

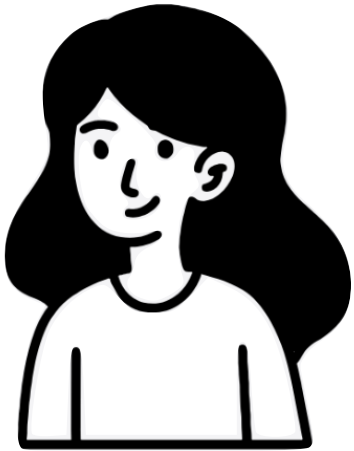


The **Hidden** Architecture of Stereotypes

Ella Qiawen Liu¹, Gary Lupyan², Xuechunzi Bai³

¹Princeton University, ²University of Wisconsin-Madison, ³University of Chicago

January 30th, 2026, RF-CDR Winter Retreat



Woman



Man



Poor



Rich

Caring

Decisive

Incompetent

Selfish



A simple line drawing of a woman with long, dark hair, wearing a simple top.

Woman

A simple line drawing of a man with short, dark hair, wearing a simple shirt.

Man

A simple line drawing of a person wearing a beanie and sleeping on a bench. There is a small table next to them with a bowl of food and a sign with a dollar sign.

Poor

A simple line drawing of a person's head with a large, cloud-like shape above it. The cloud shape contains several dollar signs (\$).

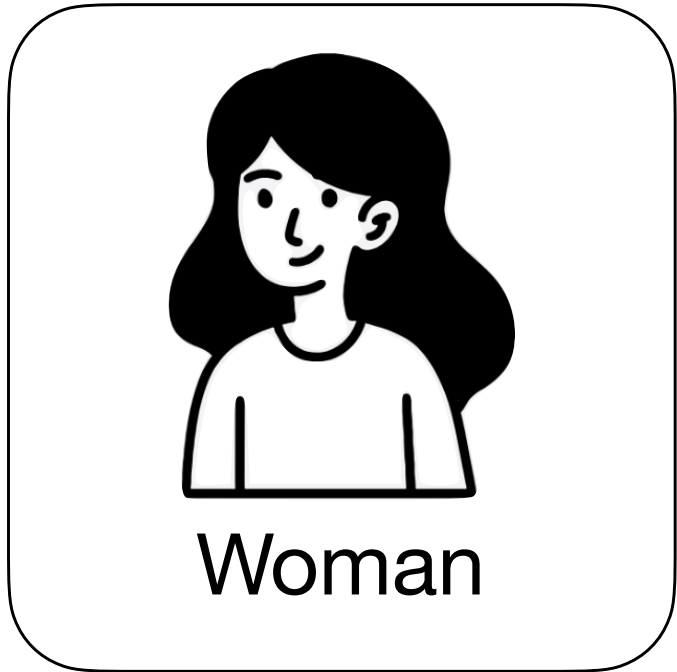
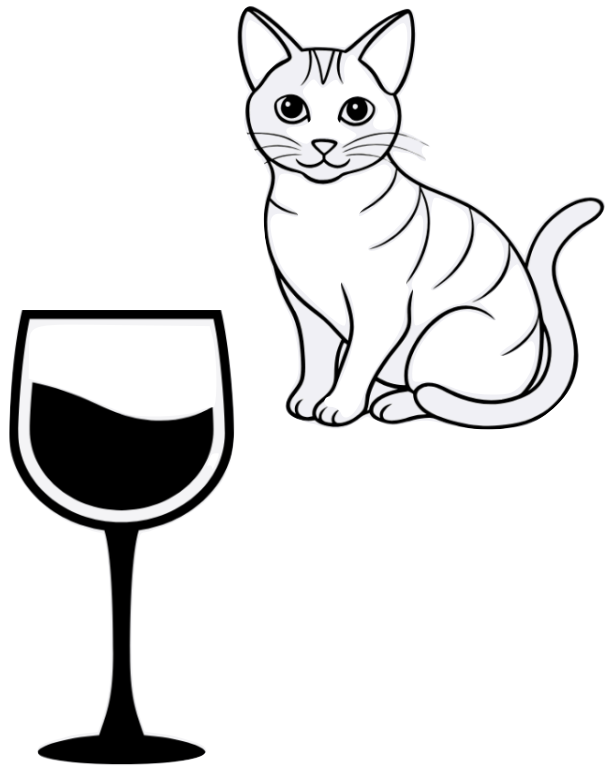
Rich

Caring

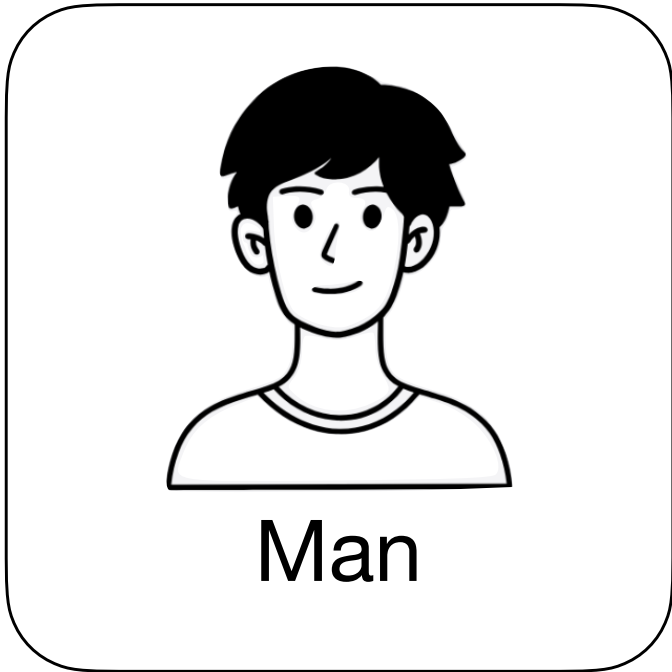
Decisive

Incompetent

Selfish



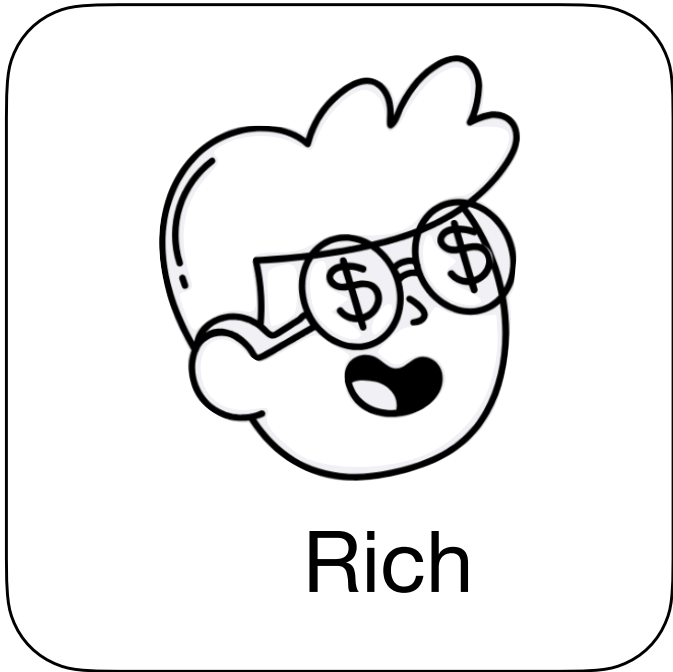
Woman



Man



Poor



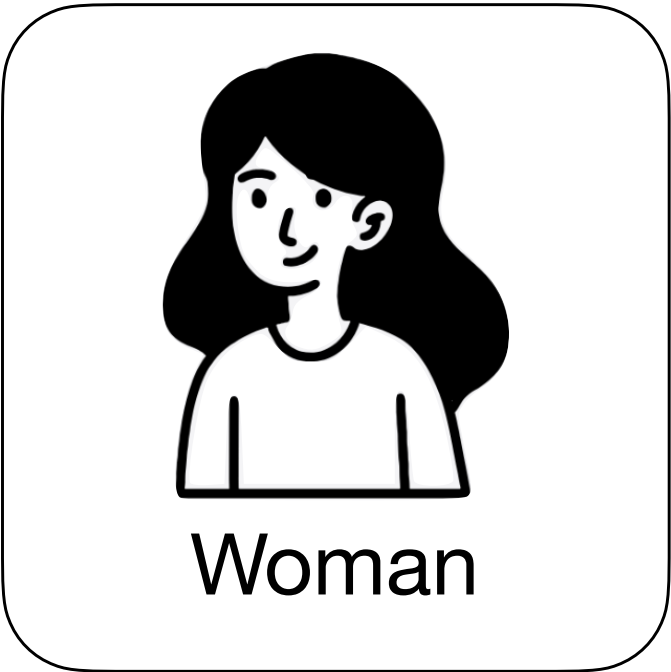
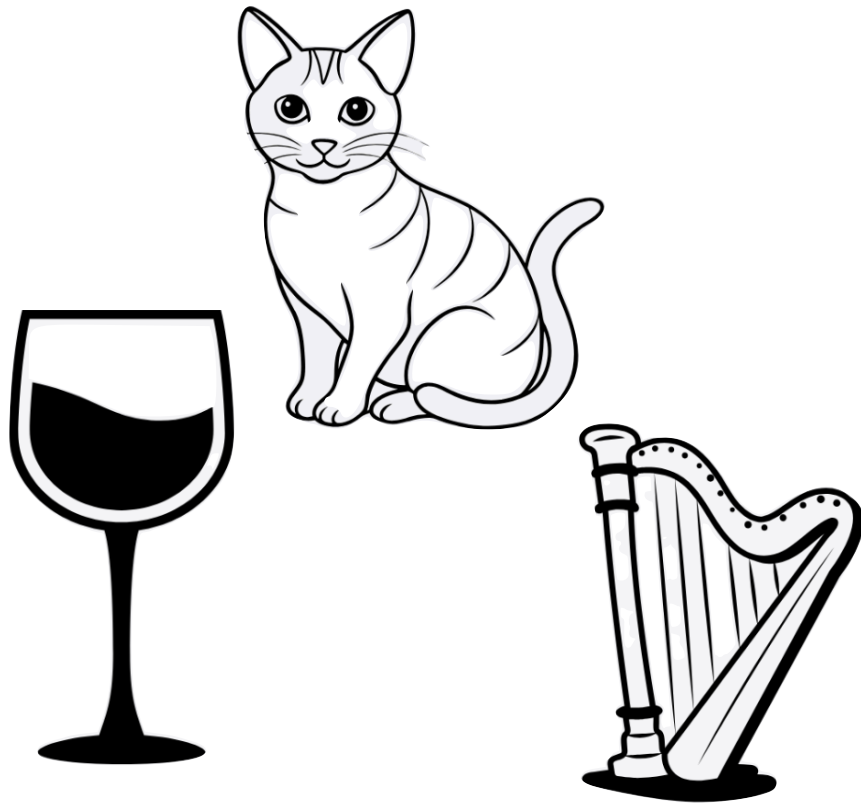
Rich

