# A User Study of Perceived Carbon Footprint

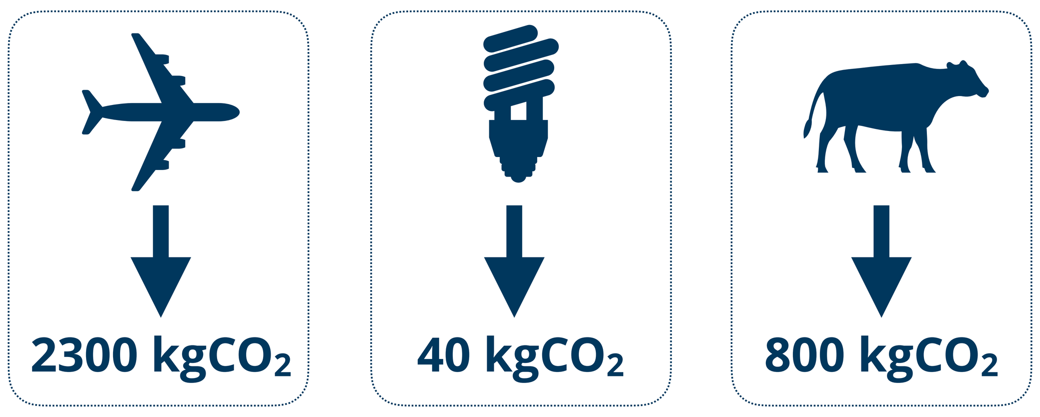
**Victor Kristof**, Valentin Quelquejay-Leclère, Robin Zbinden,
Lucas Maystre, Matthias Grossglauser, Patrick Thiran

Information and Network Dynamics Lab, School of Computer and Communication Sciences, EPFL

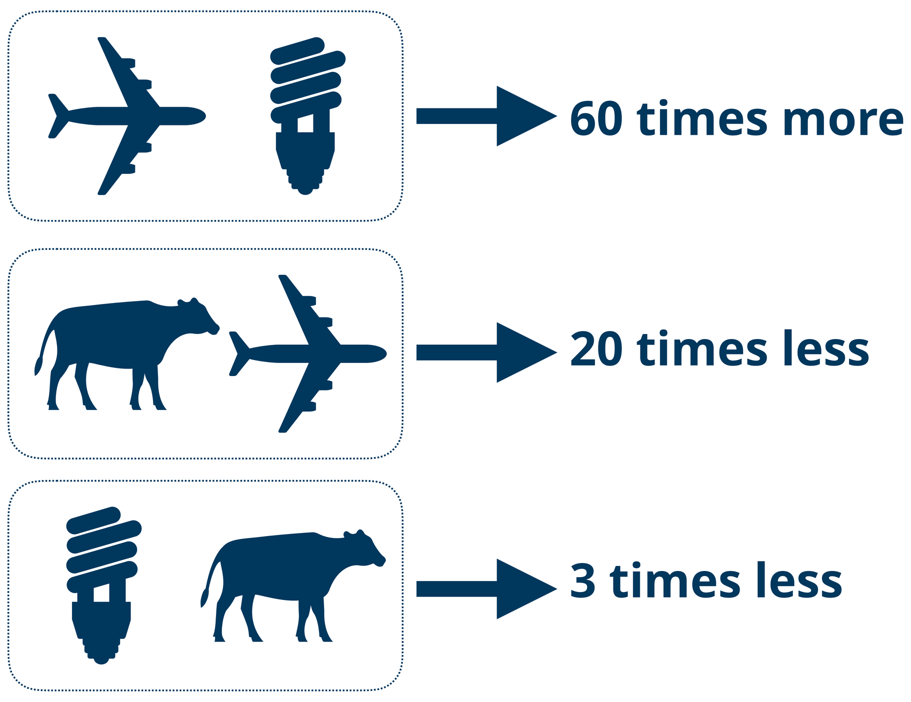
## Introduction

**Goal:** Understand how people perceive the carbon footprint of their actions.

**Problem:** Except for experts, it is very difficult to estimate $CO_2$ emissions in absolute terms.




**Idea:** It is easier to estimate the relative carbon footprint between two actions!




## Actions

Let $\mathcal{A}$ be a set of $M$ actions.

**Example:**

- Flying from London to New York
- Light your house with LED
- Eat meat for a year

Let $(i, j, y)$ be a triplet encoding that action $i$ has an impact ratio of $y$ in $\mathbf{R}$ over action $j$.

## Model

Given some parameters $w_i$ and $w_j$ representing the perceived carbon footprint of actions $i$ and $j$, we model the (log-)impact ratio as

$$\log y = w_i - w_j + \epsilon = \boldsymbol{x}^\intercal \boldsymbol{w} + \epsilon,$$

where $\epsilon$ is a zero-mean Gaussian noise with variance $\sigma_n^2$.

## Likelihood

We cast the problem of inferring a population's global perception from pairwise comparisons as a Bayesian linear regression.

