

MyBubble: Influence of Algorithms in Users' Filter Bubbles

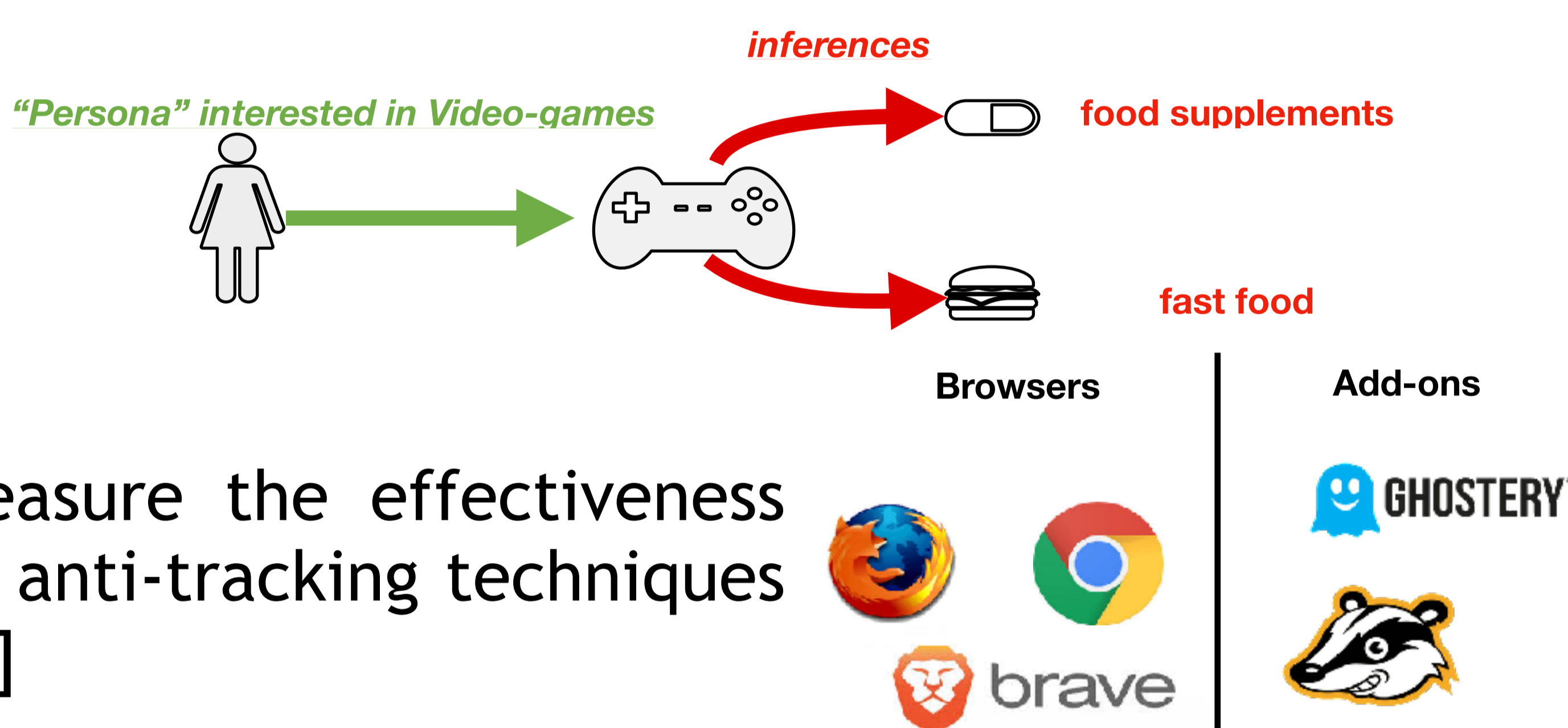
Pelayo Vallina*†, Antonio Fernández Anta*, Rubén Cuevas†, Esteban Moro†‡
IMDEA Networks Institute *, Universidad Carlos III de Madrid†, MIT Media Lab.‡

User's Personalisation

- Online services have the capacity of learning preferences of online users and personalize the content that they offer them using:
 - Tracking techniques
 - Interaction of the users with the services
- However, this personalization has an impact on the final users:
 - Influence the content that each user receives
 - May lead to discrimination (e.g., by gender, race, or geolocalization [1,2])

Goals

- Understand how online services profile users
- Identify biases in the online advertising ecosystem