Caring

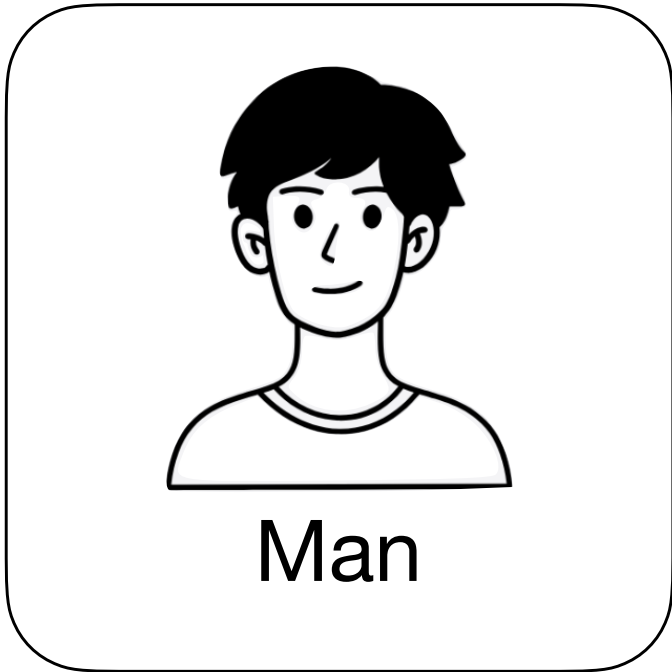
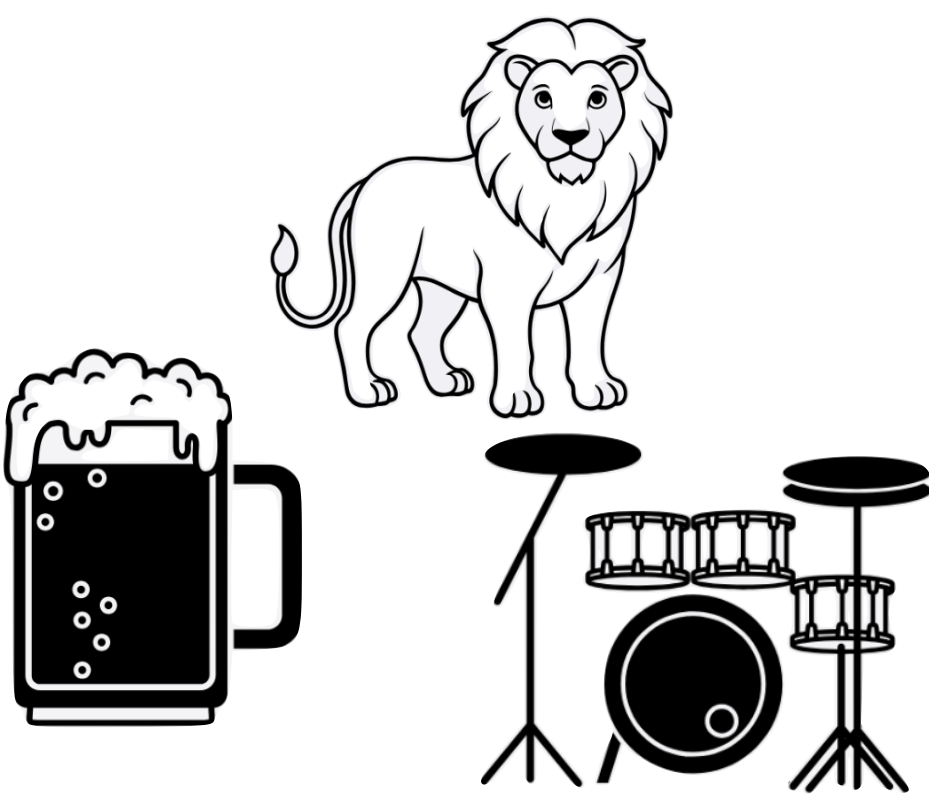
Decisive

Incompetent

Selfish



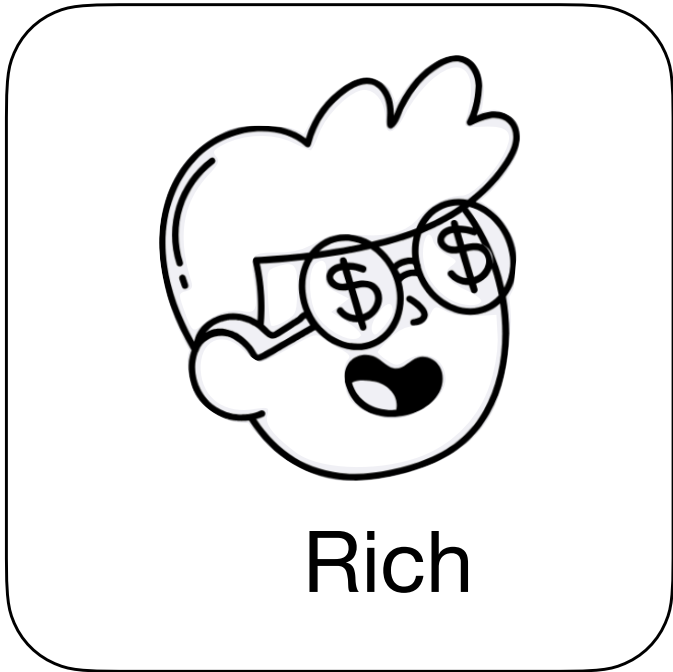
Woman



Man



Poor



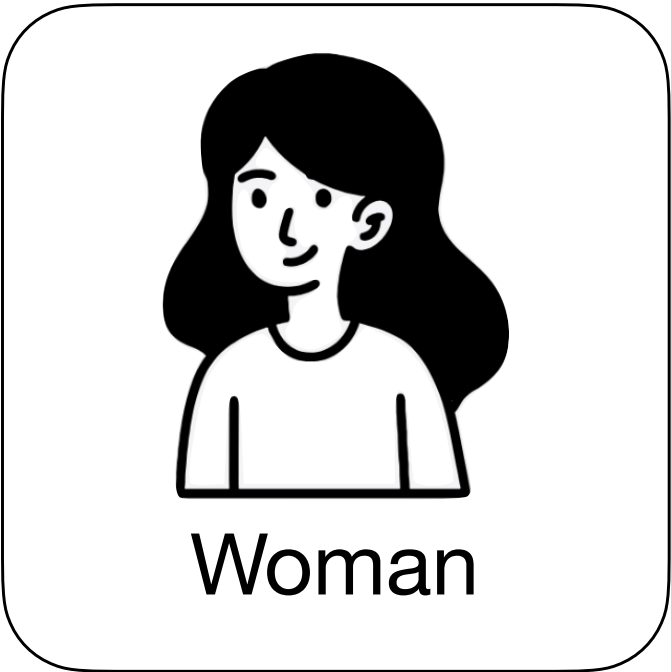
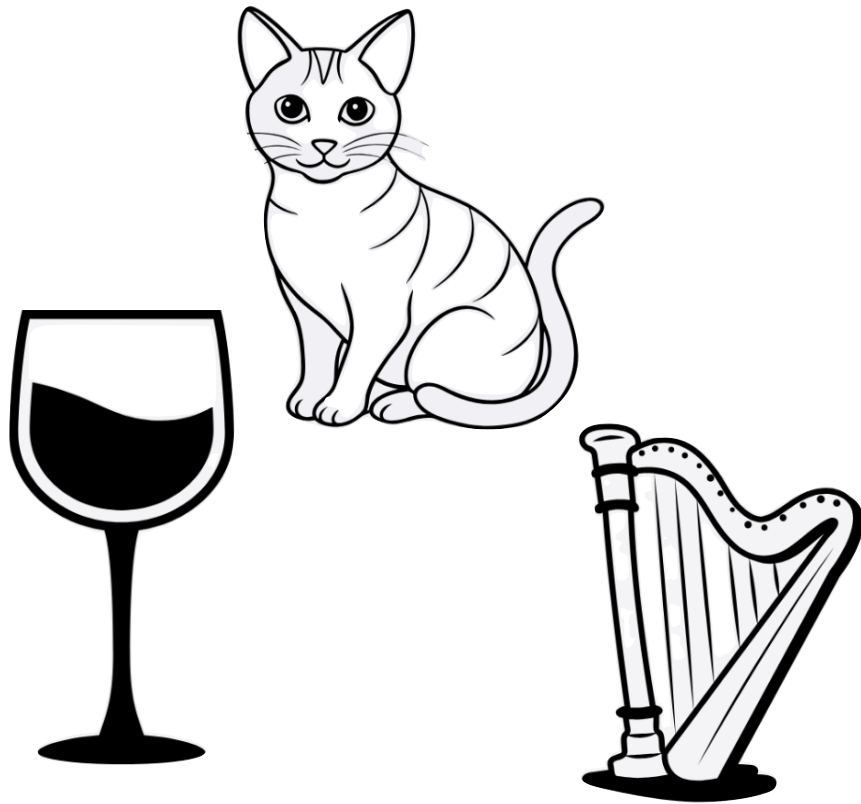
Rich

Caring

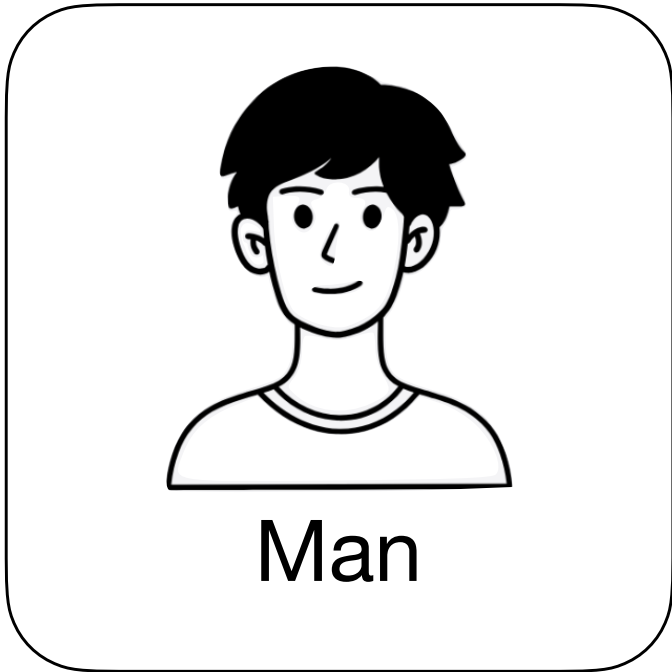
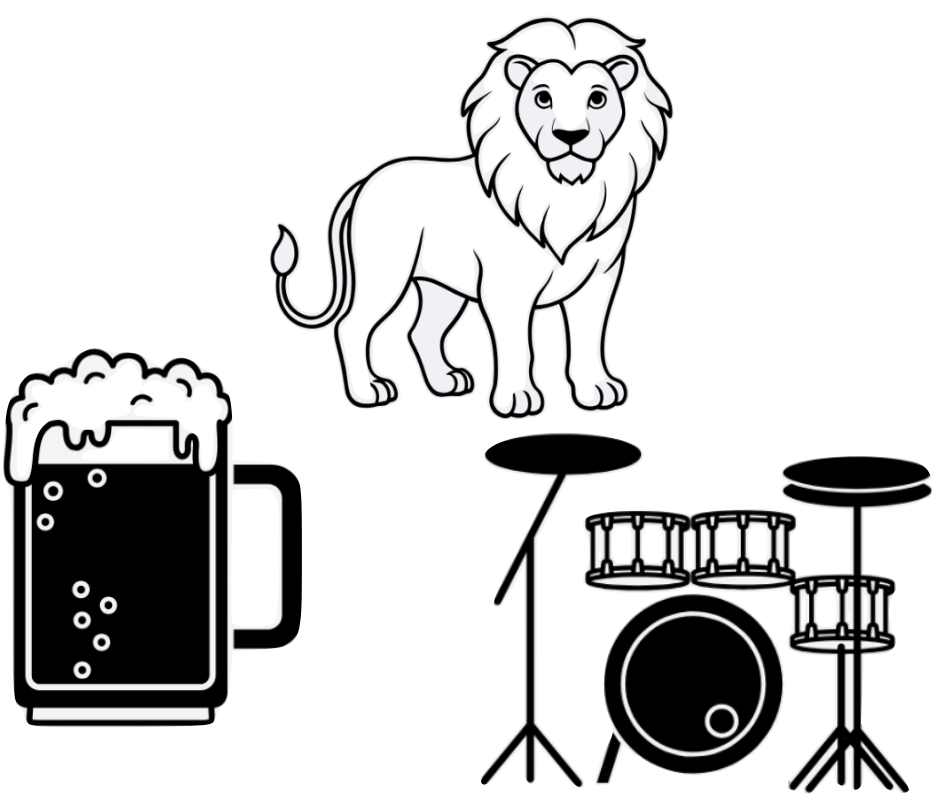
Decisive

