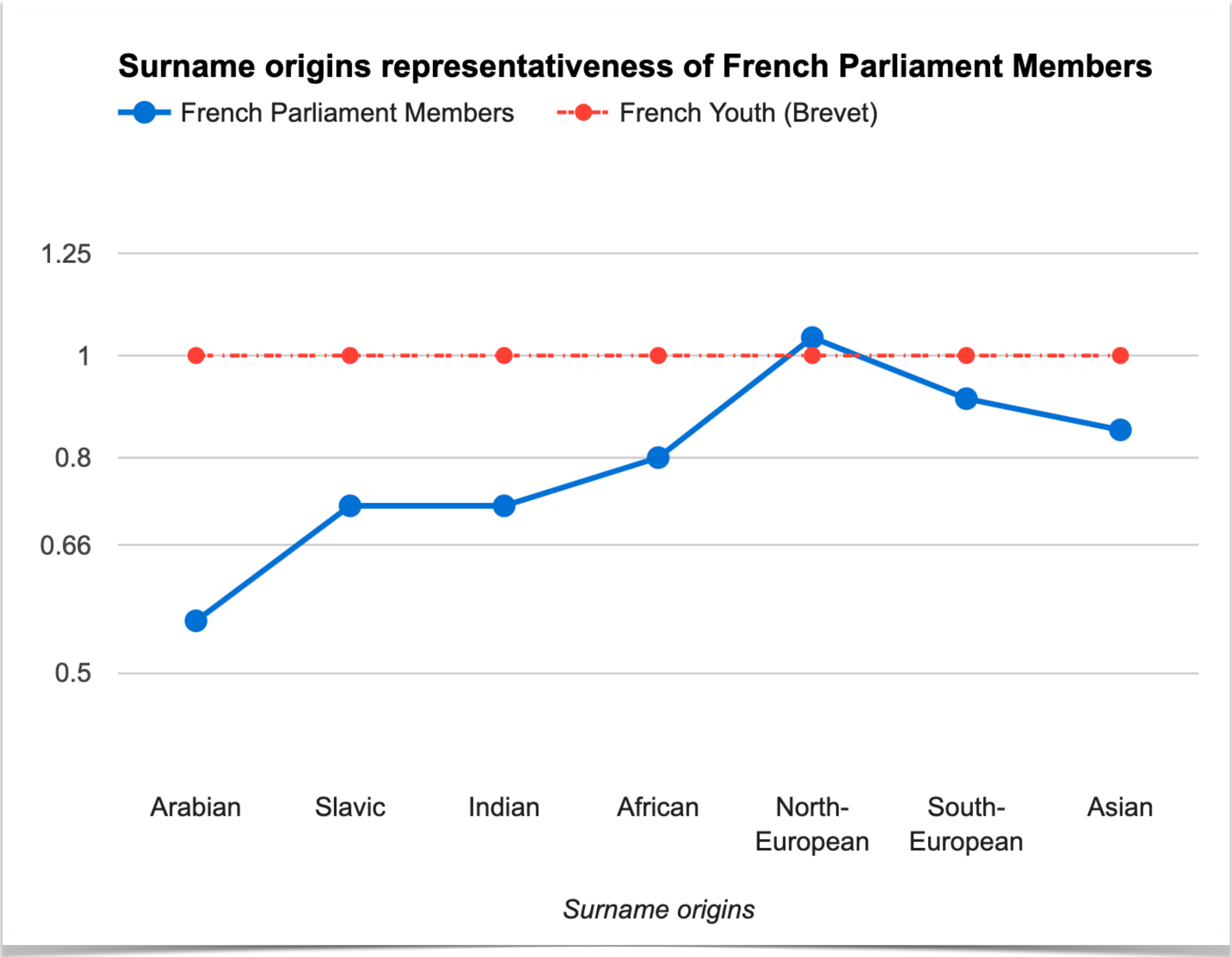


Une approche computationnelle à l'étude de la représentativité de genre dans les films populaires

Antoine Mazieres

Journée d'étude « Quantification des discriminations », 20 novembre 2020.

