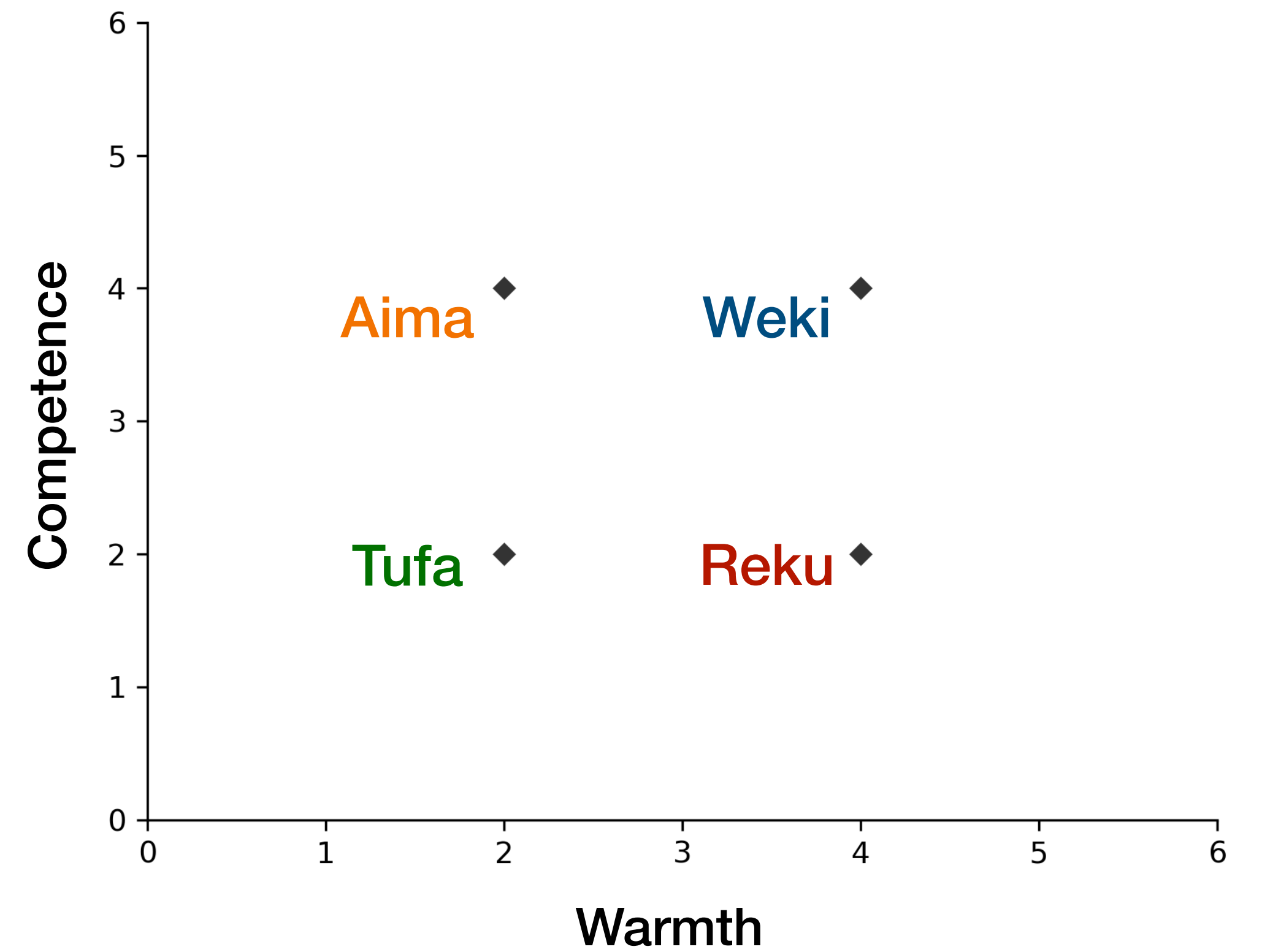


## One concrete example: Subject # 54

### Hiring choices



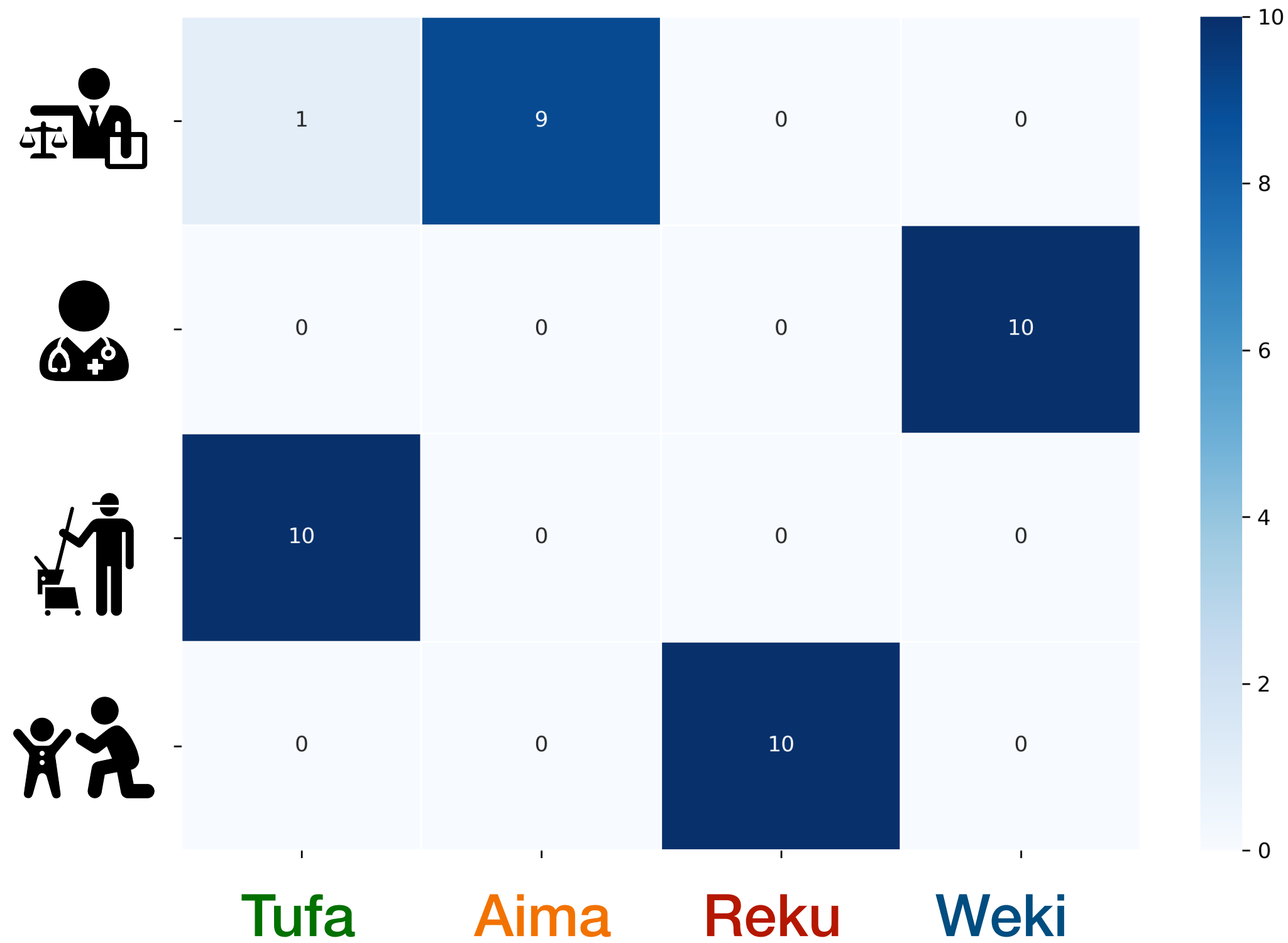
### Social stereotypes



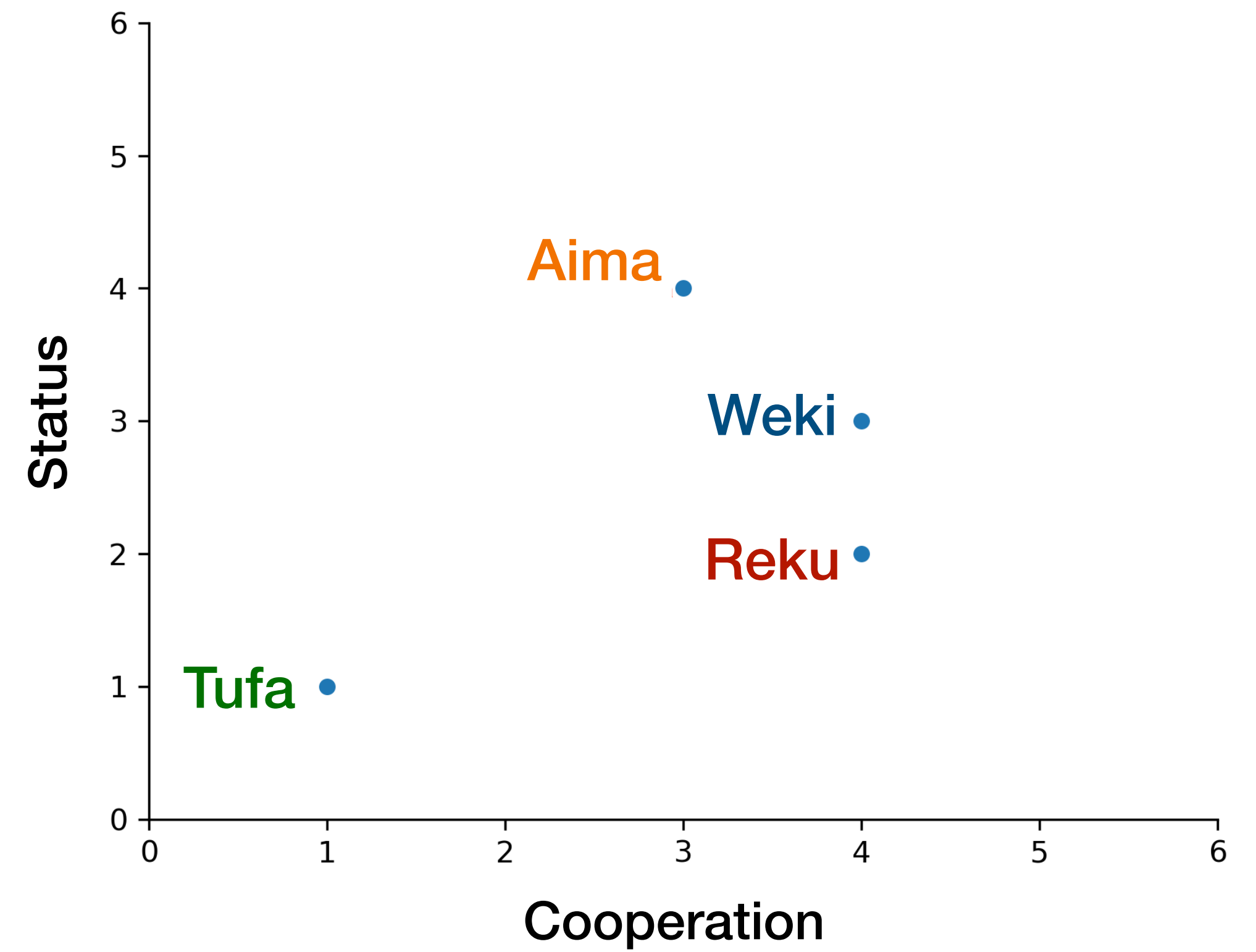


## One concrete example: Subject # 54

### Hiring choices



### Social positions



## One concrete example: Subject # 54

Arbitrary initials

“At first it was just trial and error and

Adaptive decisions

when I saw which received points I stuck with those...I had a feeling that if I recommended any of the people for any particular job, I would earn a bonus. I didn't want to chance that hypothesis being incorrect, so I stuck to the recommendations that had proven to earn a bonus on previous rounds...

the desire to earn more of a bonus affects the choices we make.”

