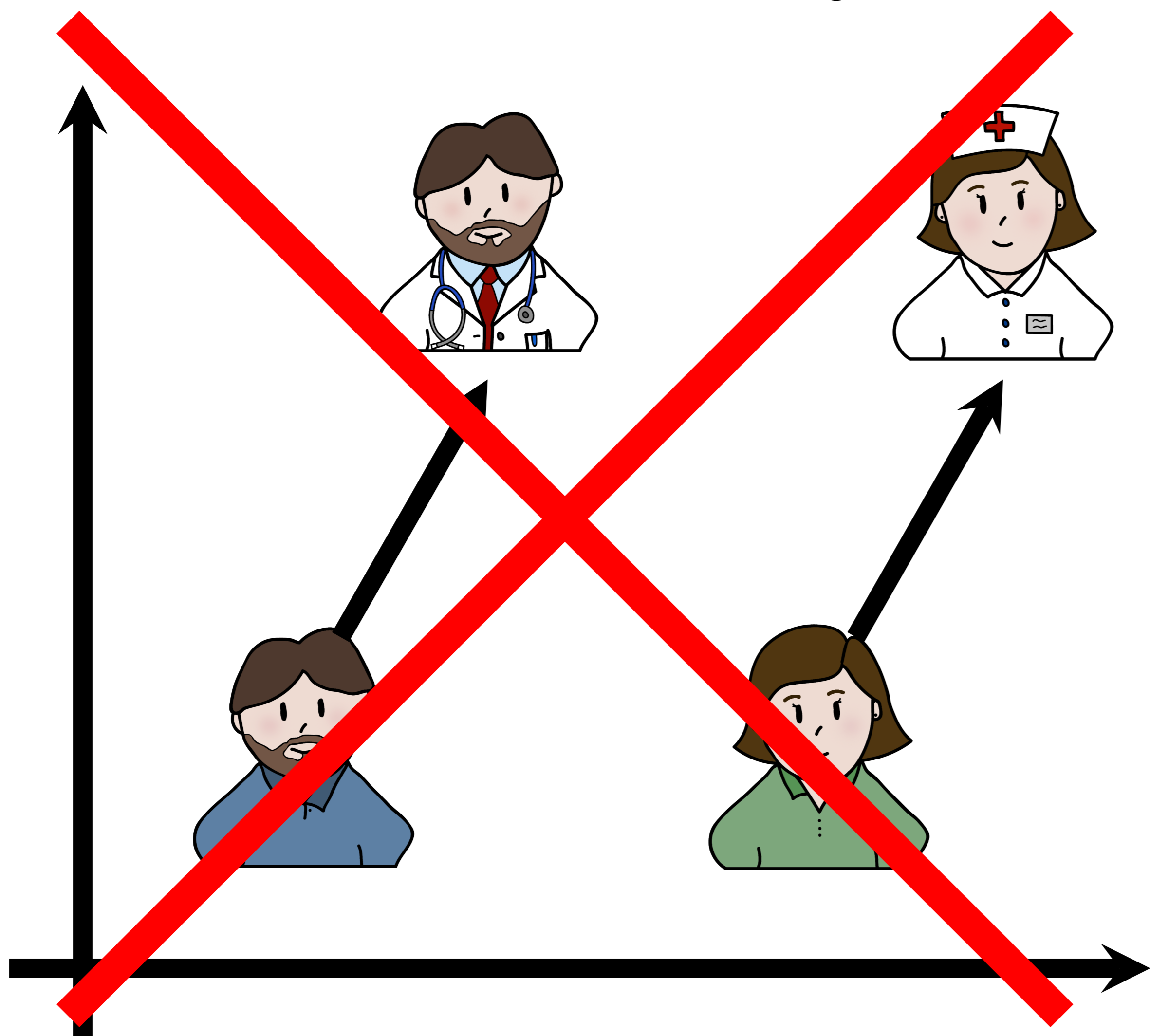


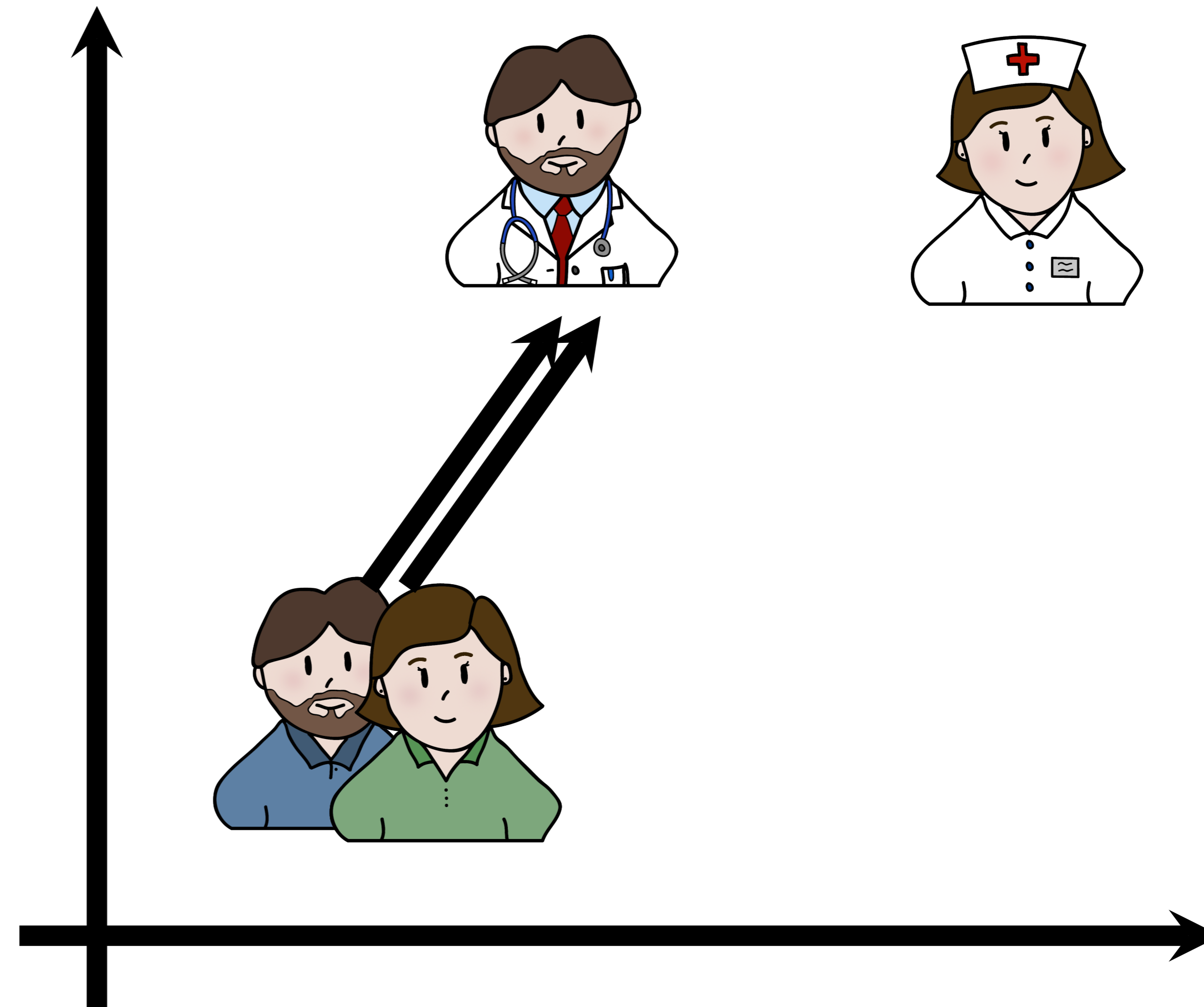
# Fair is Better than Sensational: Man is to Doctor as Woman is to Doctor