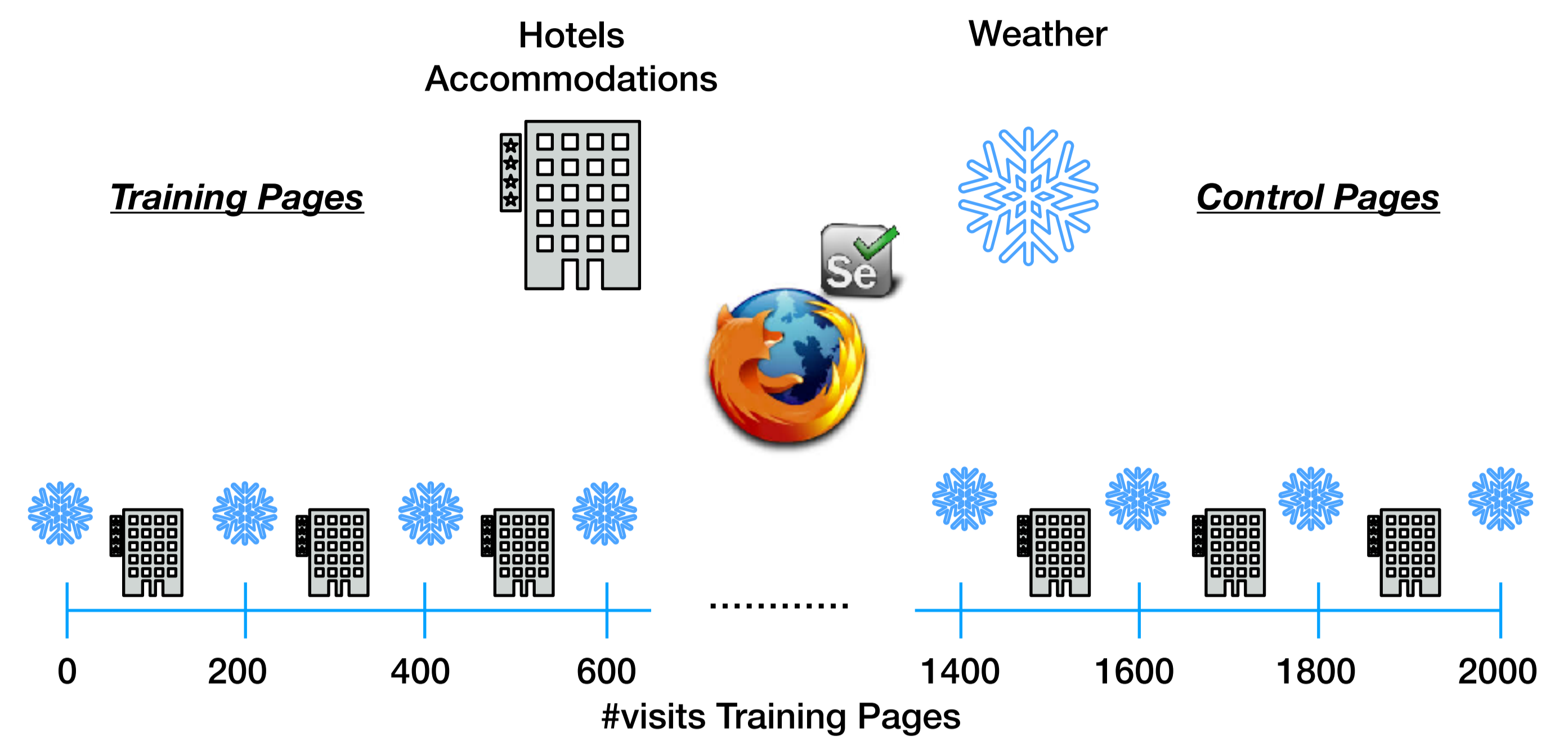


- Measure the effectiveness of anti-tracking techniques [4]

Creating Personas

- Bots with a strong behavior in a certain category or a combination of them, called "Personas"[3]
- Using the categorization of AdWords, obtain the 10 most relevant website on each category
- Individual analysis of each website using third-party services in order to measure the level of noise in our "Personas". From each domain we collect the categories associated to them using **VirusTotal**, **McAfee** and **AdWords**
- Measure how the biggest player in the online advertising ecosystem categorize our Personas
 - Open a Google session
 - Visit the 10 websites that belong to each "Persona"
 - Visit Google AdSettings and collect the categories associated to the account
 - Clean the historical information and close the session
 - Repeat the process with another "Persona"