Self-interest maximization

This talk:  
One individual

**Future: Consensus  
across individuals**

## **Our Proposal: Adaptive Exploration**

 Globally Accurate

 Morally Right

 Adaptive to self-  
interested  
decision-makers

 Detrimental to  
collective society

## **Implications: Adaptive Choices vs. Collateral Damage**

## **Groups stand in different positions and have distinct stereotypes. Why?**

- **We aim to offer one sufficient explanation.**
  - Without intentional oppressors or cognitively constrained decision makers;
  - Without group size differences or innate group differences.

## **Groups stand in different positions and have distinct stereotypes. Why?**

- We aim to offer one sufficient explanation.
  - Without intentional oppressors or cognitively constrained decision makers;
  - Without group size differences or innate group differences.
  
- **Because people learn & decide based on the (selective) past.**  
Initiated by historical events, cascading from adaptive choices, driven by self-interest.

## **Groups stand in different positions and have distinct stereotypes. Why?**

- We aim to offer one sufficient explanation.
  - Without intentional oppressors or cognitively constrained decision makers;
  - Without group size differences or innate group differences.
- Because people learn & decide based on the (selective) past.

Initiated by historical events, cascading from adaptive choices, driven by self-interest.
- **Individually adaptive, therefore, collectively alarming.**

## Real-world examples: Adaptive decisions & Collateral damage

*RESEARCH*

---

**SOCIAL SCIENCE**

**Improving refugee integration  
through data-driven  
algorithmic assignment**



## Real-world examples: Adaptive decisions & Collateral damage

*RESEARCH*

---

SOCIAL SCIENCE

Improving refugee integration  
through data-driven  
algorithmic assignment



facebook

NETFLIX

Google

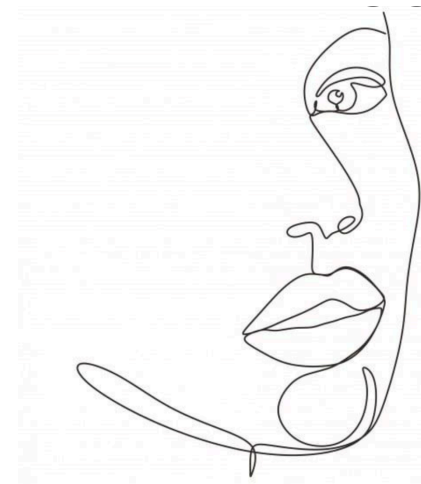
amazon

**Are we doomed because of adaptive exploration?**

**Design a system to ameliorate negative consequences from adaptive exploration.**

Are we doomed because of adaptive exploration?

## In hiring practices:



### Motivational biases

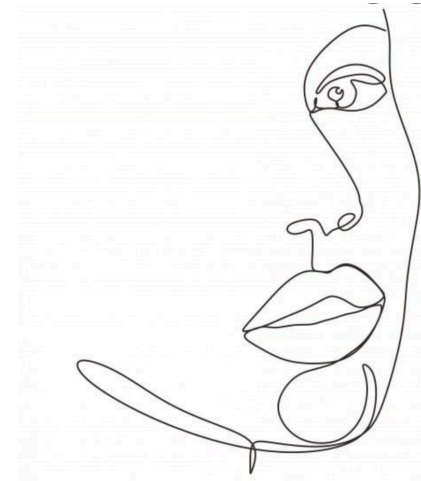
- Identity
- Dominance

### Cognitive biases

- Limited memory
- Selective attention

e.g., de-bias training

**In hiring practices:**



**Sample biases**

- Unequal group size

**Group differences**

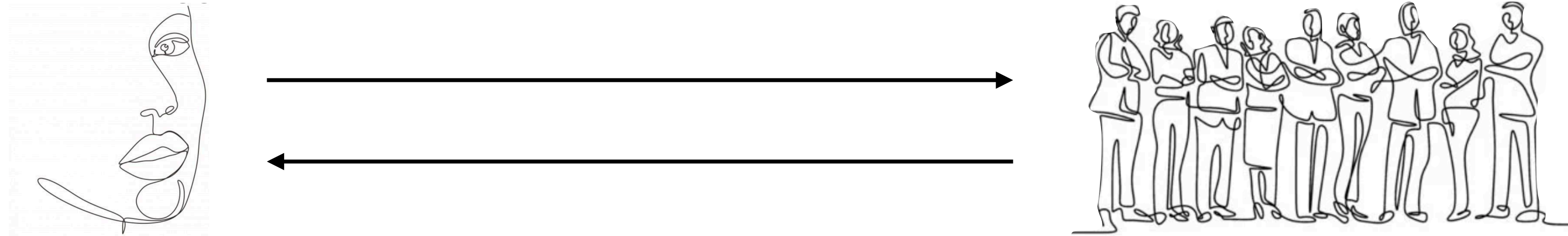
- Gender

**e.g., diversity recruitment**

**In hiring practices:**



**In hiring practices:**



**Hypothesis**

Design a system that encourages continuous exploration.

**Ideas**

Add exploration bonus? Add equity goal?

