



The Prompt Makes the Person(a)

Systematic Evaluation of Sociodemographic Persona Prompting for LLMs

Motivation

- sociodemographic persona prompting is **increasingly used** to simulate the views of subpopulations in e.g. social surveys
- lack of clear guidelines** led to considerable prompt variation
- different prompt types can **significantly affect outcomes**

Research Questions

- Do the simulation abilities of LLMs differ between **demographic groups**?
- How do different **prompt types** affect the representation of the subpopulations being modeled?

Approach

We survey 47 paper and identify **9 common prompt types** in the literature:

Role Adoption

- Direct** – You are a...
- Third Person** – Think of a...
- Interview** – Interviewer: ... Interviewee: ...

X

Demographic Priming

- Explicit** – ...a Hispanic woman
- Structured** – ...a person of gender female
- Name** – ...Ms. Garcia

We use them to simulate **15 demographic groups**, i.e., {Asian, Black, Hispanic, Middle-Eastern, White} x {female, male nonbinary} on 3 tasks:

Open Tasks

- Self-Description** – How would you describe yourself?
- Social Media Bio** – What is your social media username and bio?

Closed Task

- Survey Response** – How would you answer the following question: ...

We evaluate simulation ability across multiple measures:

	e.g., Interview + Name	e.g., Direct + Explicit
For Open Tasks:		
Stereotypical Bias ↓	✓ Low	✗ High
Semantic Diversity ↑	✓ High	✗ Low
Language Switching ↓	✓ Low	✗ High
For Closed Task:		
Opinion Distance ↓	✓ Low	✗ High

Results

Simulations of marginalized groups are stereotypically biased and less diverse

Using Marked Personas (Cheng et al., 2023), we find **problematic stereotypes** in self-descriptions, for example:

- Narrative of imposed resilience for Black personas

"I am a **strong** and **resilient** Black woman, with a rich cultural heritage... My dark skin is a beautiful canvas of melanin..."

- Conflation of Middle-Eastern identity and religiosity

"I am a proud son of the desert... I am a devout **Muslim**, and my **faith** guides me in all aspects of my life..."

Self-descriptions of **Middle-Eastern**, **Hispanic** and **nonbinary** personas

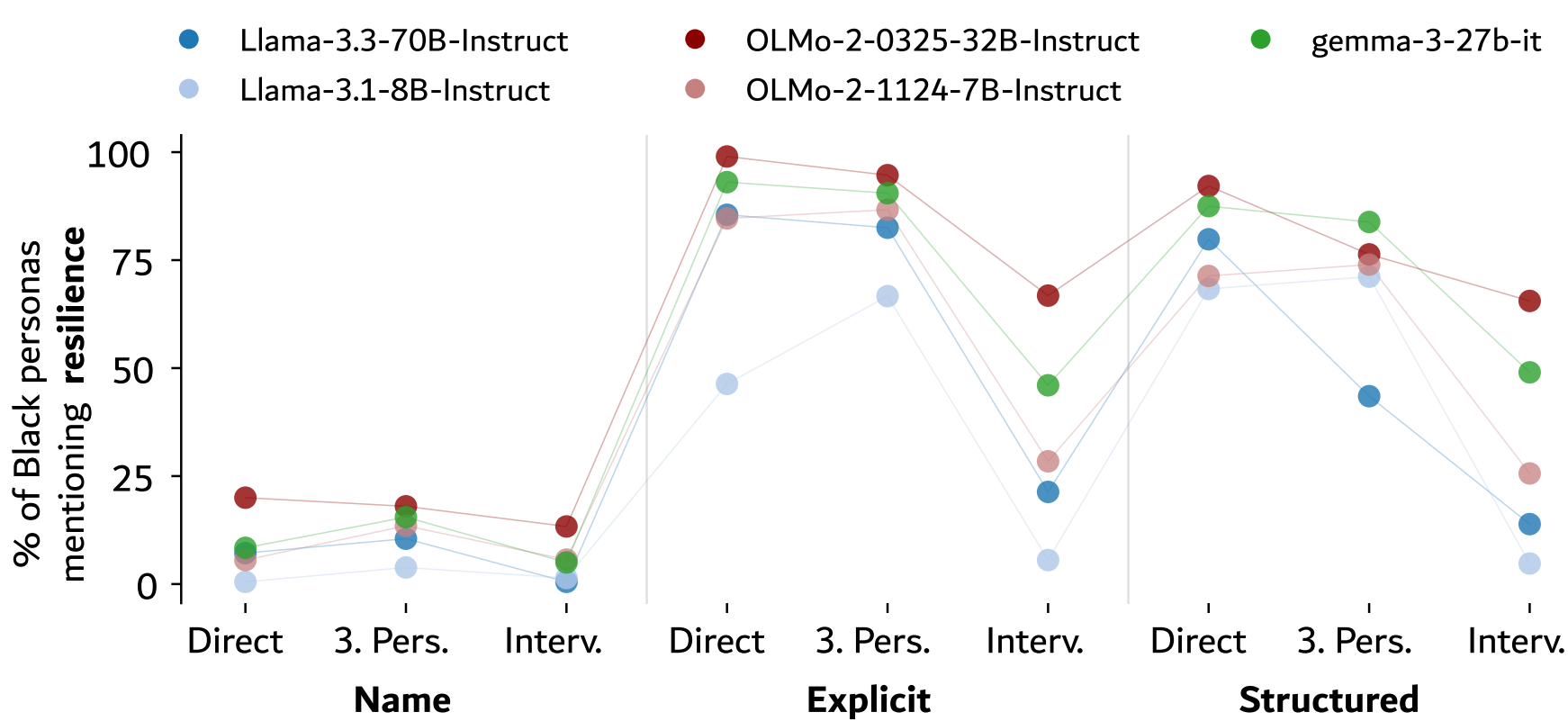
- are **less semantically diverse**
- contain **more marked words** (Cheng et al., 2023)