Intelligence artificielle et étude des discriminations ?



<https://namograph.antonomase.fr/>

Mazieres, A., & Roth, C. (2018).
**Large-scale diversity estimation through
surname origin inference.**
Bulletin of Sociological Methodology,
139(1), 59-73.

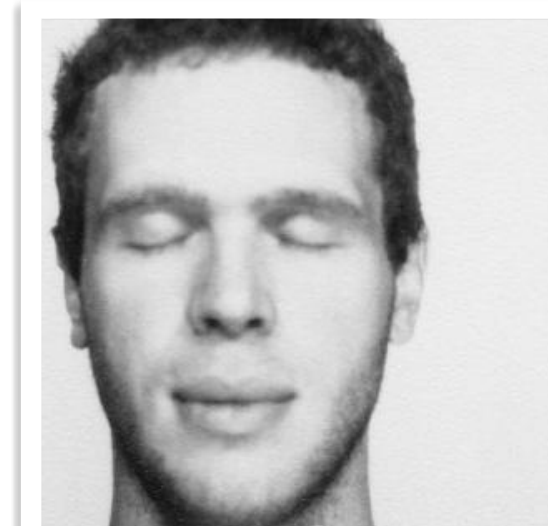
Représentativité de genre dans le cinéma et analyse d'image



Antoine Mazieres



Telmo Menezes



Camille Roth

Équipe de sciences sociales computationnelles
Centre Marc Bloch (CNRS / U. Humboldt)
Berlin

Soumis à *Humanities and Social Sciences (Nature)*

<https://arxiv.org/abs/2009.09067>

Le genre dans les médias de masse

- Longue tradition d'analyse des représentations de genre dans les médias depuis les années 1940.
- Méthodologies très changeantes : peu d'études longitudinales

Variables riches, sémantiques
Faible nombre d'observations

Vs.

Variables pauvres
Nombreuses observations

Définir la base de données

- (L'état de l'art utilise souvent le Box Office (\$) pour définir la popularité)
- Films mis à disposition par YIFY (communauté **Torrent**) avec au moins 3 *seeders* : ~13k films
- ... Et avec un certain nombre de métadonnées sur **IMDb** : Budget (med: 23m\$), Box Office (med: 43m\$), note utilisateur, niveau de censure, durée, genre (romance, crime, ...), année de sortie, etc.
- Total : **3776 films**
- **1985-2019**, minimum 50 films par an.

Extraction des données, détection des visages et inférence du genre

- Extraction de 1 image / 2 secondes : ~12,4m images.
- Modèles utilisés : *Mathematica 12*.
- 6,6m d'images avec visages détectés (~2,5k par film)
- ~ 35% femmes (34.52%, $\sigma = 9.19$)

~ 25%
La Revanche des Siths,
Pirates des caraibes,
Matrix Reloaded

~ 50%
The Hunger Games,
Rogue One,
Gravity

~ 65%
Bad Moms,
Sisters,
Cake

Évaluation et corrections des inférences



Does the **face** appearing inside the red frame **represent a woman or a man**?

☒ A woman ☐ A man ☐ I don't know ☐ There is no face inside the red frame !

Do you see **other faces** that are clearly identifiable outside the red frame?

☐ Yes ☒ No ☐ I don't know

(a) Face detection

		Humans	
		<i>Positive</i>	<i>Negative</i>
Model	<i>Positive</i>	977	23
	<i>Negative</i>	137	963