# The Psychology of How We **Make Sense of** the Social World



# The Psychology of How We **Make Sense** of the Social World

**Social Psychology**  
The “problematic” human

**Cognitive Science**  
The “intelligent” human

**Where do stereotypes come from?**  
Adaptive exploration

**Social Policy**  
Diversity and inclusion

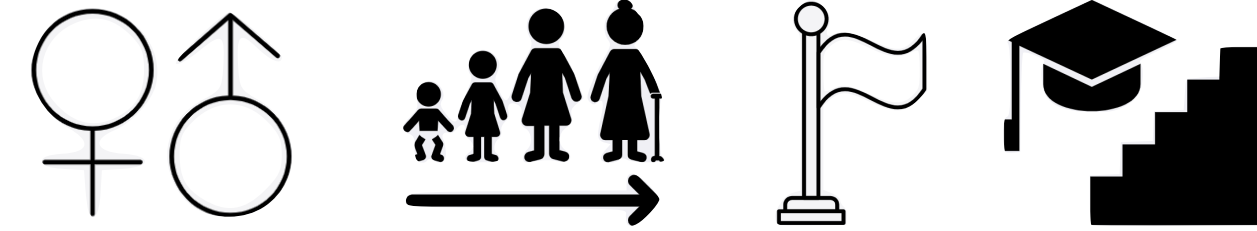
**Machine Learning**  
Fairness in A.I.

# The Psychology of How We **Make Sense of** the Social World



# The Psychology of How We **Make Sense of** the Social World

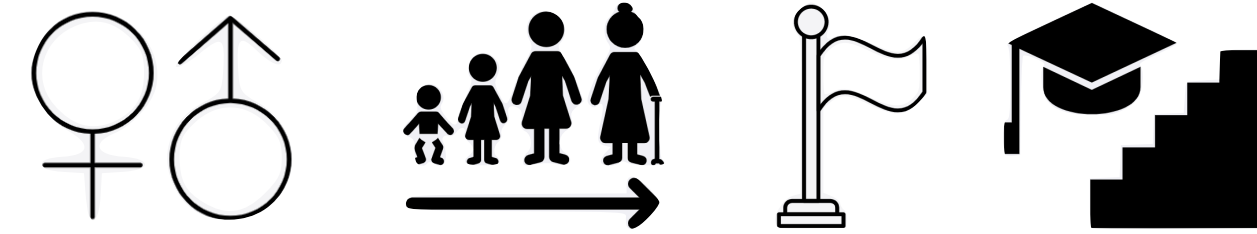
1. Which dimension(s) to use to make decisions?



**An Origin Story**

# The Psychology of How We **Make Sense of** the Social World

1. Which dimension(s) to use to make decisions?



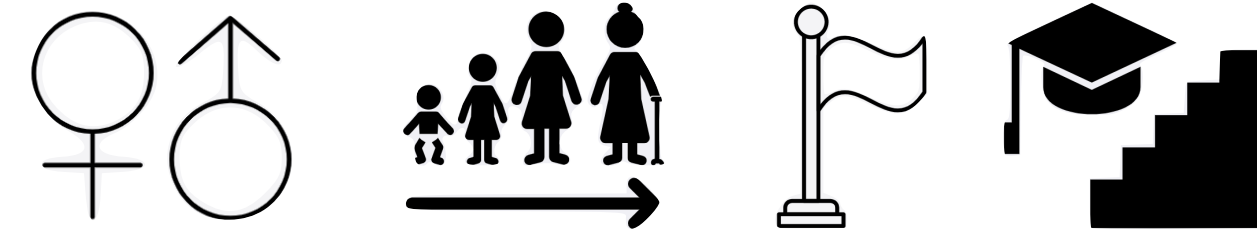
2. How about continuous and gestalt dimension(s)?