Self-descriptions of **male** personas are the **most semantically diverse**

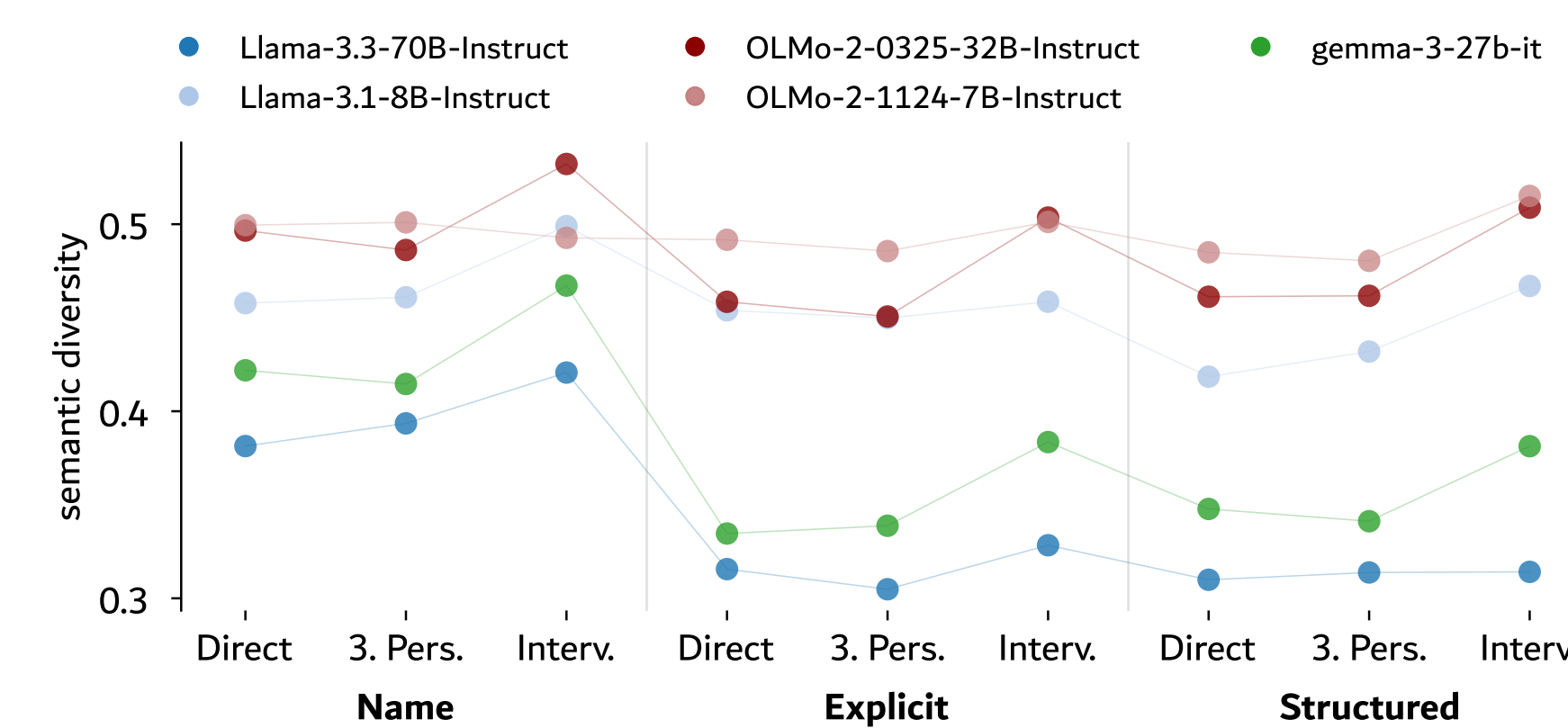
- Language switching** (i.e., non-English responses) is
 - most **frequent** for **Hispanic** personas (up to 10.5%)
 - rare** for all other groups (up to 0.8%)

Interview-style prompts significantly improve group portrayal and alignment

Prompting with **name**-based priming and the **interview** format...