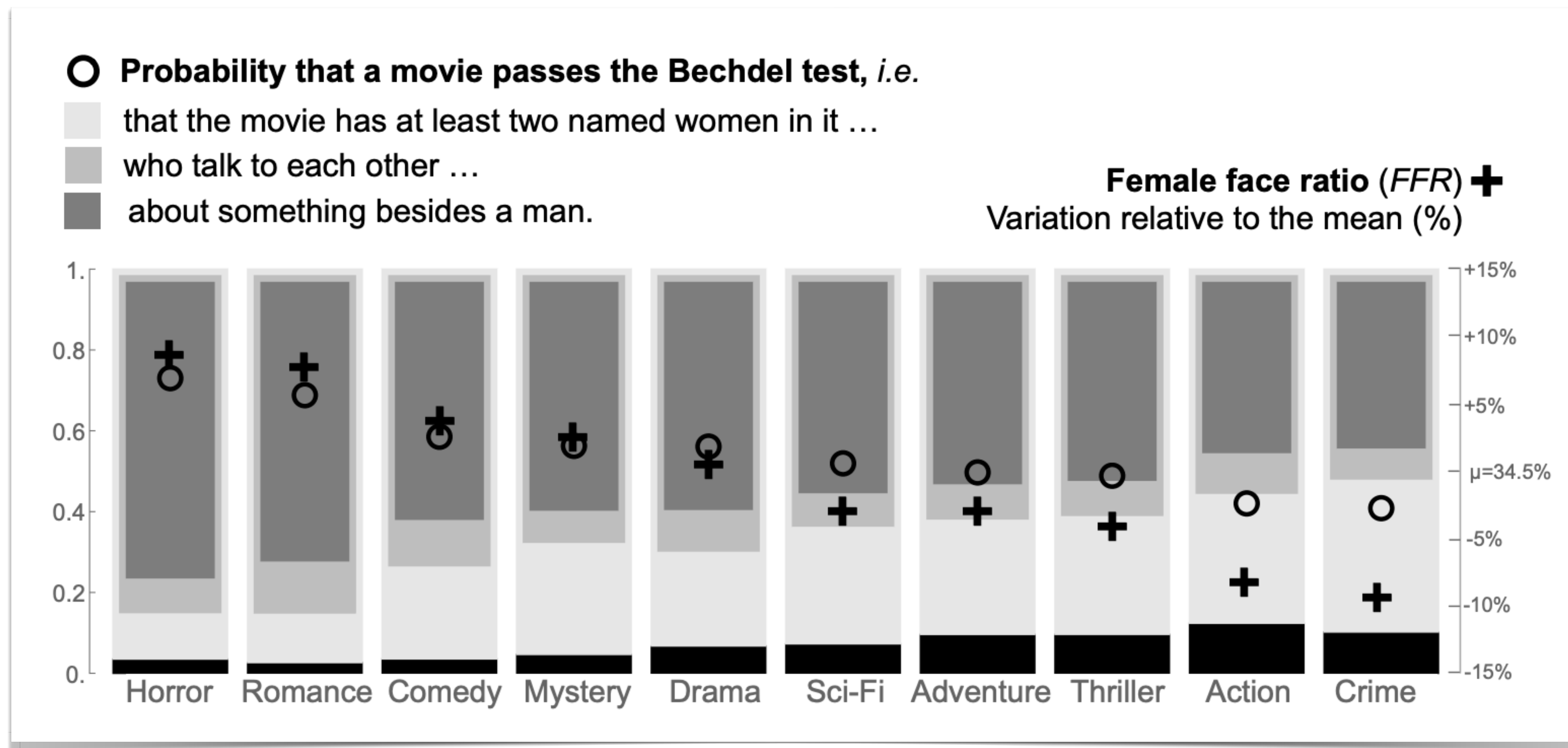
(b) Gender inference

		Humans			
		<i>Female</i>	<i>Male</i>	<i>Doubt</i>	<i>No face</i>
Model	<i>Female</i>	304	162	18	16
	<i>Male</i>	75	410	8	7