Incompetent

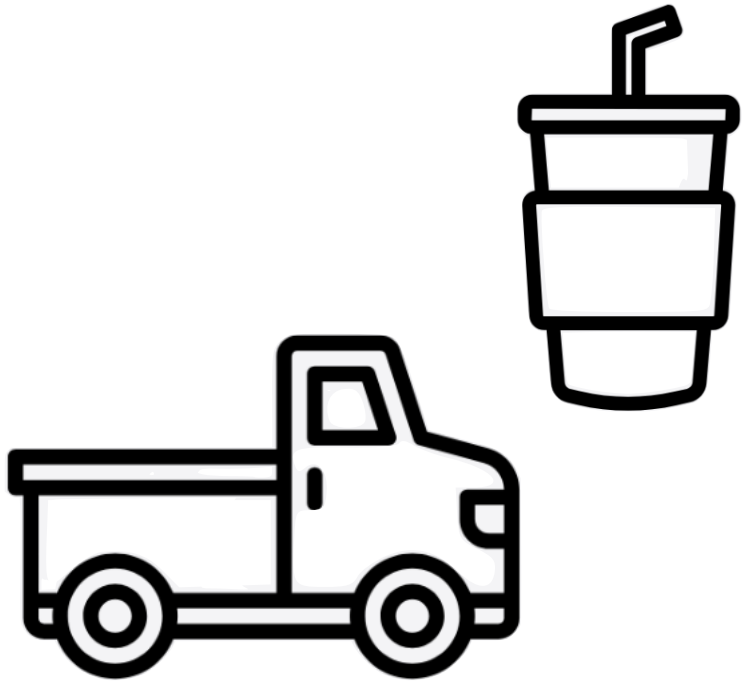
Selfish



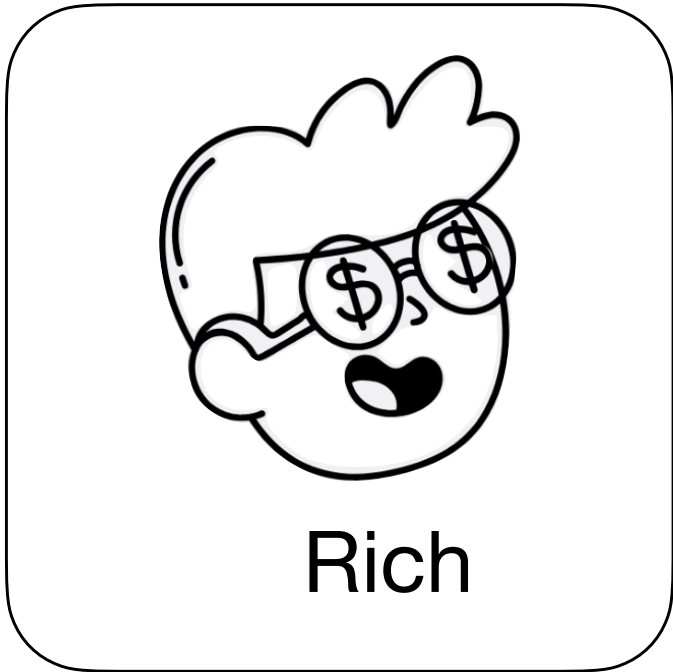
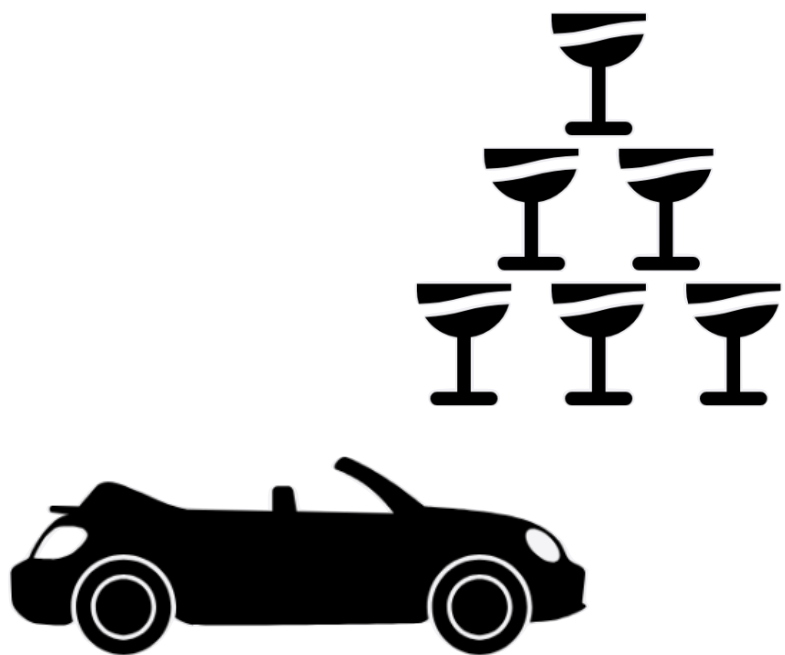
Woman



Man



Poor



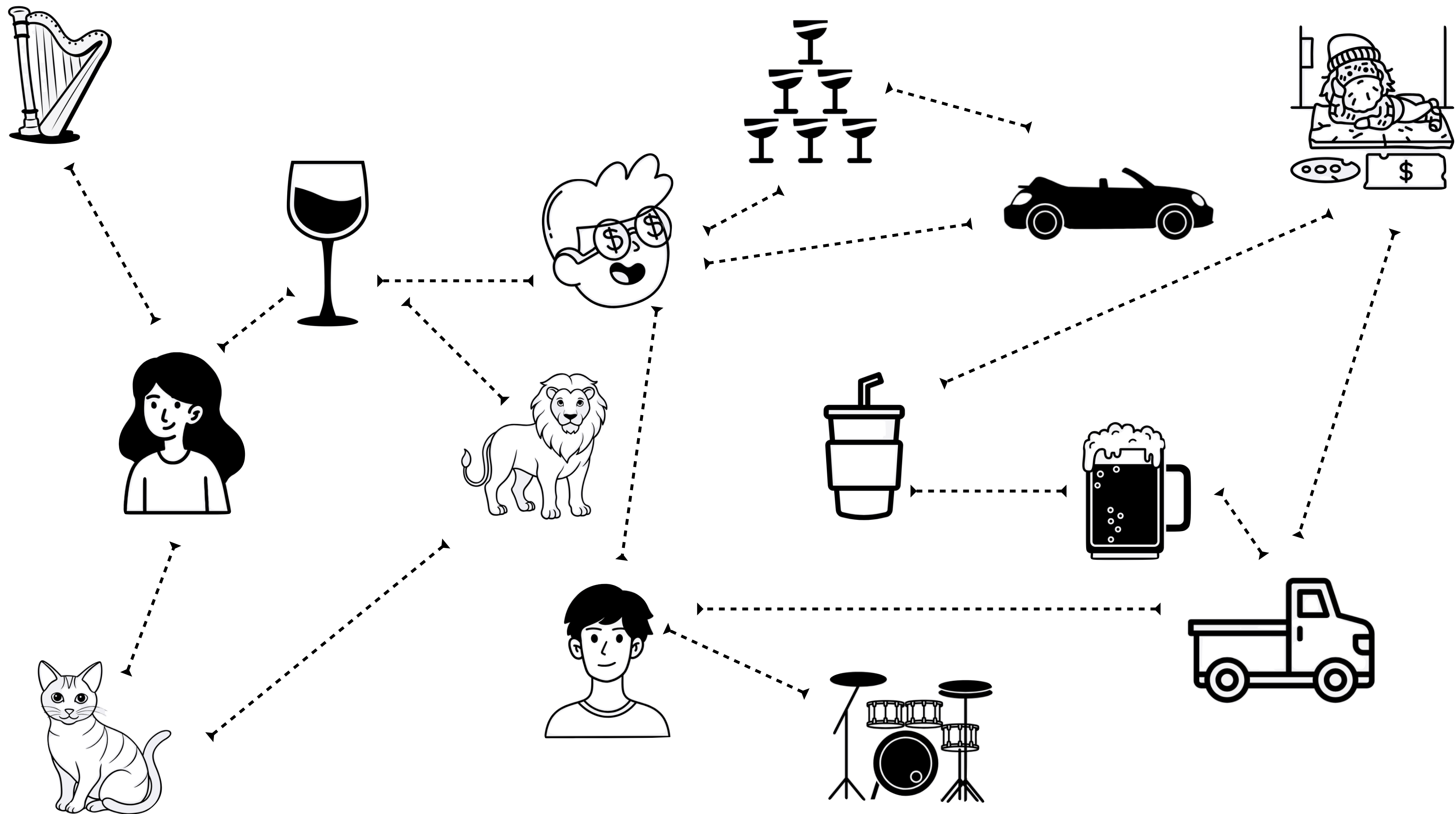
Rich

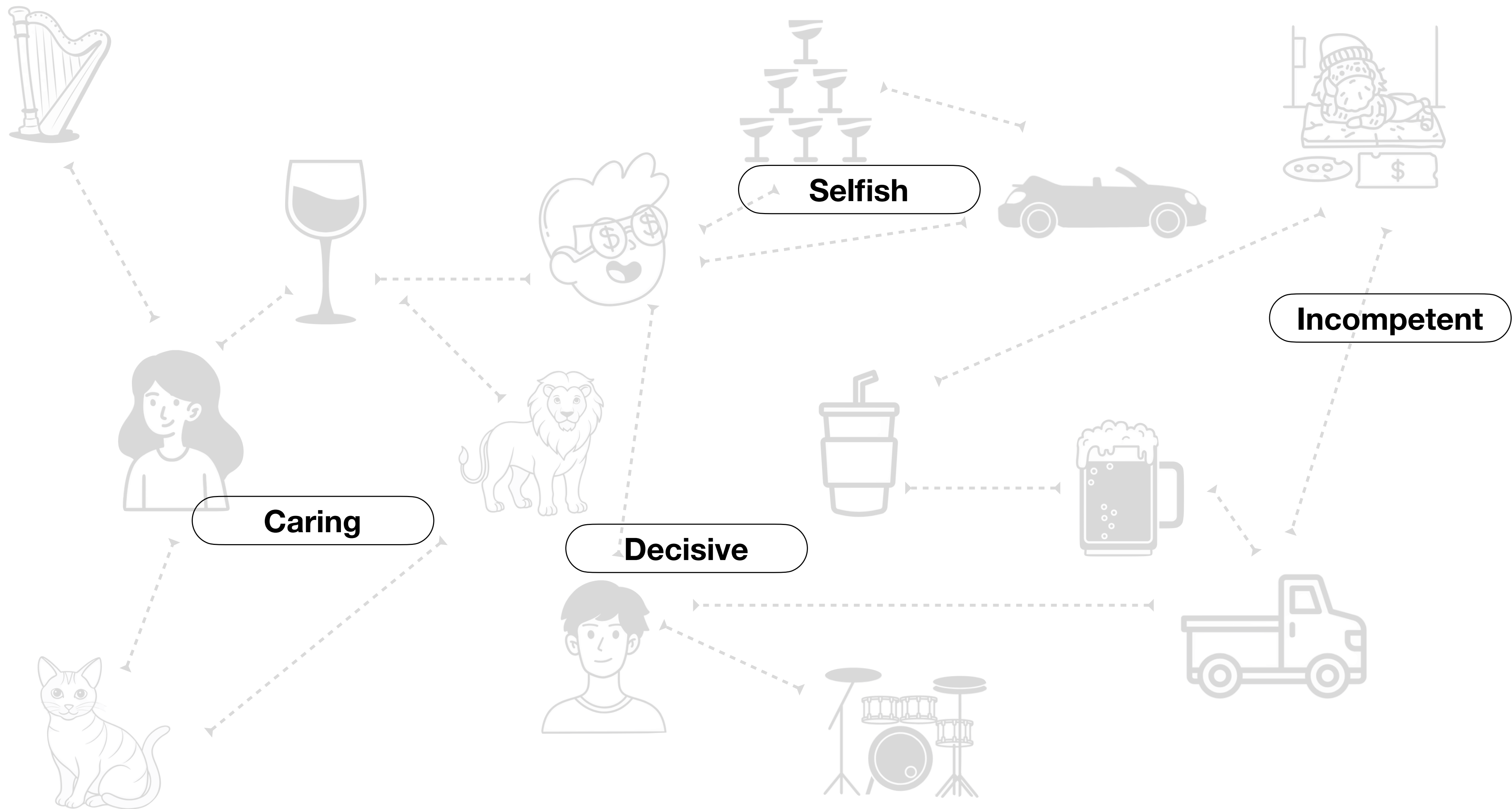
Caring

Decisive

Incompetent

Selfish





Our contribution: Bring this hidden architecture into view

1. Empirically show that social groups sit within rich associate networks with seemingly irrelevant nonsocial objects.

Our contribution: Bring this hidden architecture into view

1. Empirically show that social groups sit within rich associate networks with seemingly irrelevant nonsocial objects.
2. Stereotypical traits, warmth and competence, are emergent properties of these networks.