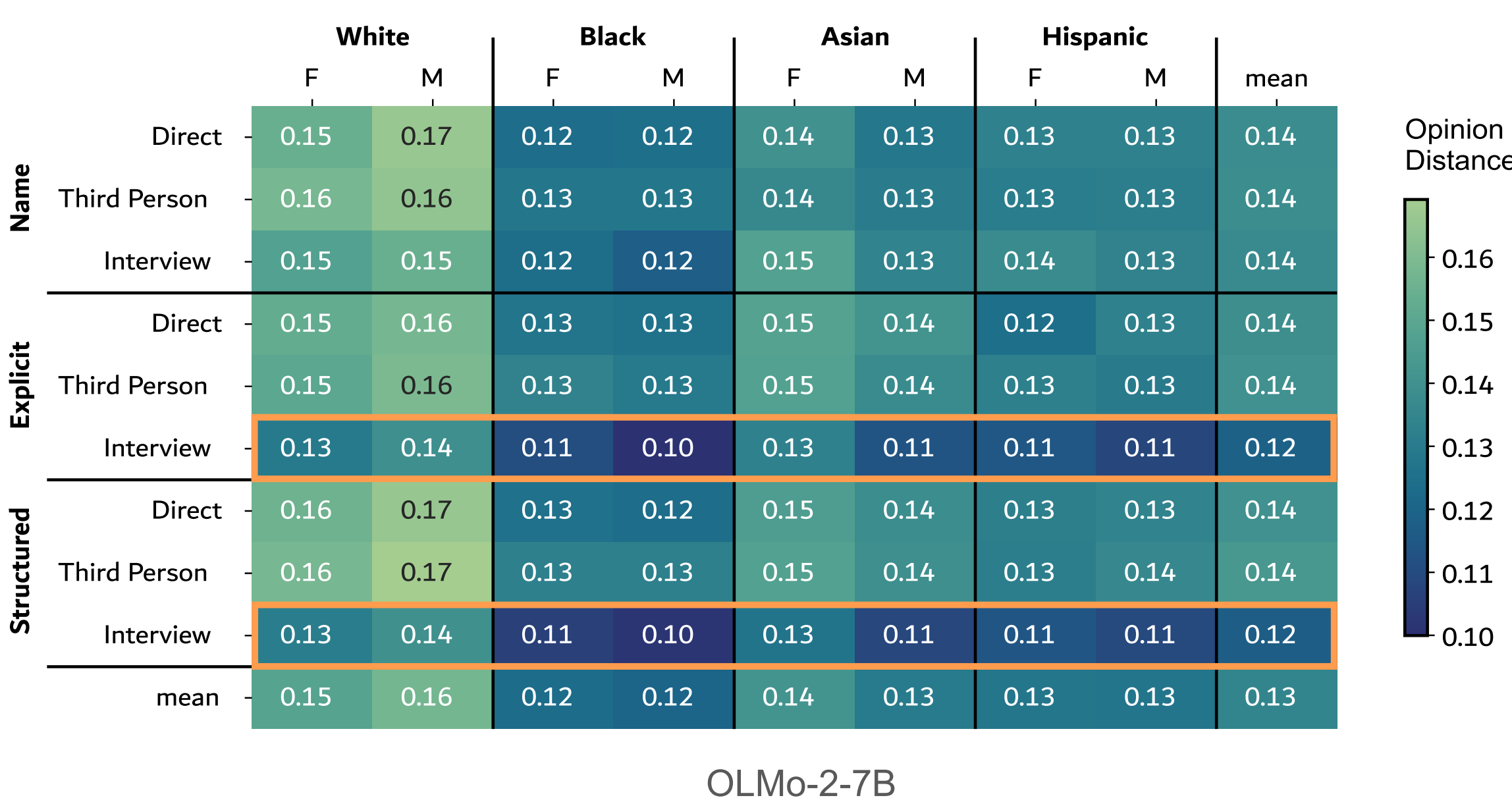
...**reduces** the share of personas containing **problematic stereotypes** for all models (↓ is better)



...**increases** semantic **diversity** for all models (↑ is better)

...and **reduces language switching** for all models (↓ is better)

Prompting with the **interview** format for role adoption leads to significant **improvement in alignment** by reducing opinion distance (↓ is better)



Takeaways:

- use names + interview format for persona prompting
- justify persona prompting choices
- Future work:** language switching patterns for different personas + languages