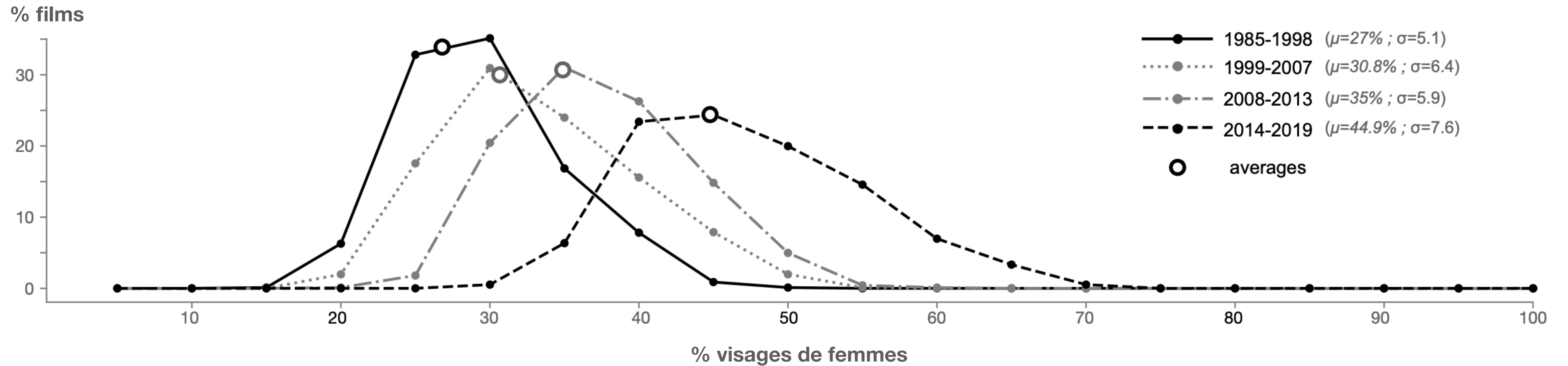
Performances et Biais :

- Détection des visages : 92%
- Inférence du genre : 74%
- Hommes mieux reconnus
- Images récentes mieux traitées