Our contribution: Bring this hidden architecture into view

1. Empirically show that social groups sit within rich associate networks with seemingly irrelevant nonsocial objects.
2. Stereotypical traits, warmth and competence, are emergent properties of these networks.
3. Editing nonsocial objects associated with a group is sufficient to change the stereotypical traits that people use to make sense of the group.

Our contribution: Bring this hidden architecture into view

1. Empirically show that social groups sit within rich associate networks with seemingly irrelevant nonsocial objects.
2. Stereotypical traits, warmth and competence, are emergent properties of these networks.
3. Editing nonsocial objects associated with a group is sufficient to change the stereotypical traits that people use to make sense of the group.
4. The magnitude of change is bigger when people engage in deep analogical reasoning, rather than superficial alignment.

Existing research is grounded in “group” - “trait” mapping

- In definition
- In measurement
- In intervention (cognitive)

Existing research is grounded in “group” - “trait” mapping

- In definition
 - ...as abstract knowledge structures that link a group with specific traits or behaviors (Hilton & von Hippel, 1996; Hamilton & Sherman, 2014)

Existing research is grounded in “group” - “trait” mapping

- In definition
- In measurement
 - ...list, rate, arrange, and react to pairs of attributes and groups (Asch, 1946; Katz & Braly, 1933; Fiske et al., 2002; Osgood, 1955; Koch et al., 2016; Nicolas et al., 2022; Greenwald et al., 1998)
 - “Write five adjectives that come to mind when you think of a rich person.”
 - “Rate Mexican immigrants on how friendly you think American society perceives them.”
 - “Sort many social groups in terms of how [dis]similar they are.”
 - “React as quickly as possible by pressing keys when shown women - friendly.”

Existing research is grounded in “group” - “trait” mapping

- In definition
- In measurement
- In intervention (cognitive)
 - Exposing counter-stereotypical pairs (Forscher et al., 2019; Weber & Crocker, 1983; Johnston & Hewstone, 1992; Dasgupta & Greenwald, 2001; Olson & Fazio, 2001; Kurdi & Banaji, 2017)
 - Present “lawyers” alongside a list of attributes “quiet, disorganized”
 - Repeated exposure of “admired Blacks” or “terrible ice cream”
 - Instruction only “you will see pleasant Lappians.”

Existing research is grounded in “group” - “trait” mapping

- In definition
- In measurement
- In intervention (cognitive)
- Anecdotal evidence associates groups with nonsocial objects
 - Female - ballet, flowers, skirt vs. caring, patient, warm (Blair & Banaji, 1996)
 - Blacks at family barbecue vs. gang incident (Wittenbrink, Judd, & Park, 2001)
 - Learning STEM in classrooms with nature posters vs. video games (Cheryan et al, 2009)

If X were a Y , what would it be?

If X were a Y, what would it be? (Liu & Lupyan, 2023; Rumelhart & McClelland, 1986; Mikolov et al., 2013; Lakoff & Johnson, 1980; Gentner, 1983)

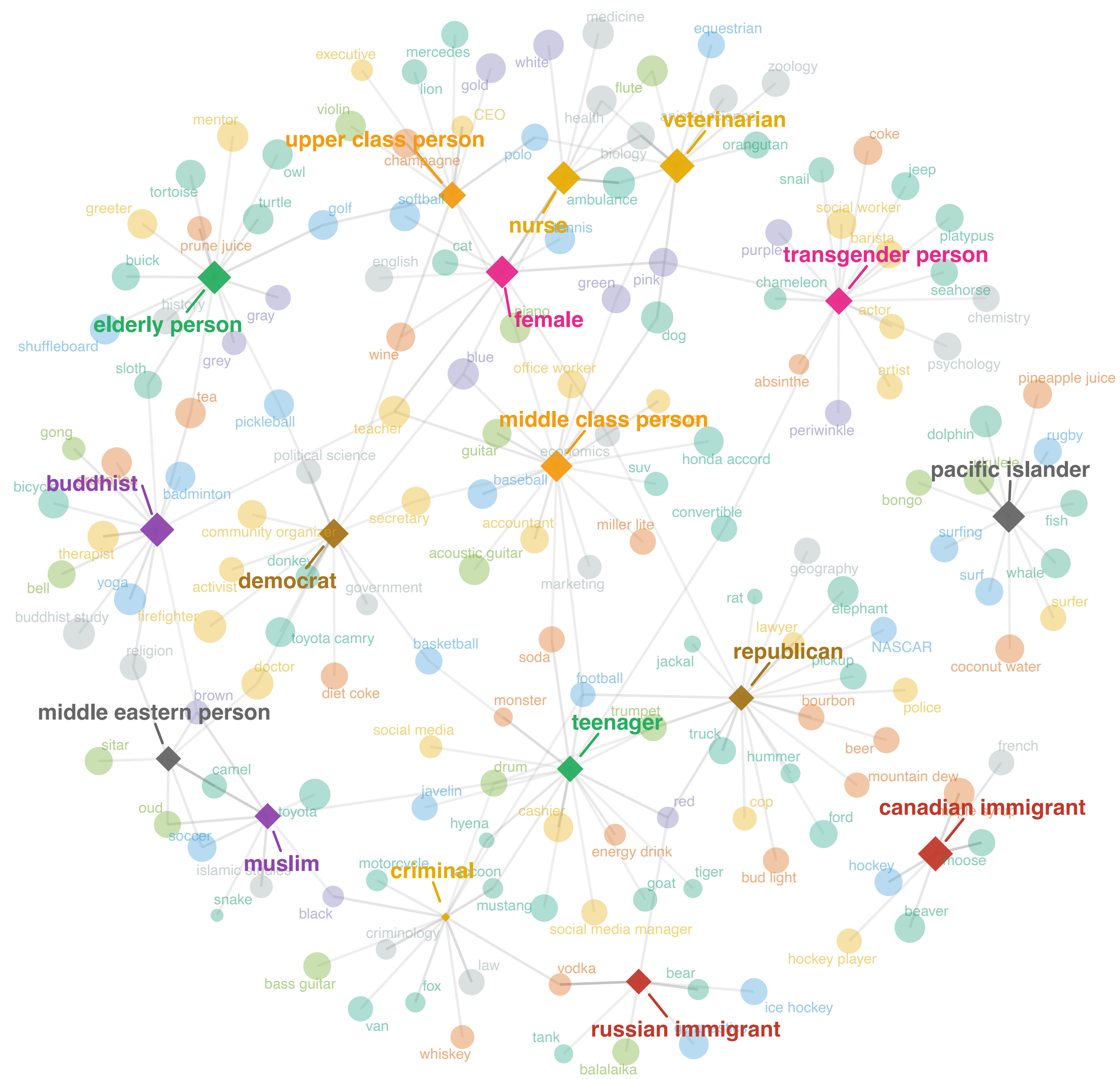
- If science were a color, what would it be?
- If a cat were a branch of knowledge, what would it be?

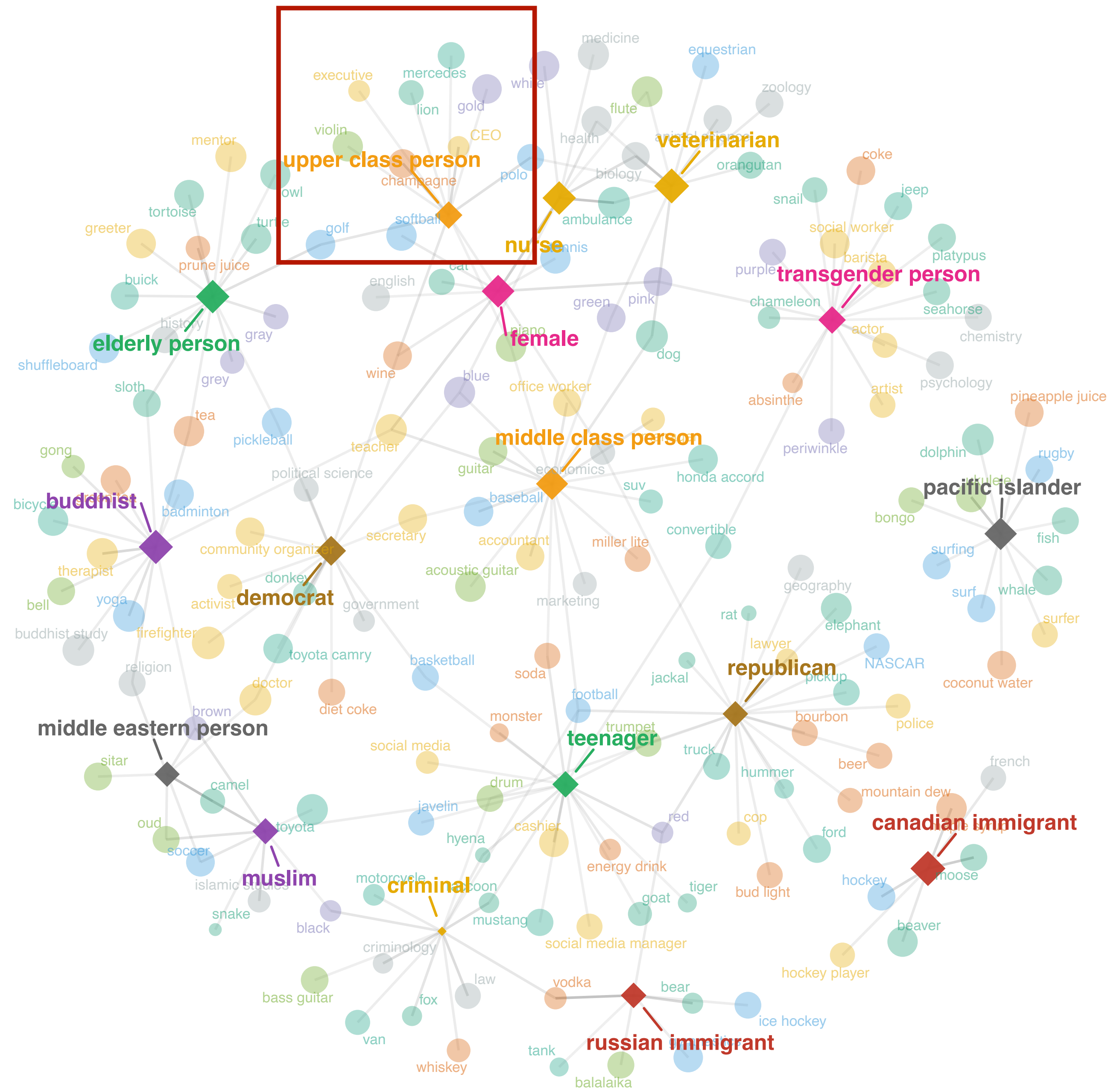
If X were a Y, what would it be? (Liu & Lupyan, 2023; Rumelhart & McClelland, 1986; Mikolov et al., 2013; Lakoff & Johnson, 1980; Gentner, 1983)

- If science were a color, what would it be?
- If a cat were a branch of knowledge, what would it be?
- If a woman were a musical instrument, what would it be?
- If a rich person were a beverage, what would it be?