**Likelihood:** For a dataset of $N$ independent triplets, the likelihood of the model is

$$p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{w}) = \prod_{i=1}^{N} p(y_i \mid \boldsymbol{x}_i^\intercal \boldsymbol{w}, \sigma_n^2) = \mathcal{N}(\boldsymbol{X}\boldsymbol{w}, \sigma_n^2 \boldsymbol{I}).$$

Comparison matrix $\boldsymbol{X} \in \mathbf{R}^{N \times M}$

Parameter vector $\boldsymbol{w} \in \mathbf{R}^{M}$

## Posterior

We assume a Gaussian prior for the weight parameters $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_p)$

$$\boldsymbol{\mu} = \mathbf{1}c, \; c = \frac{1}{M}\sum_{i=1}^{M} v_i \qquad \boldsymbol{\Sigma}_p = \sigma_p^2 \boldsymbol{I}$$

**Posterior:** The posterior distribution of the weight parameters given the data is

$$p(\boldsymbol{w} \mid \boldsymbol{X}, \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{X}, \boldsymbol{w}) p(\boldsymbol{w})}{p(\boldsymbol{y} \mid \boldsymbol{X})} = \mathcal{N}(\overline{\boldsymbol{w}}, \boldsymbol{\Sigma}),$$

$$\overline{\boldsymbol{w}} = \boldsymbol{\Sigma}(\sigma_n^{-2} \boldsymbol{X}^\intercal \boldsymbol{y} + \boldsymbol{\Sigma}_p^{-1} \boldsymbol{\mu}),$$

$$\boldsymbol{\Sigma} = (\sigma_n^{-2} \boldsymbol{X}^\intercal \boldsymbol{X} + \boldsymbol{\Sigma}_p^{-1})^{-1}.$$

## Active Learning

During one session of the quiz, a user sequentially answers comparison questions. Active learning enables us to maximize the information extracted from a session.

Let $\boldsymbol{\Sigma}_N$ and $\boldsymbol{\Sigma}_{N+1}$ be the covariance matrices of the posterior distribution when $N$ and $N+1$ comparisons have been respectively collected. Let $\boldsymbol{x}$ be the new $(N+1)$-th comparison vector. We want to select the pair of actions to compare that maximizes the total information gain

$$\Delta S = S_N - S_{N+1} = \frac{1}{2} \log(1 + \sigma_n^{-2} \boldsymbol{x}^\intercal \boldsymbol{\Sigma}_N \boldsymbol{x}).$$

Entropy of multivariate Gaussian $\qquad \boldsymbol{\Sigma}_N = [\sigma_{ij}^2]_{i,j=1}^{M}$

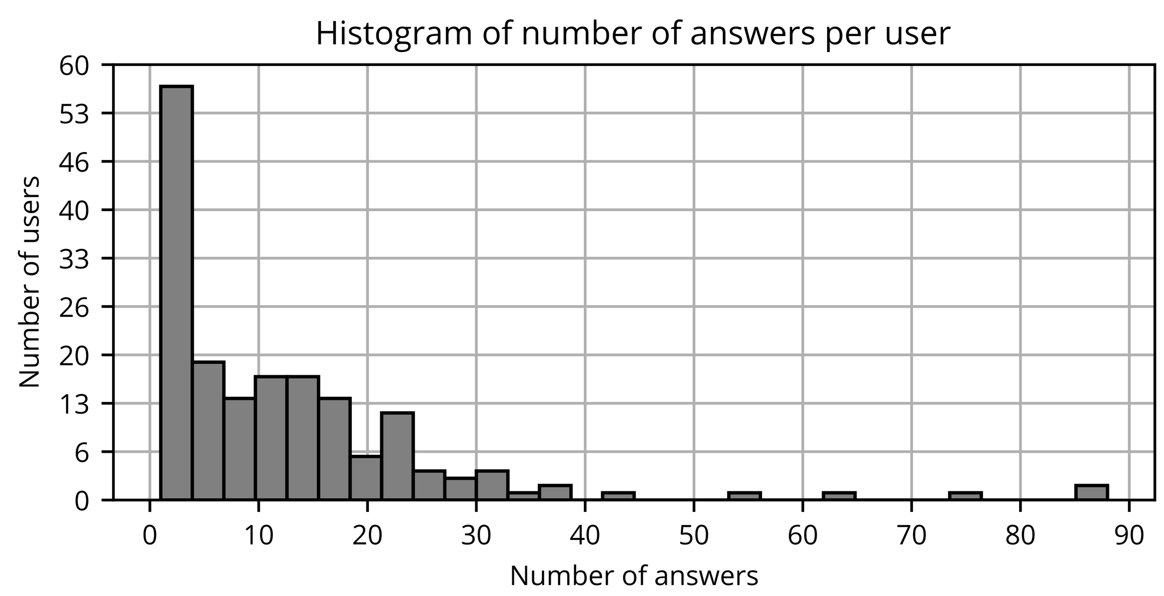
To maximize $\Delta S$, we maximize $\boldsymbol{x}^\intercal \boldsymbol{\Sigma}_N \boldsymbol{x}$ for all possible $\boldsymbol{x}$ in our dataset. We seek, therefore, to find

$$(i^\star, j^\star) = \underset{i,j}{\operatorname{argmax}} \{\sigma_{ii}^2 + \sigma_{jj}^2 - 2\sigma_{ij}^2\}.$$

## Data

We compile a set $\mathcal{A}$ of $M = 18$ actions.

We collect a dataset of $N = 2183$ triplets from a population of 178 users on a university campus (mostly students between 16 and 25 years old).




## Results