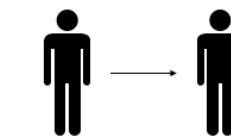
**An Origin Story**

# The Psychology of How We **Make Sense of** the Social World

1. Which dimension(s) to use to make decisions?



2. How about continuous and gestalt dimension(s)?

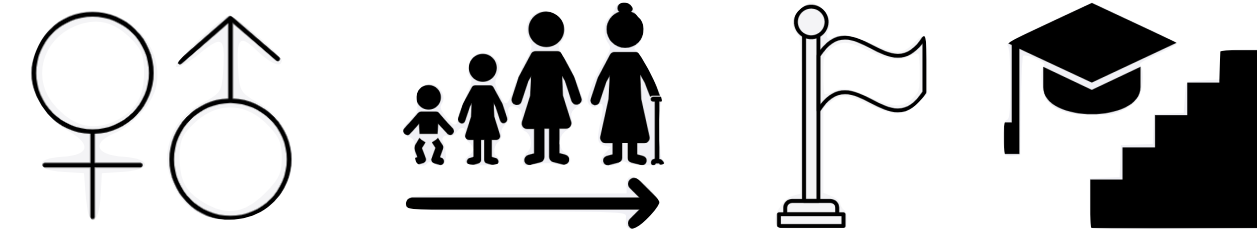


3. How do stereotypes transmit across individuals and generations?

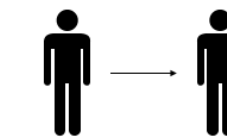
## An Origin Story

# The Psychology of How We **Make Sense of** the Social World

1. Which dimension(s) to use to make decisions?



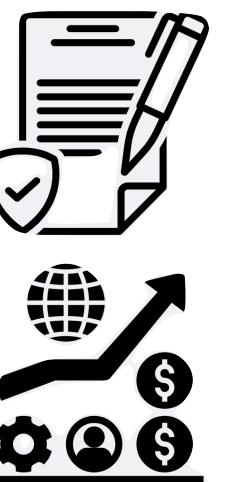
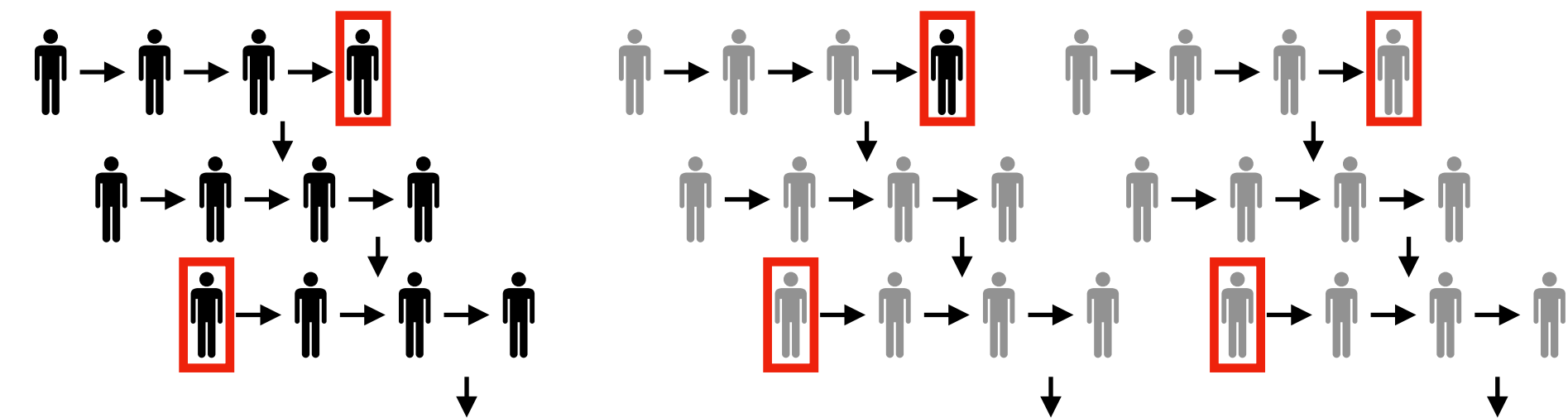
2. How about continuous and gestalt dimension(s)?



3. How do stereotypes transmit across individuals and generations?

4. How does adaptive exploration interact with other mechanisms?

**An Origin Story**



# The Psychology of How We **Make Sense of** the Social World

**Social Psychology**  
The “problematic” human

**Cognitive Science**  
The “intelligent” human

**Where do stereotypes come from?**  
Adaptive exploration

**Social Policy**  
Diversity and inclusion

**Machine Learning**  
Fairness in A.I.

# The Psychology of How We **Make Sense of** the Social World

1. Ideological legacy and stereotypes about the working class



**Leveraging the context**

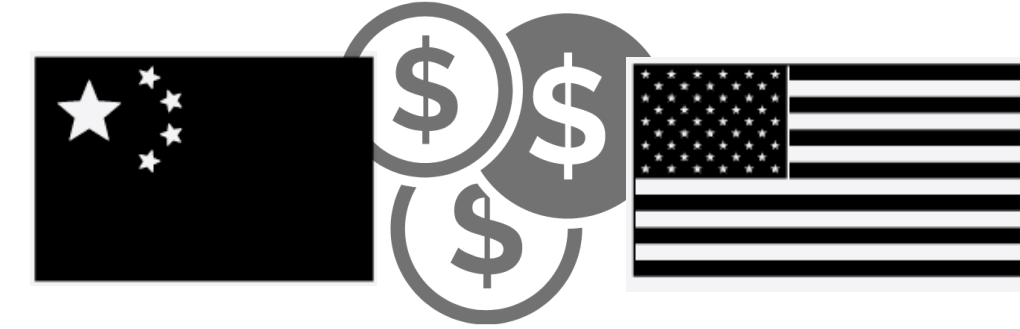


# The Psychology of How We **Make Sense of** the Social World

1. Ideological legacy and stereotypes about the working class



2. Cultural narratives and stereotypes about the upper class



**Leveraging the context**

# The Psychology of How We **Make Sense of** the Social World

1. Ideological legacy and stereotypes about the working class



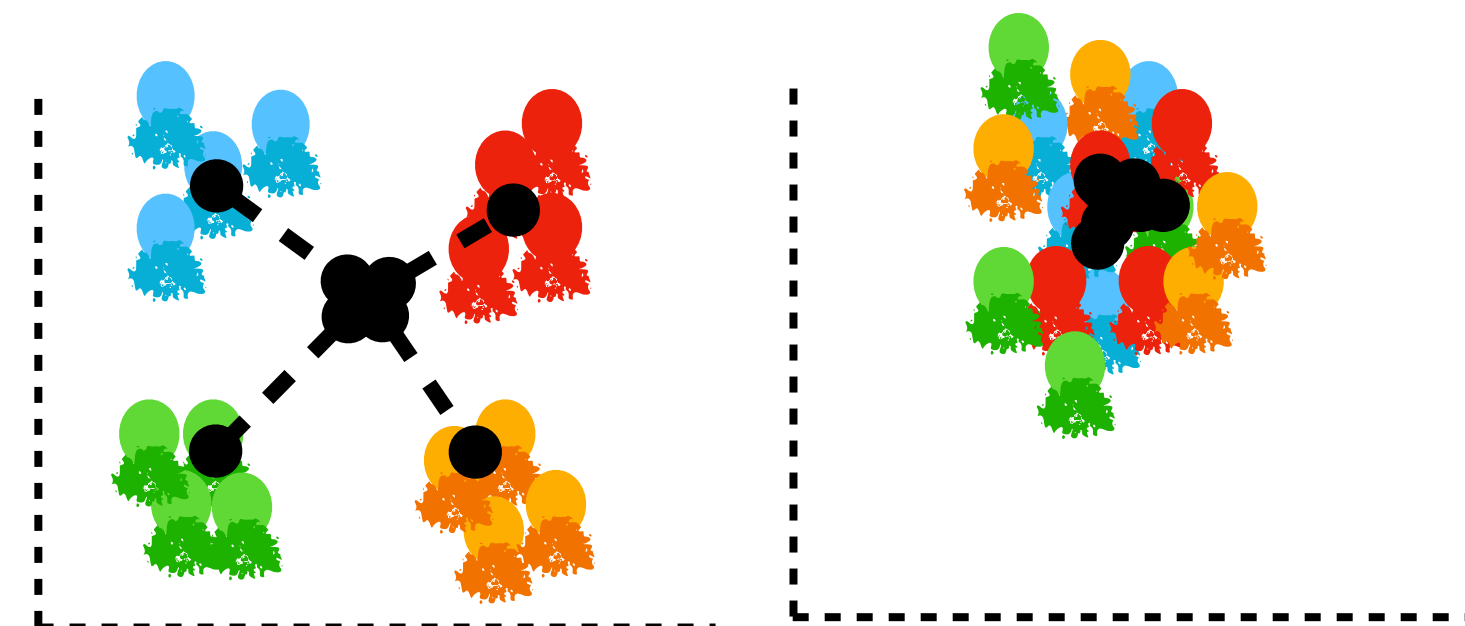
2. Cultural narratives and stereotypes about the upper class