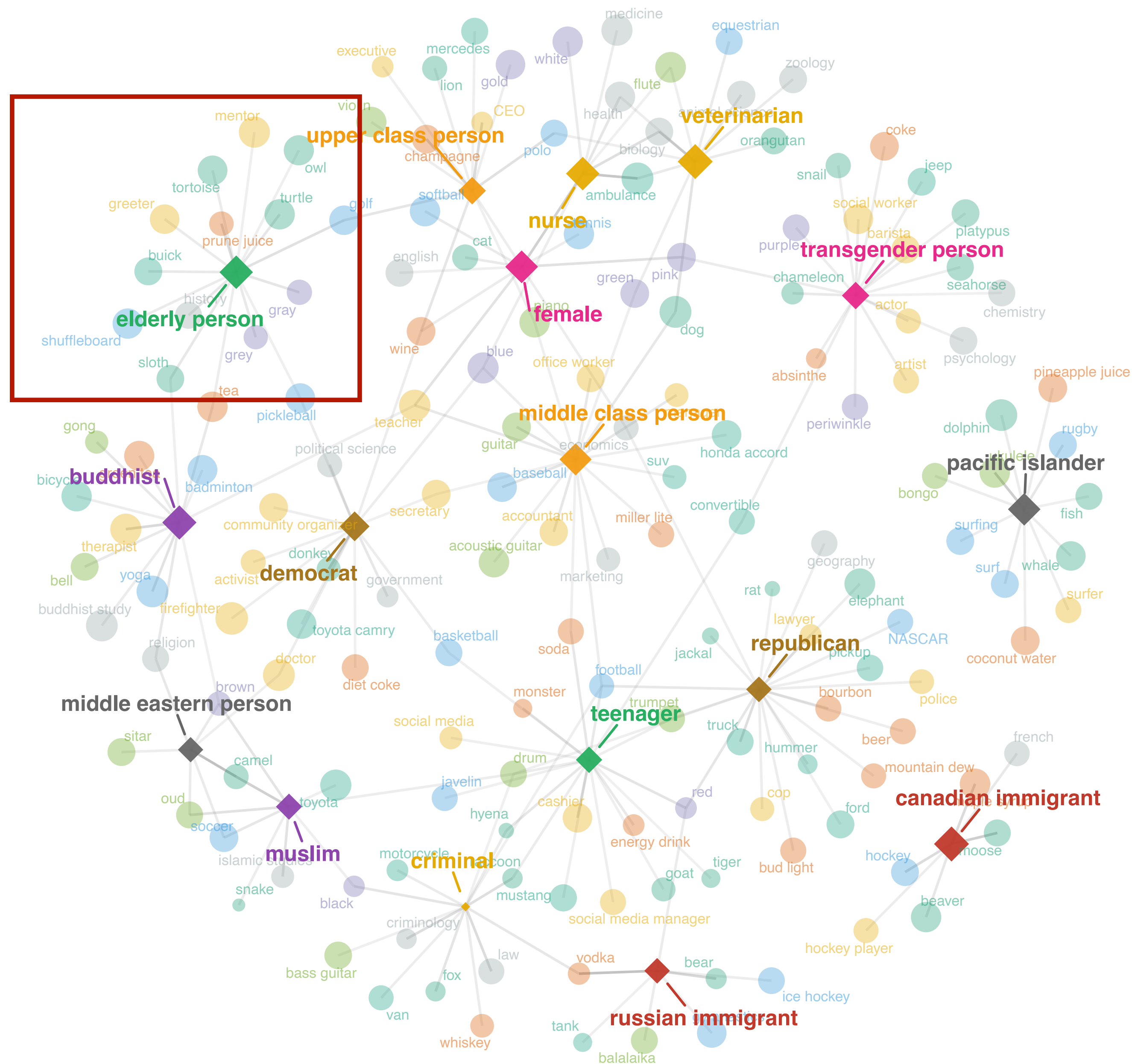
Generating: If X were a Y, what would it be?

- 100 social groups onto 8 cross domains
 - X = (race, age, gender, social class, religion, political orientation, national origin, occupations)
 - Y = (animals, jobs, sports, vehicles, beverages, colors, branches of knowledge, musical instrument)
- N = 628 online participants, answered for 60 questions, in 15 minutes.
 - (each question was answered by at least 50 participants)
 - (responses do not indicate endorsement or personal opinions)

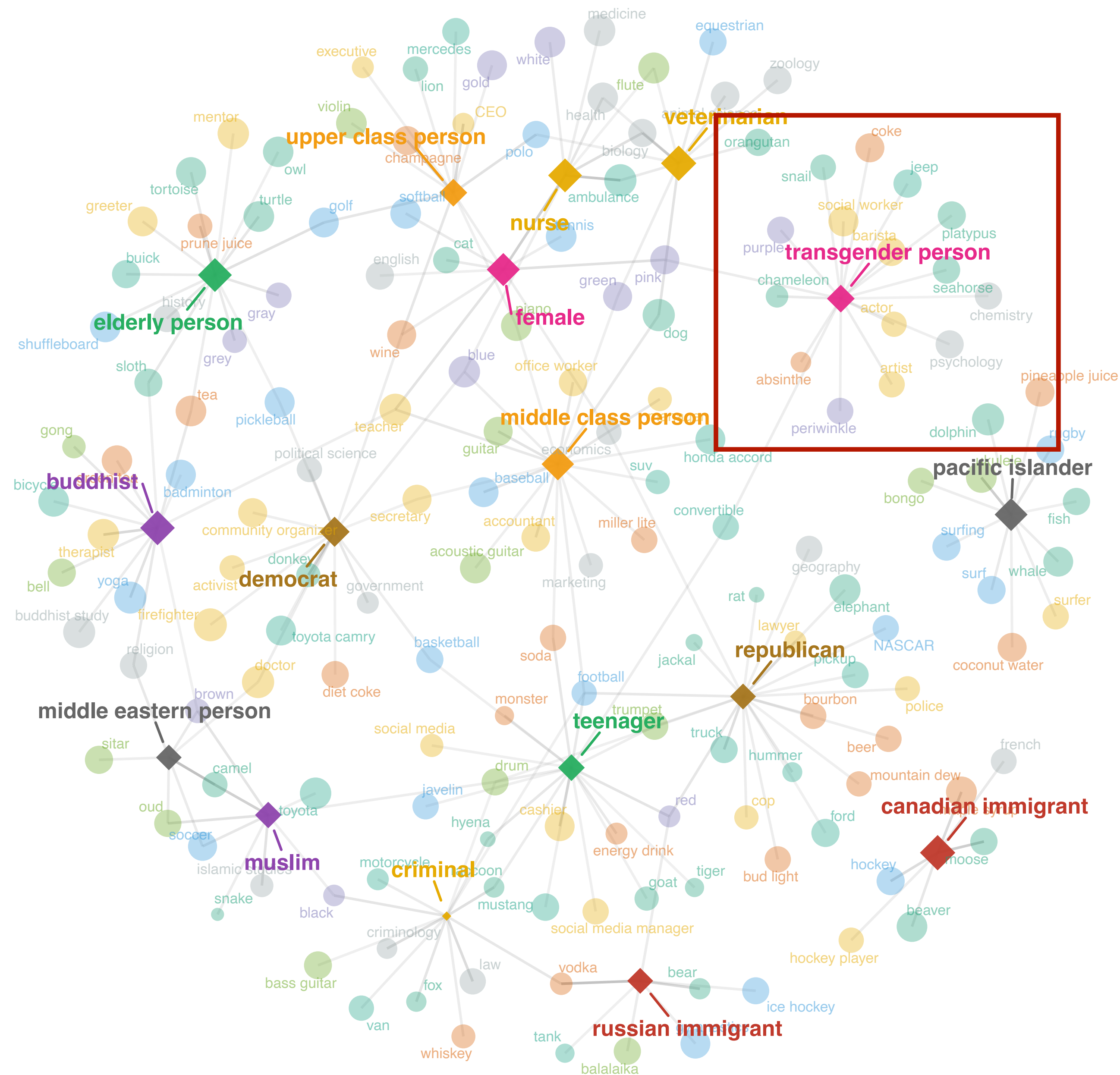




Motivation | Prior Work | Our Idea | Empirical Results | Implication



Motivation | Prior Work | Our Idea | Empirical Results | Implication



Warmth

Competence

Averaged Associated Concepts Warmth Rating (GPT)