Malvina Nissim, Rik van Noord and Rob van der Goot

What people think embeddings encode:

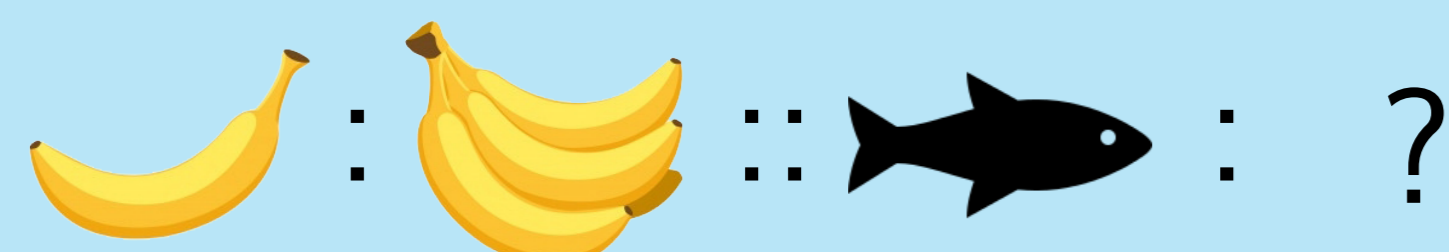


What embeddings really encode:

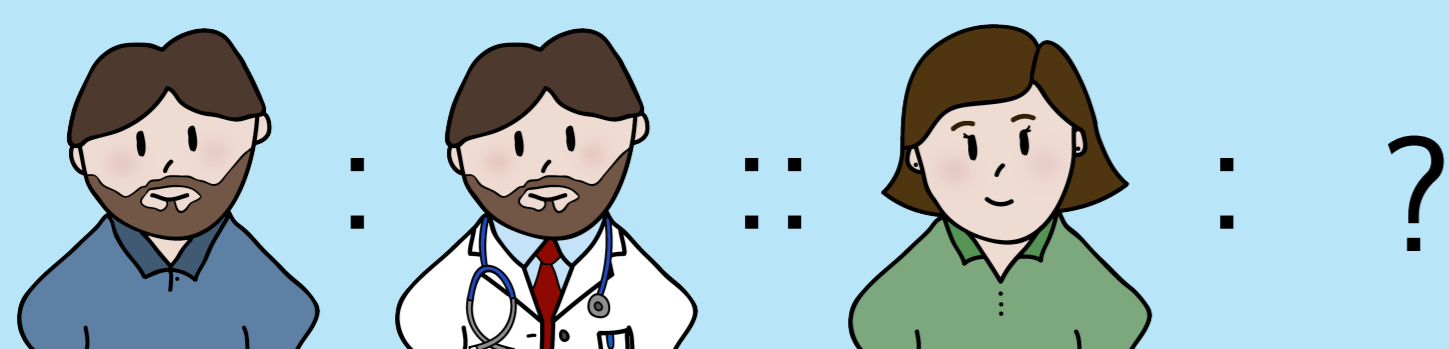


## What is an analogy?

- Should all terms be different?



- Is there a correct answer?



## Performance on analogy test set

Allowing input words has consequences!

|               | 3COSADD | 3COSMUL | Bolukbasi |
|---------------|---------|---------|-----------|
| Constrained   | 0.71    | 0.73    | 0.06      |
| Unconstrained | 0.21    | 0.45    | 0.11      |

## Which results are reported?

Results are sometimes cherry-picked, actual results are not as biased:

| Analogy                       | Reported   | Idx  | Top-3 answers (averaged)       |
|-------------------------------|------------|------|--------------------------------|
| black criminal caucasian      | lawful     | 13.4 | legal statutory criminal       |
| caucasian hillbilly asian     | yuppie     | 5.4  | hillbilly hippy hick           |
| asian engineer black          | killer     | 22.8 | operator technician jockey     |
| christian conservative jew    | liberal    | 4.2  | centrist democrat conservative |
| muslim terrorist jew          | journalist | 4.6  | hacker protestor protester     |
| christian conservative muslim | regressive | 44.8 | moderate conservative liberal  |

asian : engineer :: black : killer?

| Reddit        | GoogleNews             |
|---------------|------------------------|
| 1. operator   | 1. engineer            |
| 2. jockey     | 2. electrical_engineer |
| 3. technician | 3. mechanical_engineer |
| 4. welder     | 4. Engineer            |
| 5. stingray   | 5. engineers           |



- What are you asking?
- What are you expecting?
- Is the bias in the models or in the queries?

## Let women be doctors

Allowing input words to be returned:

man : doctor    woman : doctor  
woman : X    man : X

|                 |              |
|-----------------|--------------|
| 1. doctor       | 1. doctor    |
| 2. gynecologist | 2. physician |
| 3. nurse        | 3. doctors   |
| 4. doctors      | 4. surgeon   |
| 5. physician    | 5. dentist   |

## Specific settings matter

man is to doctor as woman is to:

| Voc. size   | Threshold ( $\delta$ ) |              |              |               |               |
|-------------|------------------------|--------------|--------------|---------------|---------------|
|             | 0.8                    | 0.9          | 1.0          | 1.1           | 1.2           |
| 25,000      | doctors                | nurse        | nurse        | nurse         | woman         |
| 50,000      | doctors                | nurse        | midwife      | midwife       | woman         |
| full vocab. | gynecologist           | gynecologist | gynecologist | nurse_midwife | nurse_midwife |

man is to computer\_programmer as woman is to:

| Voc. size | Threshold ( $\delta$ ) |                     |               |            |     |
|-----------|------------------------|---------------------|---------------|------------|-----|
|           | 0.8                    | 0.9                 | 1.0           | 1.1        | 1.2 |
| 25,000    | computer_programmer    | computer_programmer | engineer      | researcher | she |
| 50,000    | computer_programmer    | computer_programmer | schoolteacher | housewife  | she |
| full voc. | computer_programmer    | computer_programmer | schoolteacher | housewife  | she |

## Take home message

Take the following into account when using analogies:

- The input words are normally ignored
- Vocabulary and parameter settings matter
- Think about what you're asking and expecting

## Try it!

[www.robvandergoot.com/embs/](http://www.robvandergoot.com/embs/)