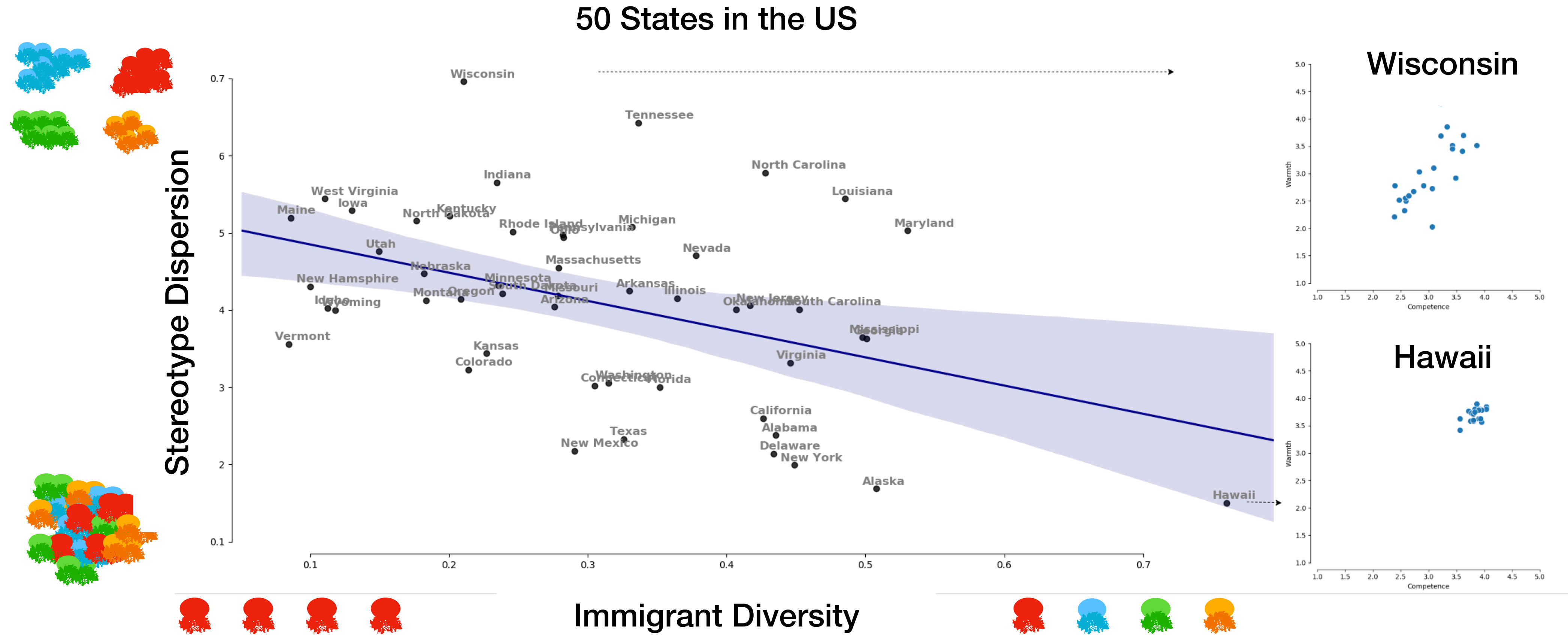
3. Immigrant diversity and perceived similarity