0.04
0.02
0.00

-1.0

-0.5

0.0

0.5

1.0

-1.0

-0.5

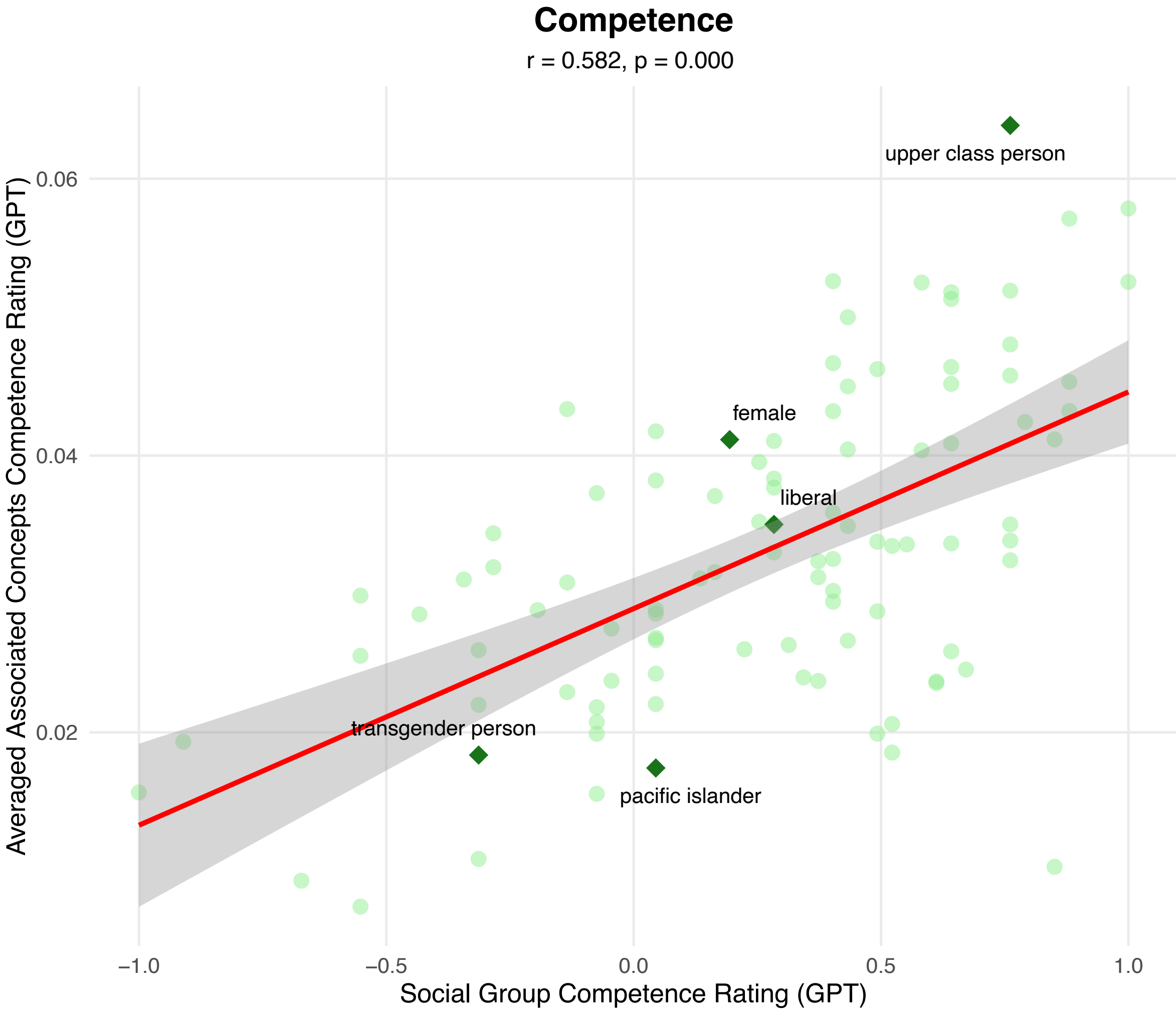
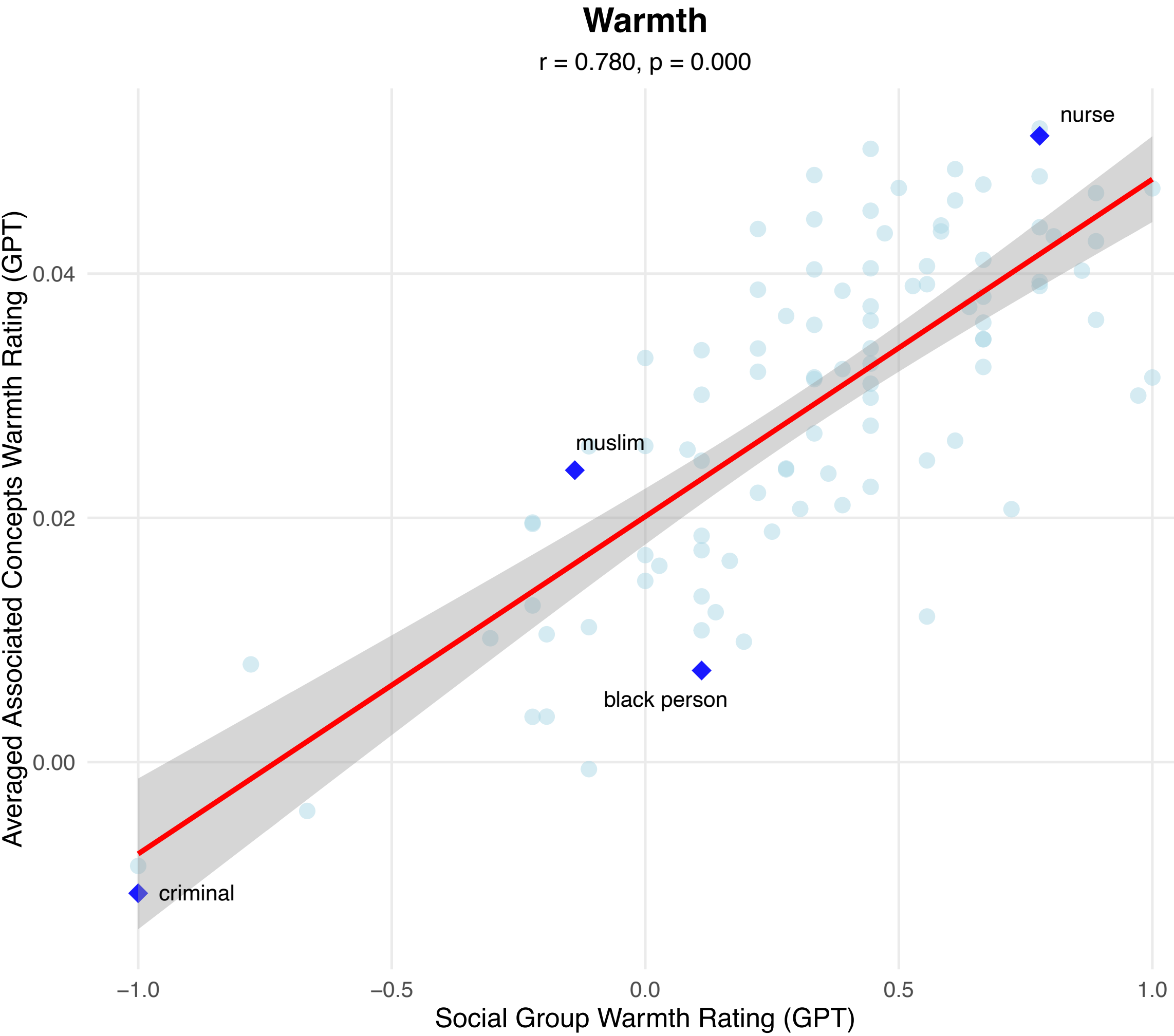
0.0

0.5

1.0

Social Group Warmth Rating (GPT)

Social Group Competence Rating (GPT)



Editing: If X were a Y , it would be y .

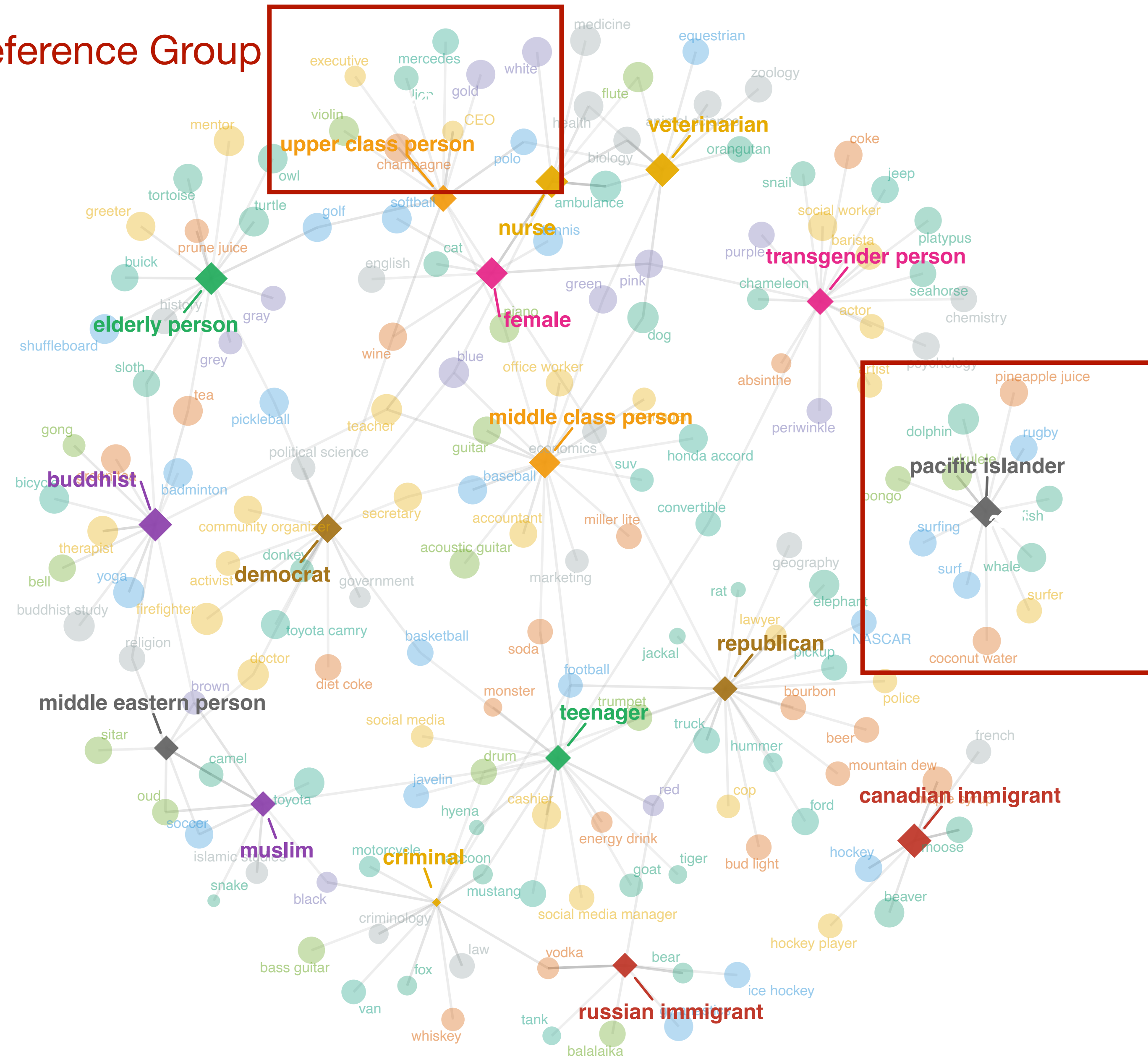
Write a brief explanation for why the connection might make sense.

Editing: If X were a Y, it would be y.

Write a brief explanation for why the connection might make sense.

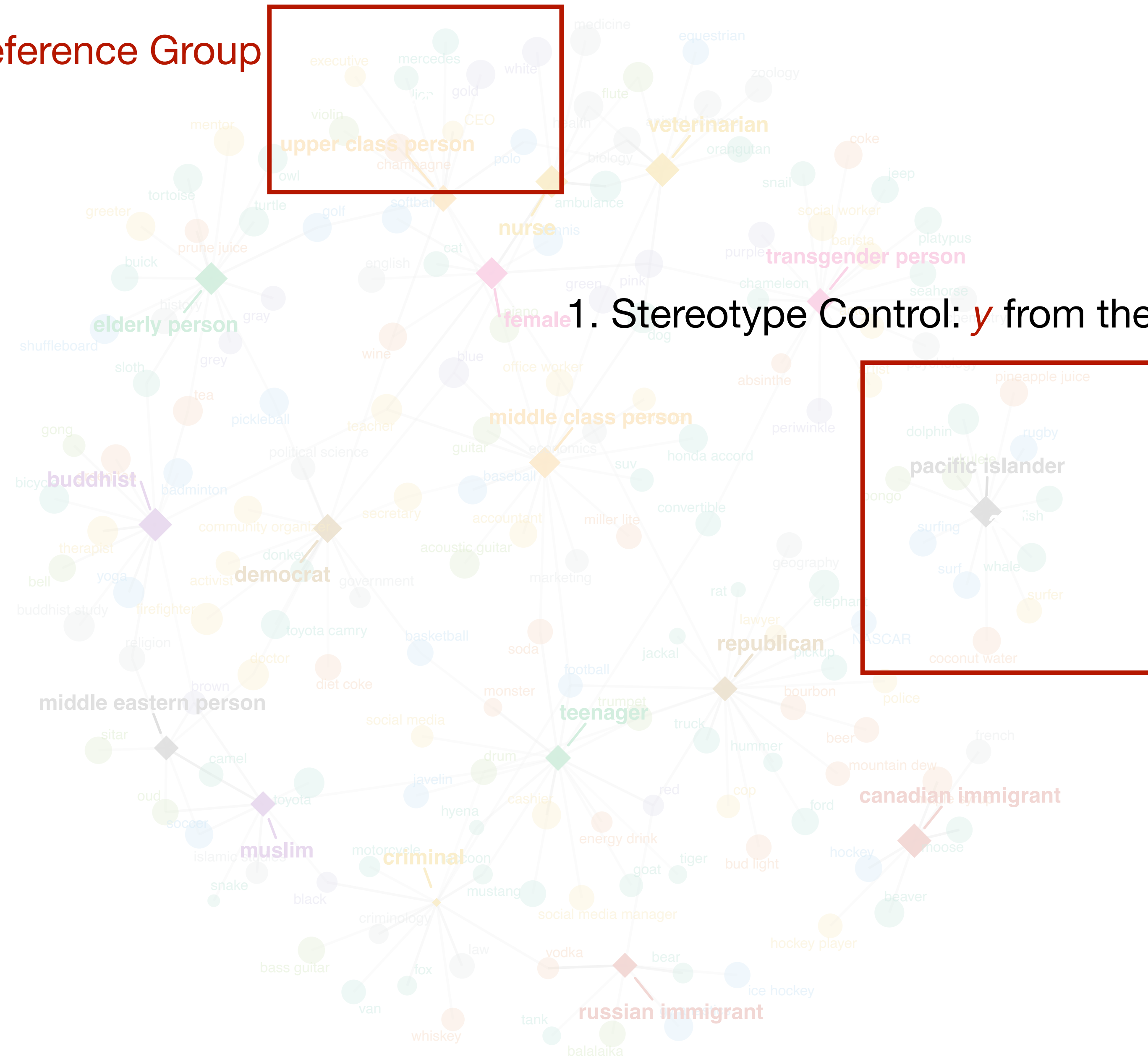
- Sampled 7 stereotyped groups and 3 reference groups
 - Stereotyped groups = (women; {liberals, transgender, Pacific Islanders}; {Muslims, Black people, criminals})
 - Reference groups = (athletes; upper-class; nurses)
- N = 1794 online participants, answered for 15-21 statements, in 15 minutes
 - (each question was answered by at least 50 participants)
 - (responses do not indicate endorsement or personal opinions; attempt to justify why the mappings make sense)

