



HAL
open science

How do our demographics shape our trajectories ?

Hippolyte Dubois, Patrick Le Callet, Antoine Coutrot

► **To cite this version:**

Hippolyte Dubois, Patrick Le Callet, Antoine Coutrot. How do our demographics shape our trajectories?. International Conference on Spatial Cognition, Sep 2021, Rome, Italy. Proceedings of the 8th International Conference on Spatial Cognition. hal-03379165

HAL Id: hal-03379165

<https://hal.science/hal-03379165v1>

Submitted on 14 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

How do our demographics shape our trajectories ?

Hippolyte Dubois, Patrick Le Callet, Antoine Coutrot

Université de Nantes, CNRS, LS2N, F-44000 Nantes, France

Introduction

Trajectories are complex signals produced by individuals when confronted with a constrained environment in which they have to complete a task. As such, they are traces of the cognitive processes involved in space-integration, strategy-planning and continuous actualisation. Demographics, among other factors, have been shown to have an influence on those cognitive processes, influence that should therefore be reflected in the resulting trajectories.

To analyse this phenomenon and better understand how demographics and degenerative diseases shape our trajectories, the Sea Hero Quest project has collected the trajectories of over 4 million players in the virtual environments of the eponym phone and tablet game[2].

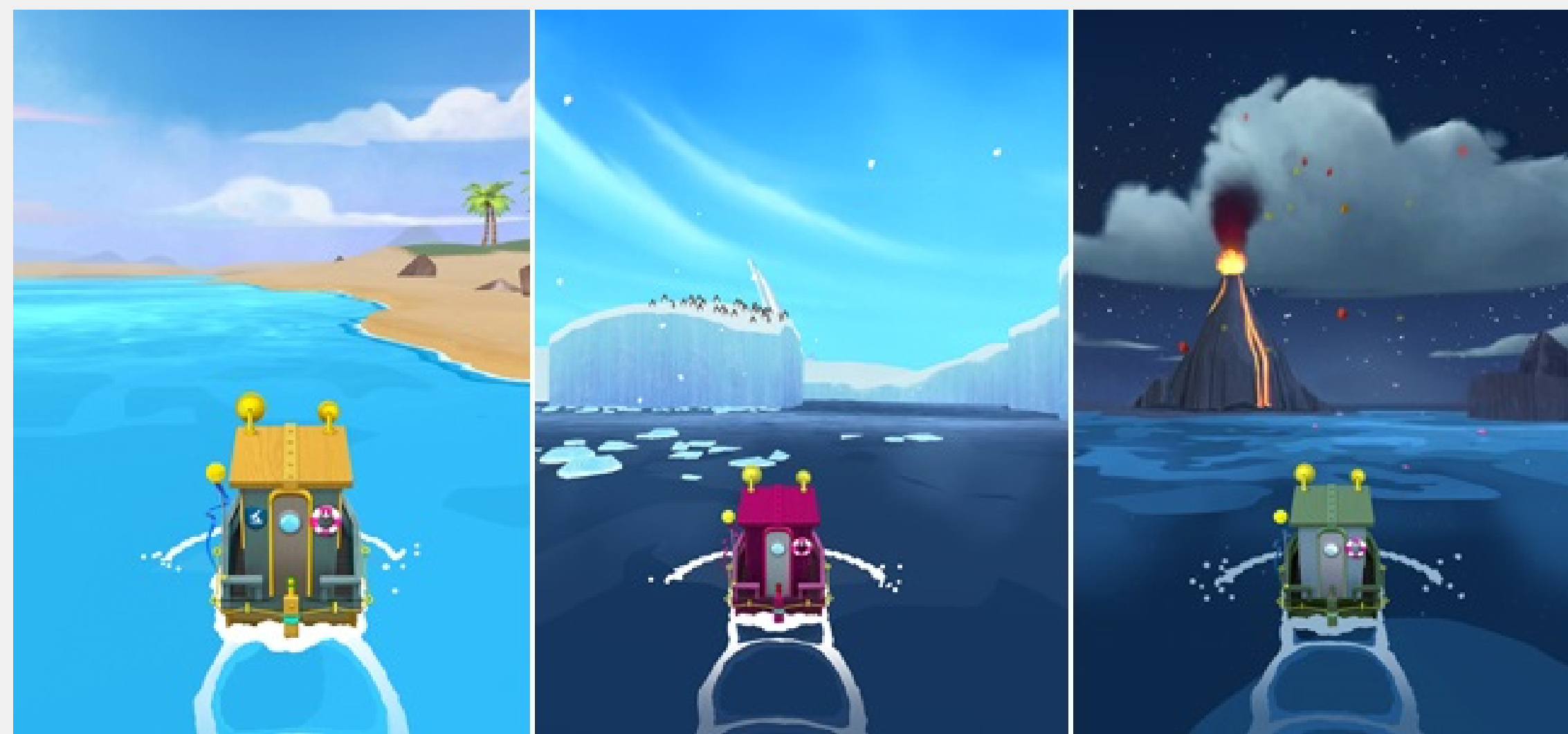


Figure 1. Screenshots from Sea Hero Quest

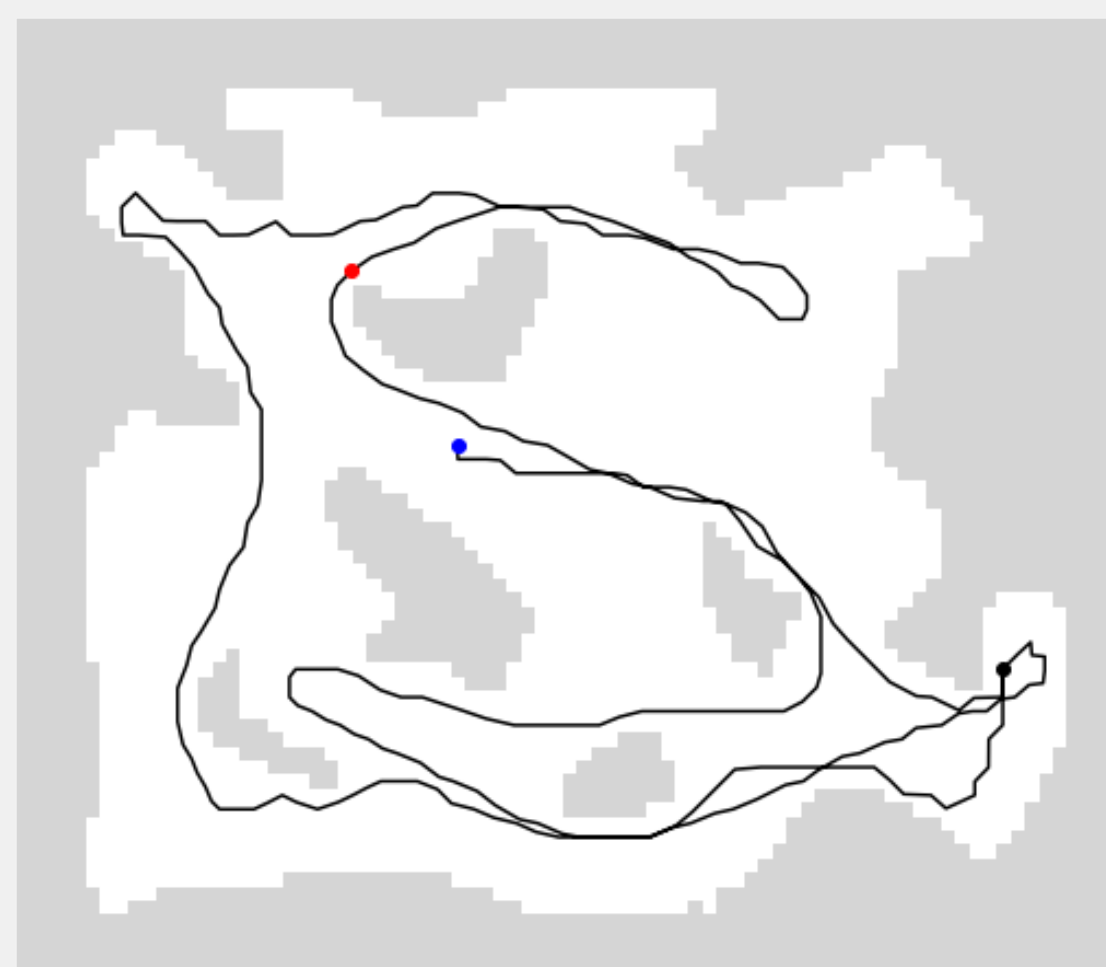


Figure 2. Trajectory example

Aims

In this study, we aim at designing a set of trajectory metrics that captures how demographics are associated with spatial navigation behaviour. In this context, a good metric allows to differentiate trajectories and can be predicted from the demographics of the person that generated this trajectory.

Dataset

The SHQ dataset is comprised of two parts.

1. **Trajectories**, timeseries of points regularly sampled in $\{x, y\} \in \mathbb{N}^2$ and then resampled to a length of 300 points sampled in \mathbb{R}^2 . A trajectory belongs to a specific level and is produced by a specific player. We augment this signal by adding four additional dimensions, namely $\sin(\theta)$, $\cos(\theta)$, $\sin(d\theta)$, $\cos(d\theta)$, to capture the orientation of the trajectory at each point. We later refer to a single trajectory as t .
2. **Demographics**, vector of self-declaratory characteristics:
 - Age
 - Gender
 - Country
 - Education level
 - Dominant hand
 - Neighbourhood type
 - Orientation skills self-assessment
 - Hours of sleep per night
 - Commute time per day

In this experiment we drop the country and only keep the trajectories of players who filled every fields. We later refer to a single demographic vector as u .

In this experiment we use 23,095 trajectories and associated demographic vectors.

Method

Because they are timeseries made of 300 samples of 6 dimensions each, we can't just throw the trajectory in our model and hope for it to be able to predict it from the demographics. Instead, we compute several metrics, its length and its likelihoods given all other trajectories (later referred to as entropy). Likelihoods are computed in x, y, θ , and $d\theta$. Each trajectory is therefore represented by a vector of dimension 4.

Using this vector we label each trajectory as normal or outlier using IsolationForest algorithm. We then train a model f to predict this label from the demographic vector.

$$l = \text{IsolationForest}(\text{metrics}(t)) \quad (1)$$

$$f : u \rightarrow l \quad (2)$$

We can then look at the model's learned parameters to analyse which demographic features contribute the most to the prediction, using *permutation importance*[1].