**Leveraging the context**

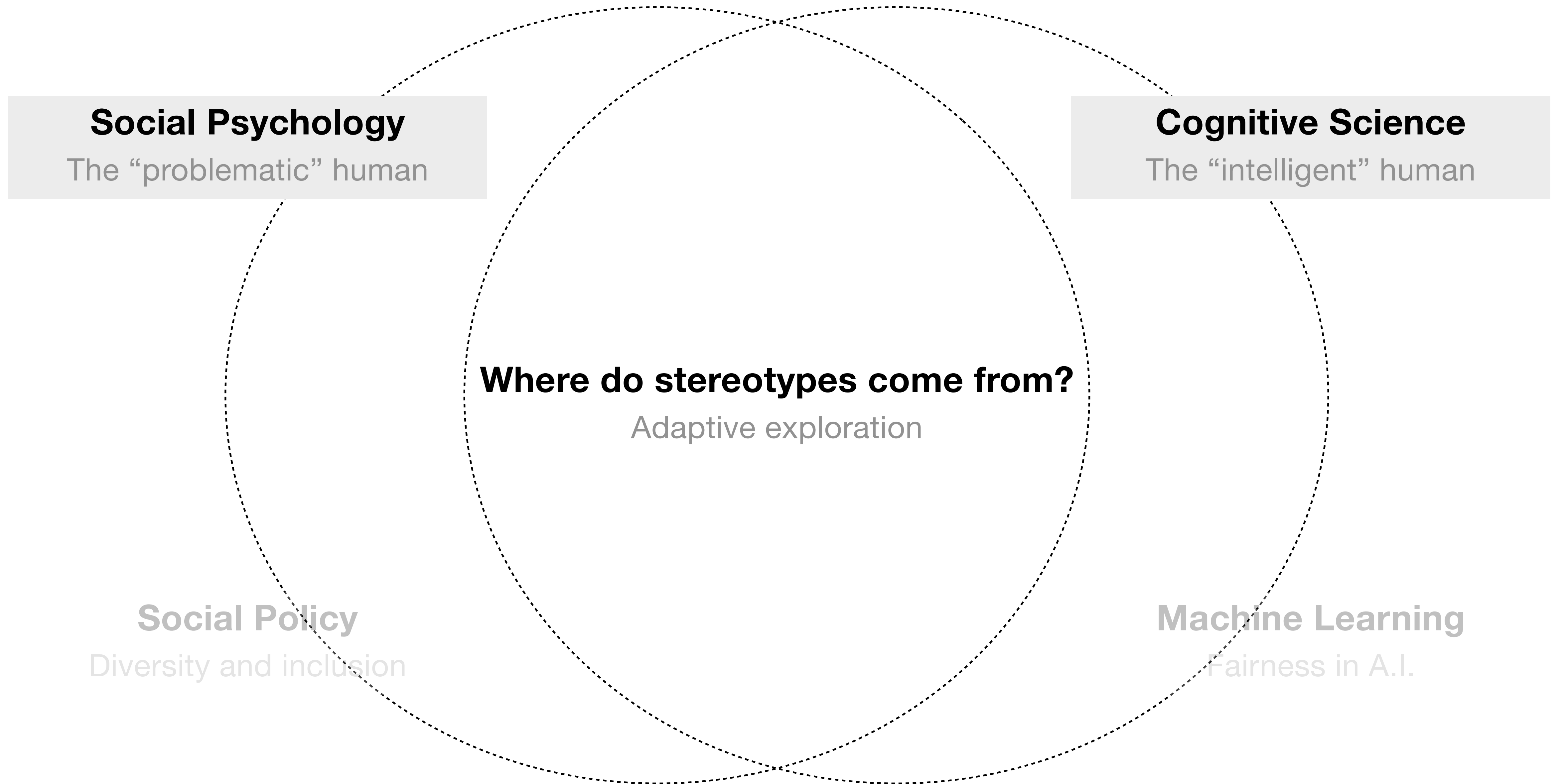


# The Psychology of How We **Make Sense of** the Social World



$r(48) = -.384, p = .006; b = -.282, 95\% CI [-.478, -.086], p = .008$

# The Psychology of How We **Make Sense of** the Social World



# The Psychology of How We **Make Sense** of the Social World

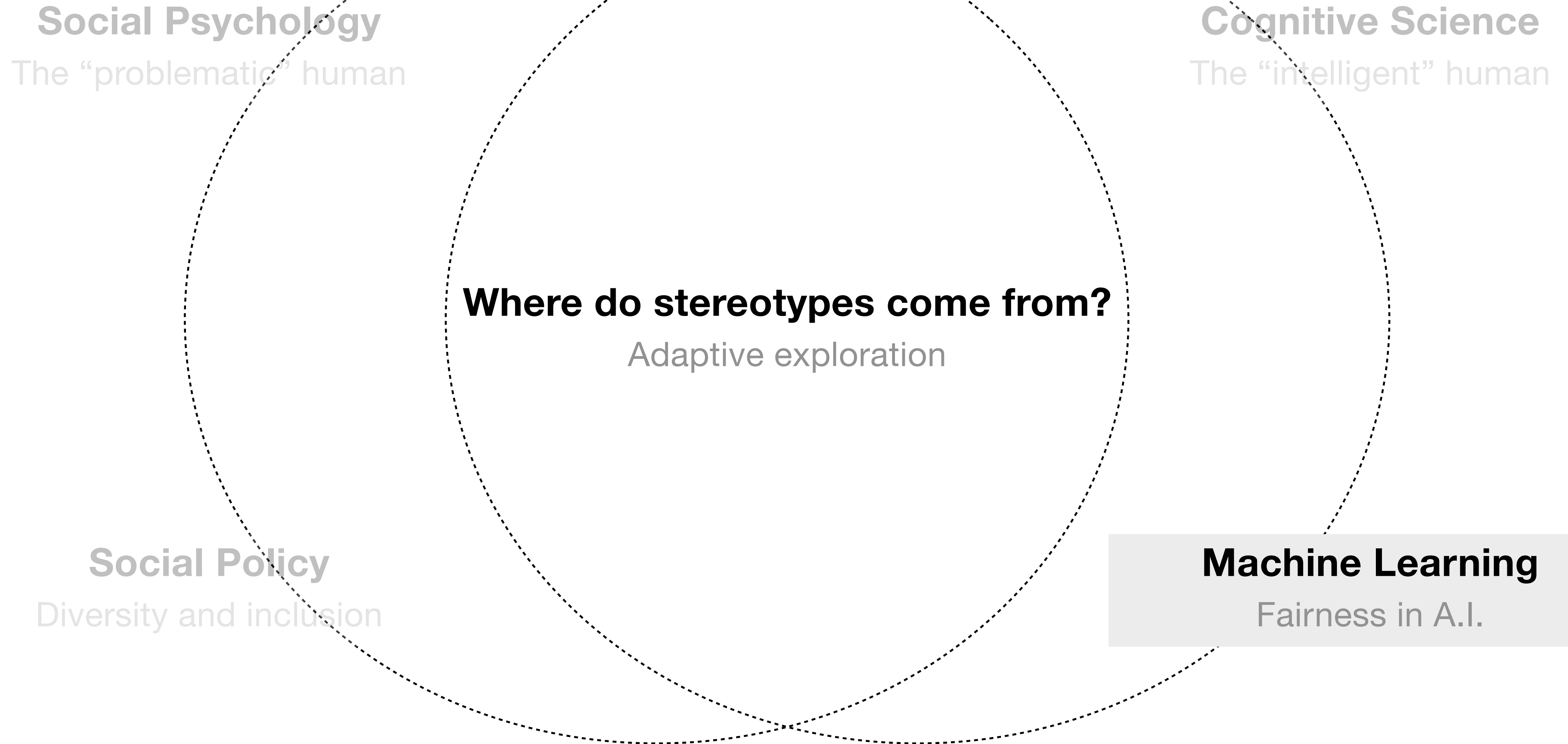
## **Social Psychology**

The “problematic” human

## **Cognitive Science**

The “intelligent” human

# The Psychology of How We **Make Sense of** the Social World



# The Psychology of How We **Make Sense of** the Social World

1. How to design individually intelligent and socially responsible human-AI systems?

**Machine Learning**

Fairness in A.I.