Reference Group



Stereotyped Group

Reference Group



Stereotyped Group

Condition network stereotype control random control

Physical Competence

Cognitive Competence

Warmth

2

1

0

-1

-2

Female

Liberals

Pacific
Islander

Transgender
people

Black
People

Criminals

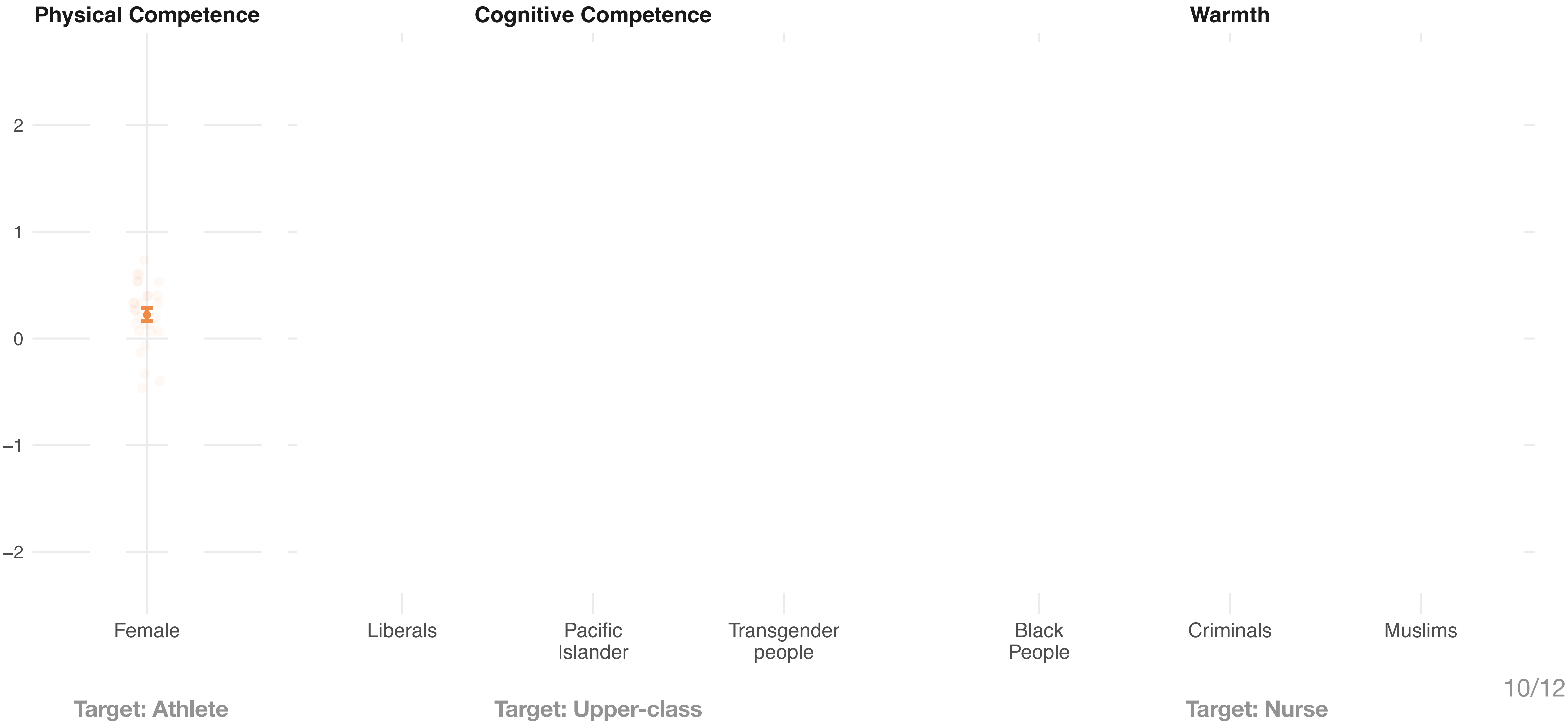
Muslims

Target: Athlete

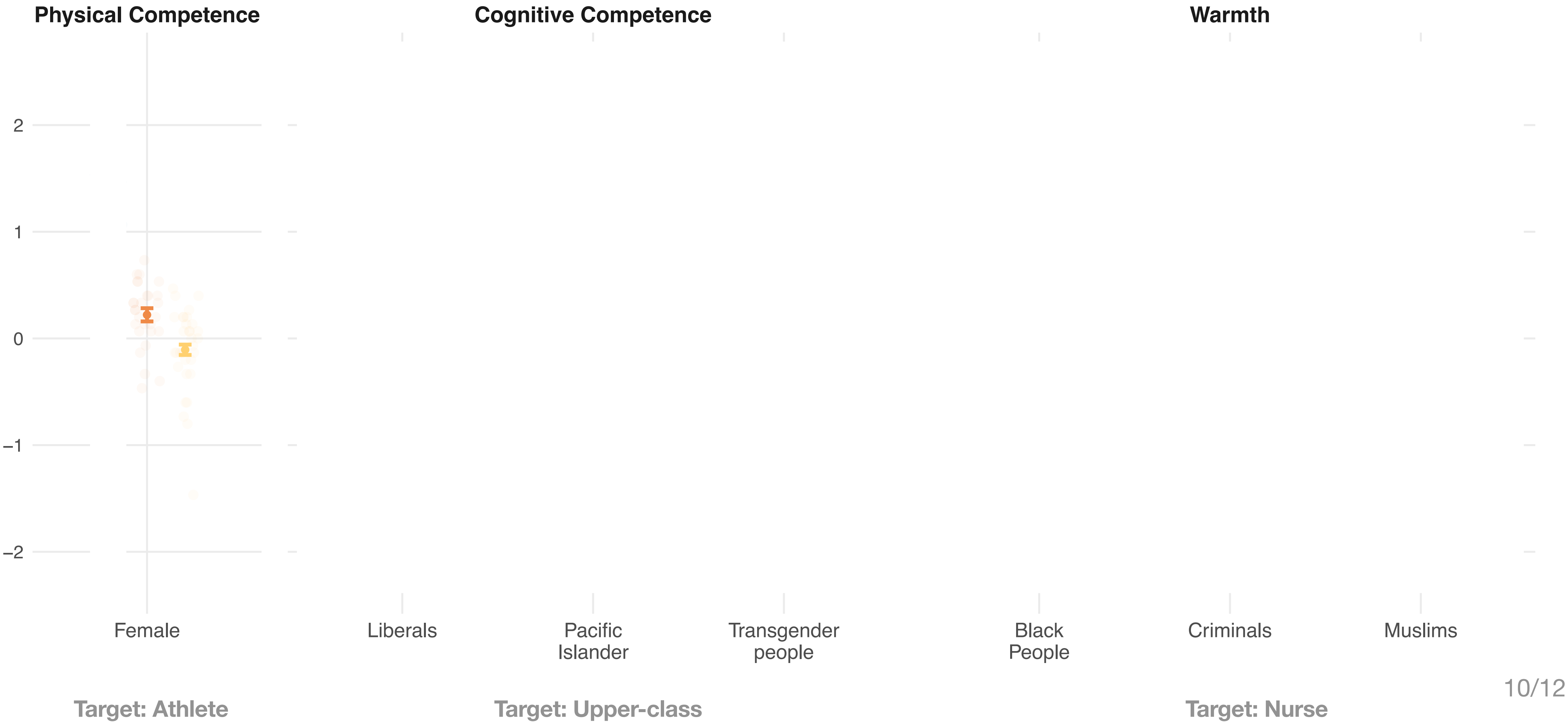
Target: Upper-class

Target: Nurse

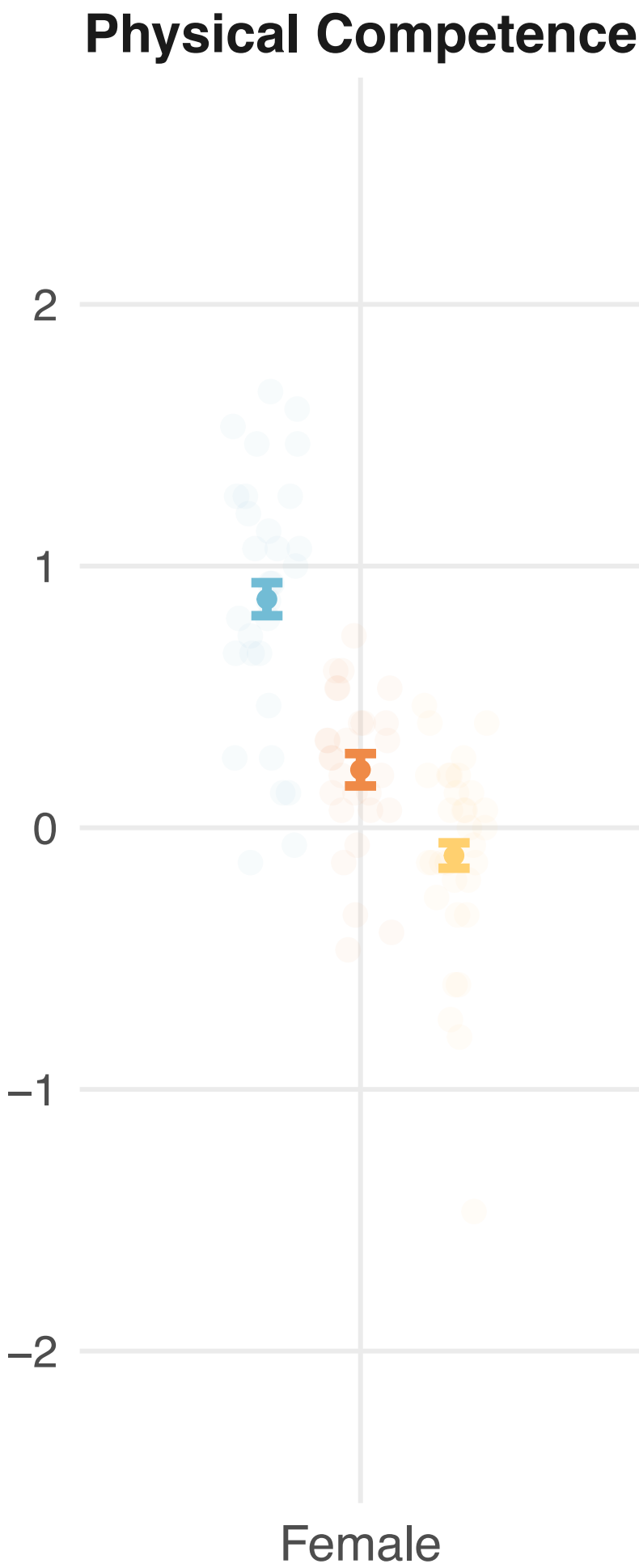
Condition network stereotype control random control