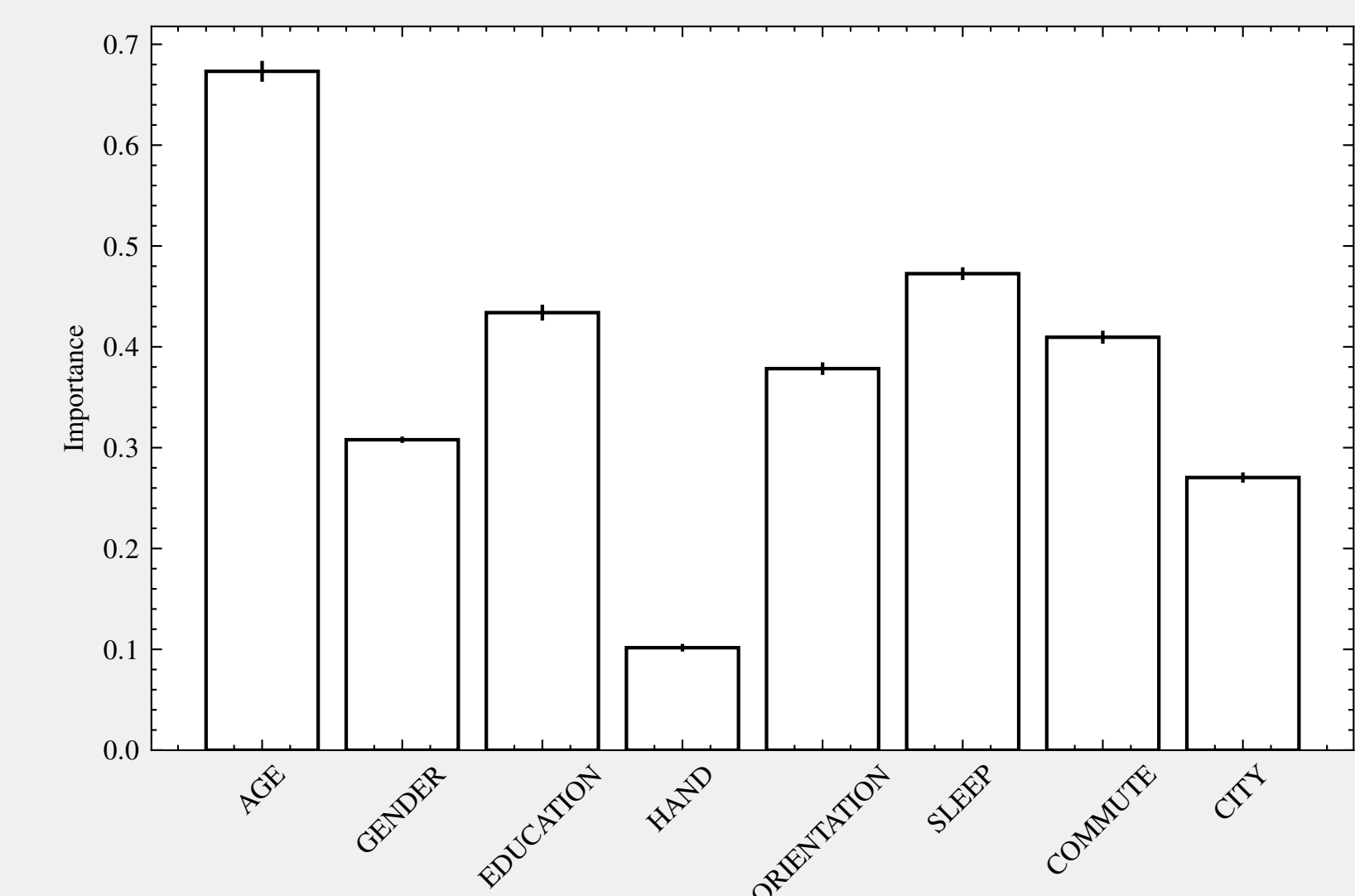
Evaluation

By construction, the 2 classes defined by the IsolationForest algorithm are highly imbalanced. To compensate for this imbalance, we trained our model on synthetic data (SmoteTOMEK algorithm) in addition to the original data, increasing the number of samples from 23,095 to 42,478. The evaluation is only based on the original data.

To be able to compare different parametrization of the IsolationForest algorithm which results in different class balance, we score the model's performance using Cohen's Kappa, as it is robust to data imbalance.

Results

Classifier	Mean Cohen's Kappa	95% CI
SVM RBF	0.092	0.045
Gradient Boosting	0.112	0.009
Random Forest	0.667	0.101



Discussion

Having to rely on synthetic data to overcome the strong data-imbalance caused by the outlier detection framework isn't a good thing, especially when dealing with categorical features in the demographics. This requires investigating how to design a better labeling algorithm for this task.

Nonetheless, we were able to show that several demographic characteristics have an impact on spatial behaviour (mainly the age), while other have seemingly no impact (namely the handedness).

References

- [1] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [2] Antoine Coutrot, Ricardo Silva, Ed Manley, Will de Cothi, Saber Sami, Véronique D. Bohbot, Jan M. Wiener, Christoph Hölscher, Ruth C. Dalton, Michael Hornberger, and Hugo J. Spiers. Global determinants of navigation ability. *Current Biology*, 28(17):2861–2866.e4, September 2018.