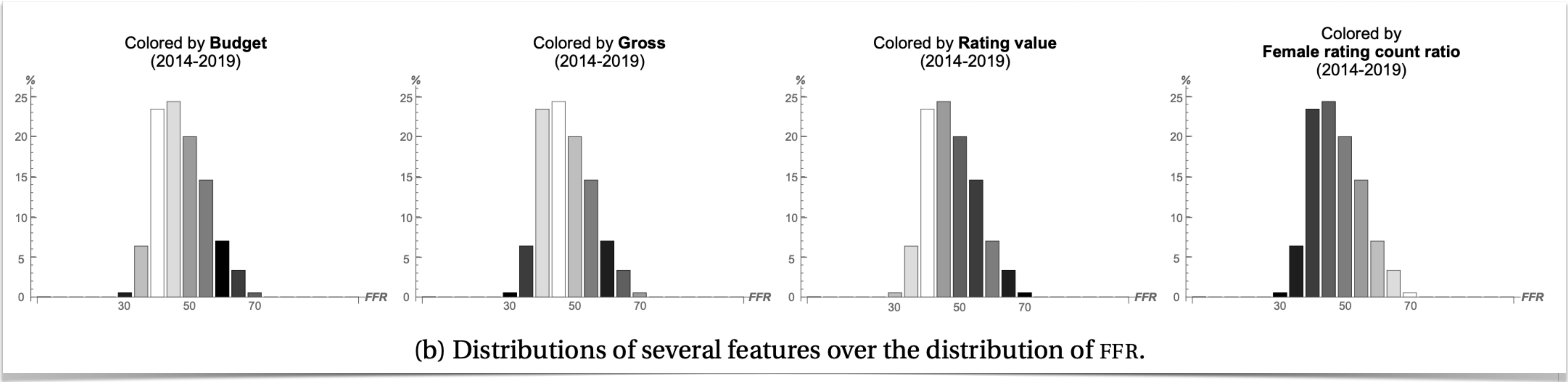
Comparaison avec le Bechdel Test



Analyse longitudinale



Distributions de plusieurs métadonnées



Biais de genre dans la mise-en-cadre

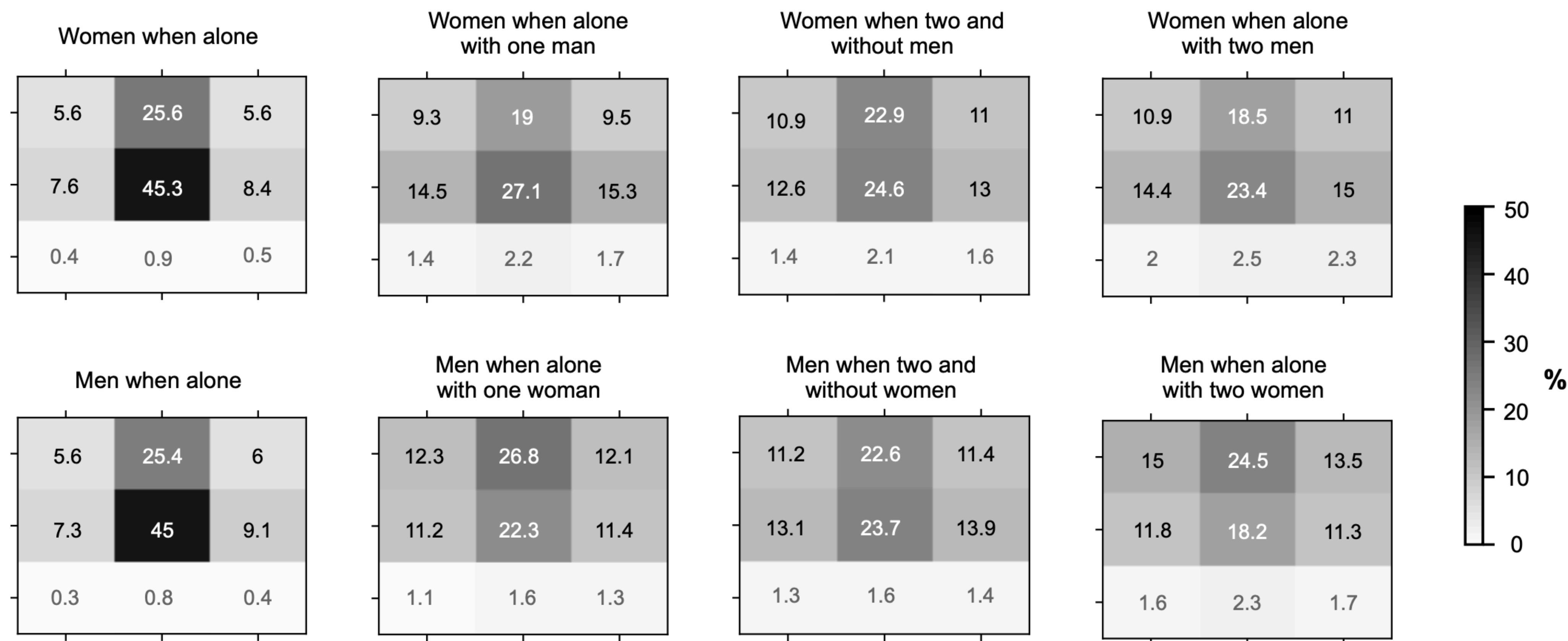


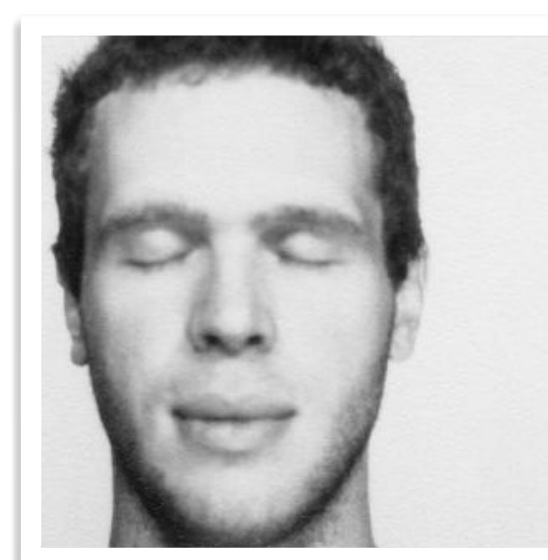
Figure 6: **Distribution of faces position on-screen** (2014-2019).



Antoine Mazieres



Telmo Menezes



Camille Roth

Équipe de sciences sociales computationnelles
Centre Marc Bloch (CNRS / U. Humboldt)
Berlin

antoine.mazieres@gmail.com

@mazieres

<https://antonomase.fr>