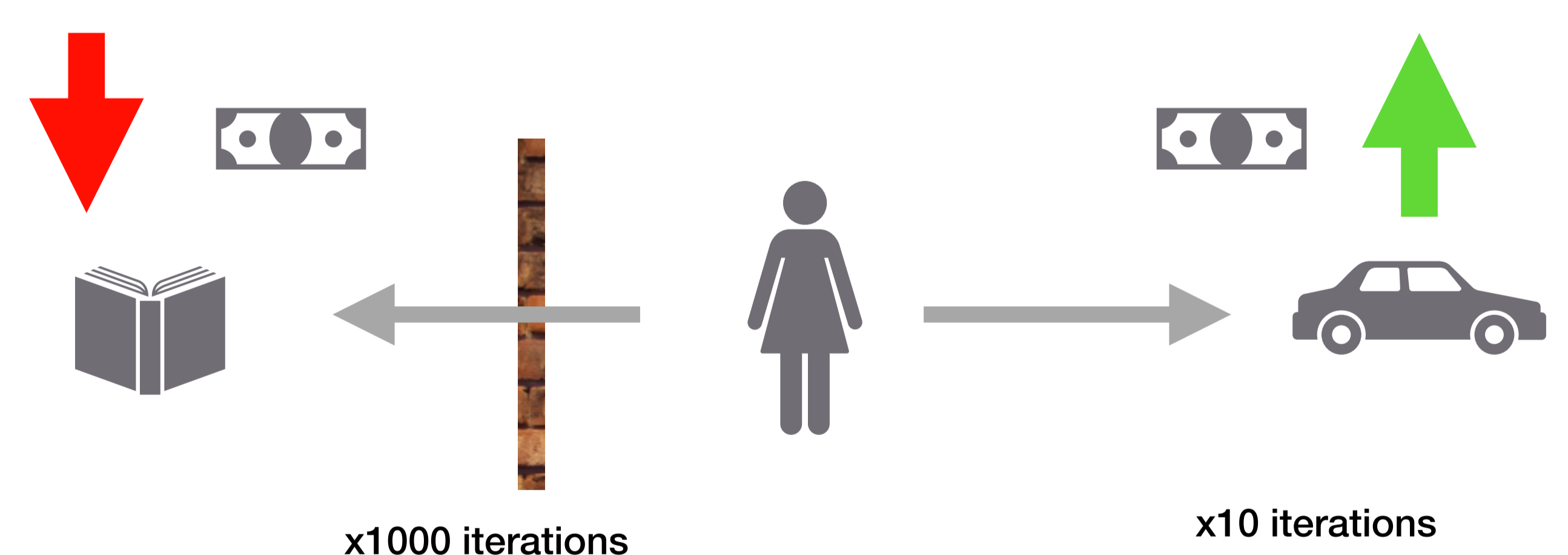
Experimental Methodology



1. Create a new and clean Firefox session using Selenium WebDriver
2. Select the behavior that we want to provide to our "Persona"
3. Visit all the "Training Pages", the web pages that will give the behavior to our "Persona"
4. Every 200 iterations visit the "Control Pages" and collect the ads that appear on each website
5. Repeat the process using other "Persona"

Example Use Case

- Measure how easy/hard is to transit to different categories with a different cost for advertisers



Acknowledgment

Partially supported by the Regional Government of Madrid (CM) grant Cloud4BigData (S2013/ICE-2894) cofounded by FSE & FEDER, Spanish Ministry of Science, Innovation, and Universities grant DiSCoEdge (TIN2017-88749-R), and "la Caixa" Foundation MISTI Global Seed Funds grant MyBubble.

Reference

- [1] Hannák, Anikó, et al. Bias in Online Freelance Marketplaces: Evidence from TaskRabbit and Fiverr. En CSCW. 2017. p. 1914-1933.
- [2] Hannák, Anikó, et al. Measuring price discrimination and steering on e-commerce web sites. IMC'14.p. 305-318.
- [3] Carrascosa, Juan Miguel, et al. I always feel like somebody's watching me: measuring online behavioural advertising. CoNext'15.
- [4] Muhammad Ahmad Bashir , et. al. How Tracking Companies Circumvent Ad Blockers Using WebSockets, ConPro 2018