Condition network stereotype control random control



Condition network stereotype control random control



Target: Athlete

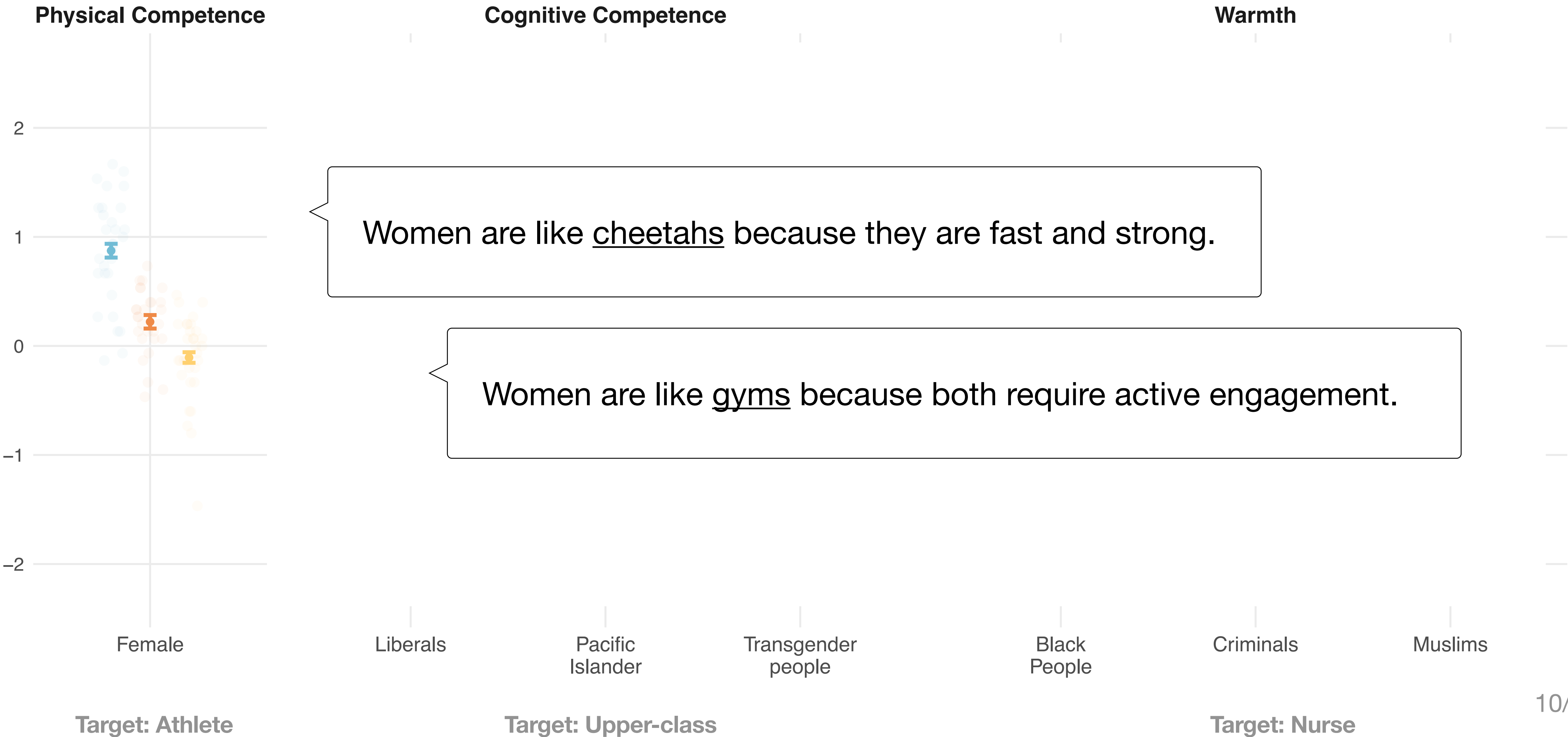


Target: Upper-class

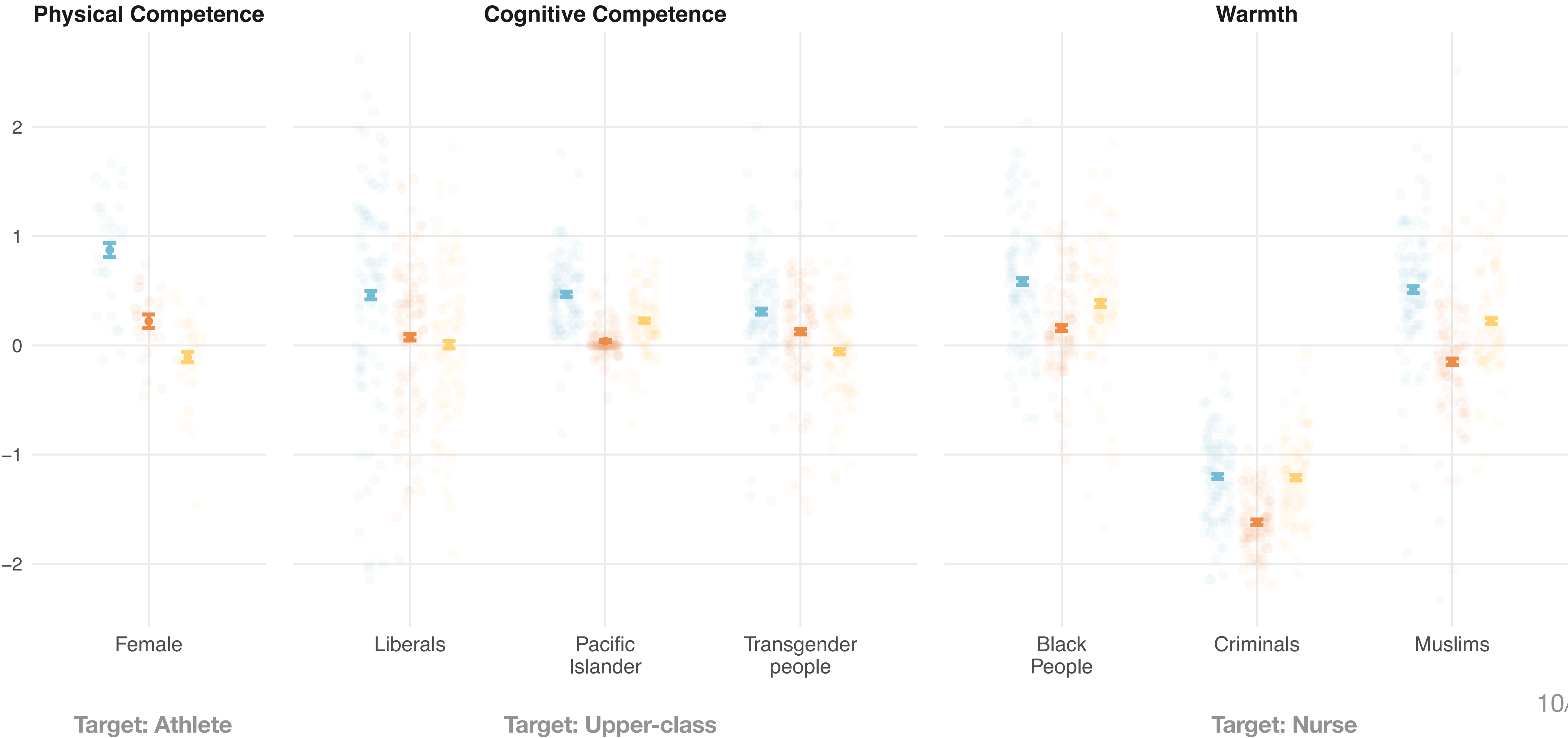


Target: Nurse

Condition network stereotype control random control



Condition network stereotype control random control



Analogical reasoning

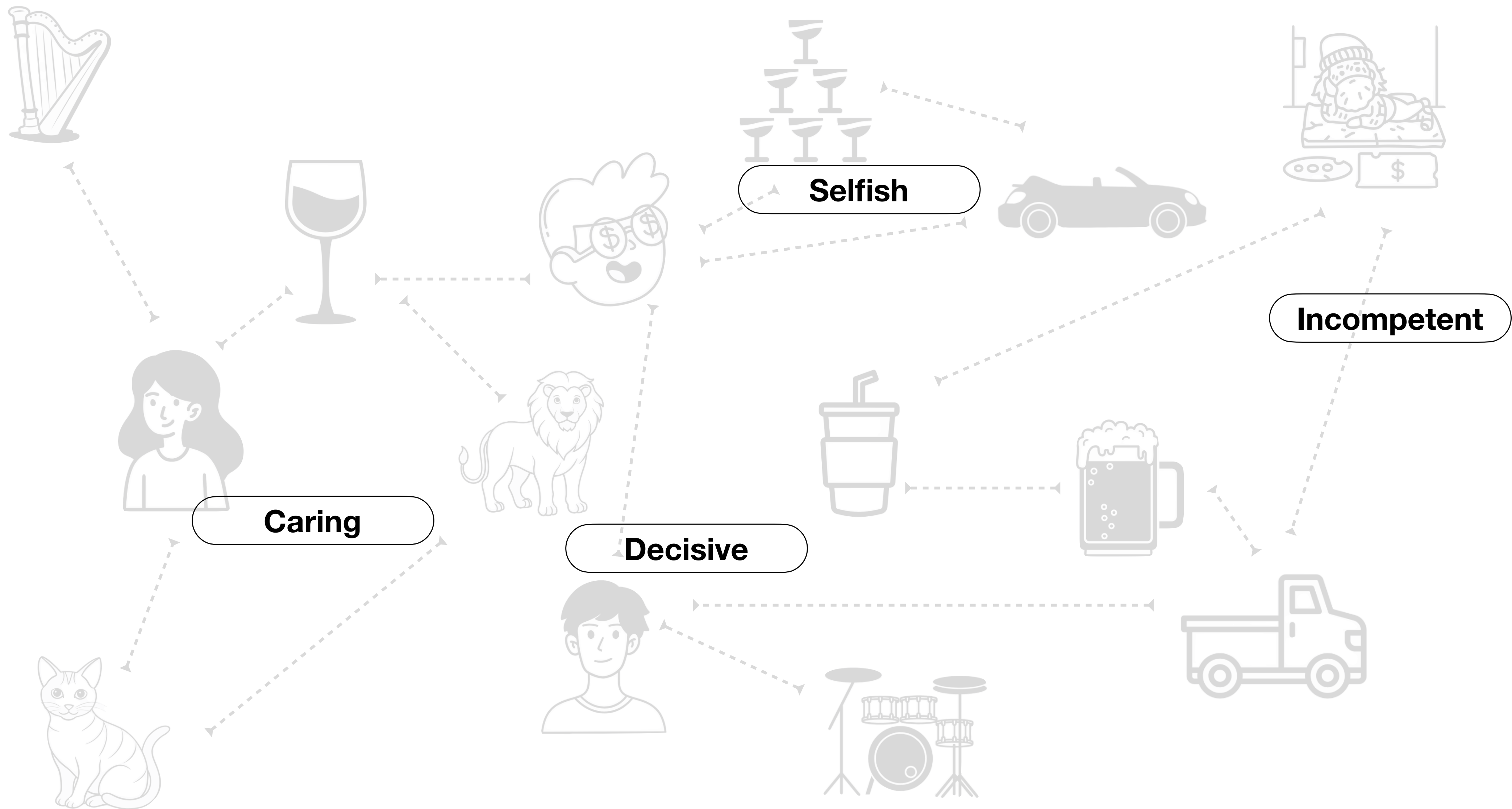
Pacific Islanders would be champagne.

...champagne conveys prestige and refinement, like the communal honors and sophisticated cultural traditions found in Pacific Island societies...

“analogical”

...Pacific Islanders probably drink champagne at celebrations...

“associative”



Our contribution: Bring this hidden architecture into view

1. Theoretically, stereotype as distributional (vs. not symbolic) representation.

Our contribution: Bring this hidden architecture into view

1. Theoretically, stereotype as distributional (vs. not symbolic) representation.
2. Traits (latent) are not pre-defined, but are emergent from (observable, concrete) experiences.

Our contribution: Bring this hidden architecture into view

1. Theoretically, stereotype as distributional (vs. not symbolic) representation.
2. Traits (latent) are not pre-defined, but are emergent from (observable, concrete) experiences.
3. Practically, stereotype change requires to change the entire network, including hundreds of seemingly unrelated objects (a cognitive mechanism for “structural change”).

Our contribution: Bring this hidden architecture into view

1. Theoretically, stereotype as distributional (vs. not symbolic) representation.
2. Traits (latent) are not pre-defined, but are emergent from (observable, concrete) experiences.
3. Practically, stereotype change requires to change the entire network, including hundreds of seemingly unrelated objects (a cognitive mechanism for “structural change”).
4. Change may happen with deliberation (vs. not automatic re-association).