# The Psychology of How We **Make Sense of** the Social World

1. How to design individually intelligent and socially responsible human-AI systems?
2. Can we use insights from human social interactions to inform AI designs, vice versa?

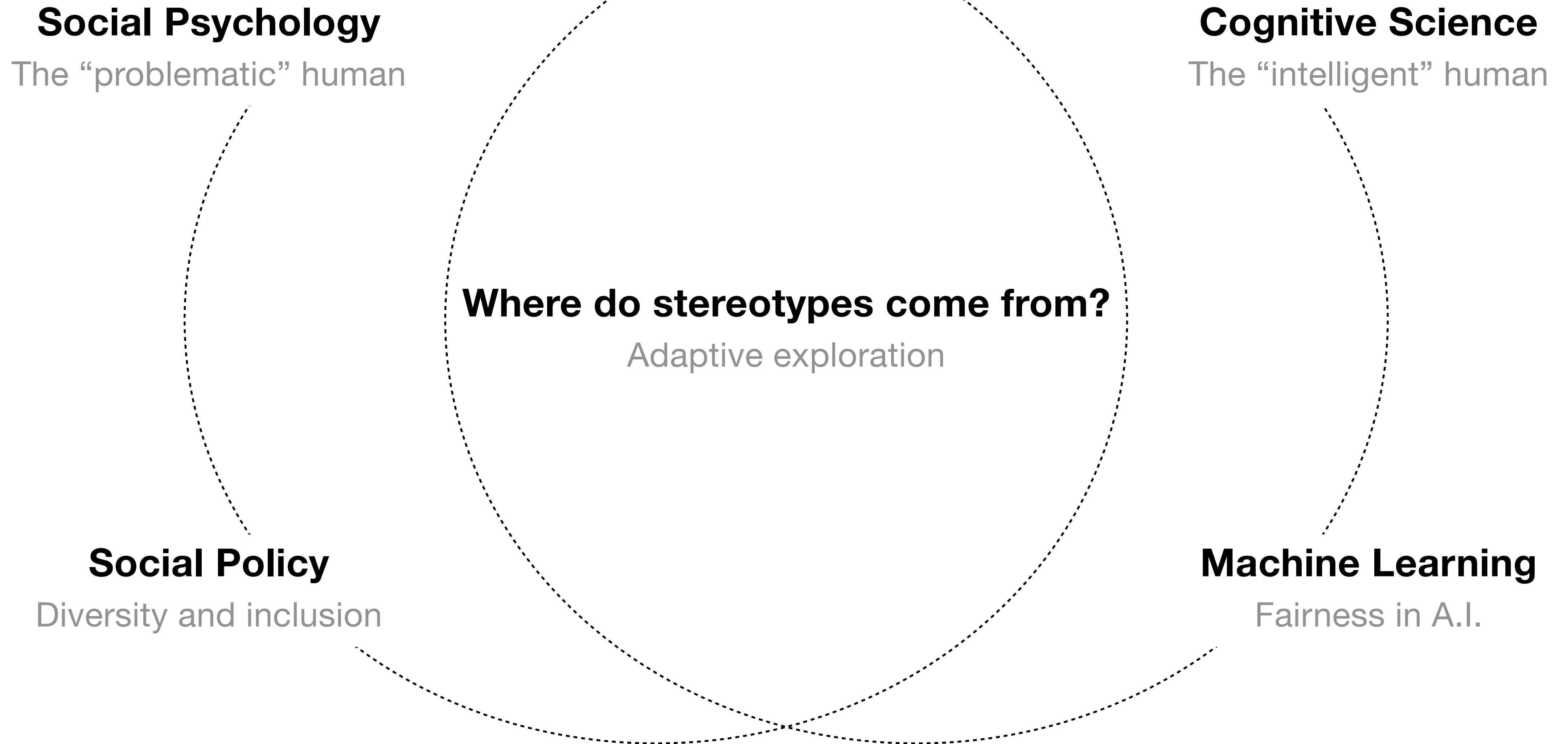


**Machine Learning**

Fairness in A.I.



# The Psychology of How We **Make Sense of** the Social World



Thank You!



Susan Fiske



Tom Griffiths



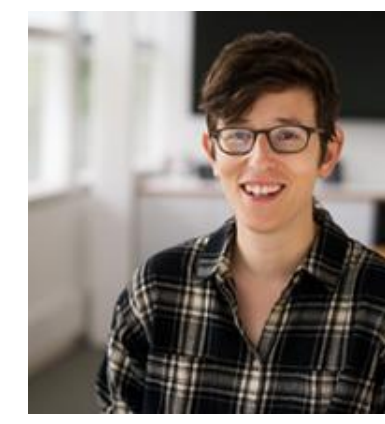
Alex Todorov



Kristina Olson



Stacey Sinclair



Diana Tamir



Gandalf Nicolas



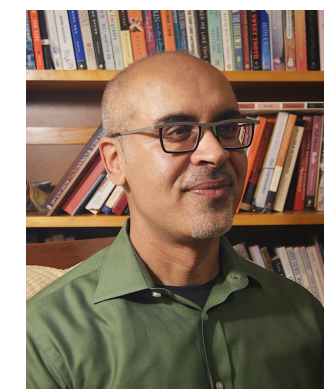
Sherry Wu



Lucy Grigoryan



Miguel Ramos



Varun Gauri



Stefan Uddenberg



Kevin McKee



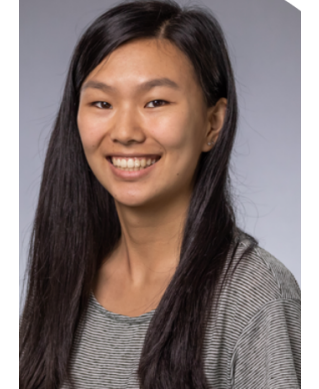
Kim Knipprath



Ted Summers



Mayank Agarwal



Angelina Wang