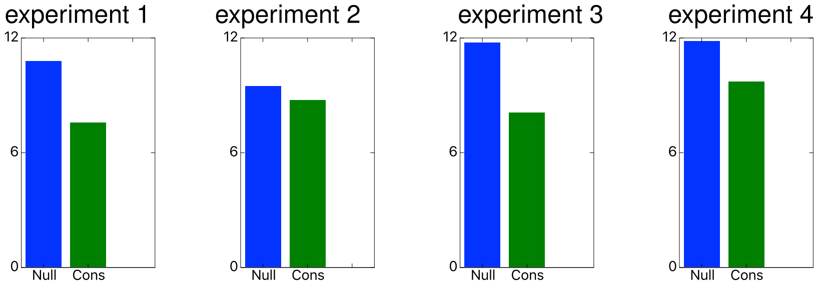
Model prediction errors



Situations

- Predictions under null model
- Predictions via our model

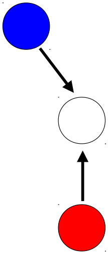
Modelling influence and opinion evolution in online collective behaviour

Samuel Martin
Joint work with **Corentin Vande Kerckhove, Pascal Gend, Julien Hendrickx, Jason Rentfrow, Vincent Blondel**
Centre de Recherche en Automatique, NANCY, CNRS-Uni Lorraine
UCLouvain, UCambdridge

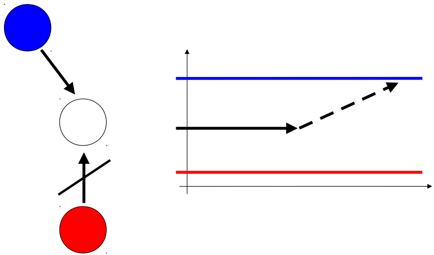
2016

Why modeling opinion dynamics ?

Why modeling opinion dynamics ?



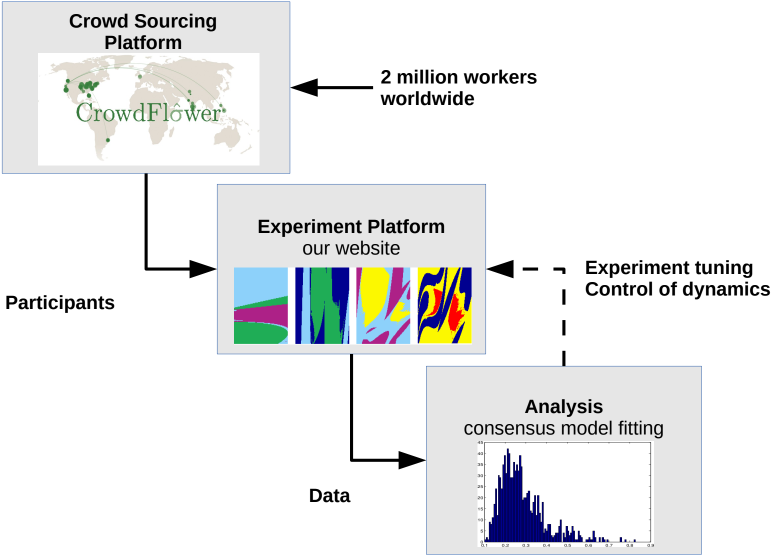
Why modeling opinion dynamics ?

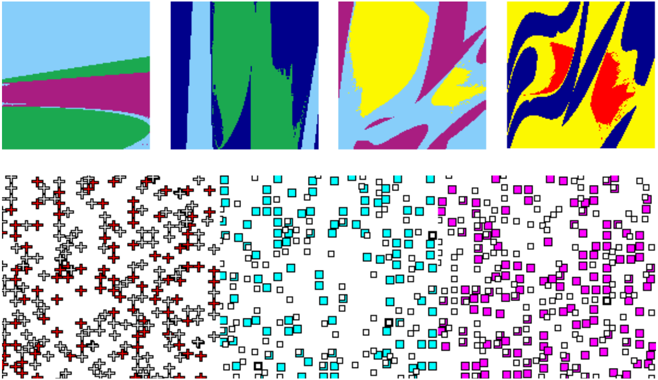


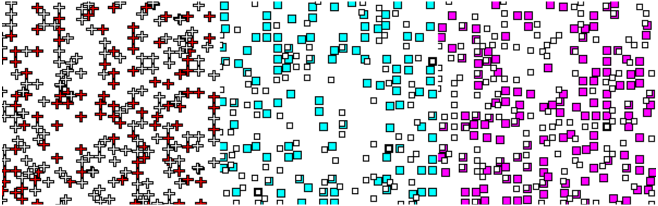
Research questions

- Can we model opinion evolution as a result of interactions ?
- How good can we expect predictions to be ?

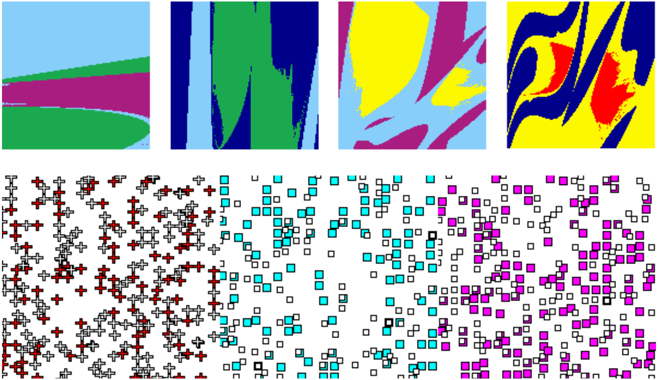
How to get opinion dynamics data ? An in vitro experiment







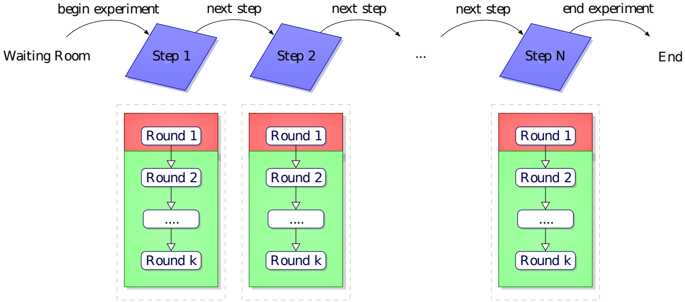
64 groups × 6 participants × 30 pictures × 3 rounds of estimations
71 groups × 6 participants × 30 pictures × 3 rounds of estimations



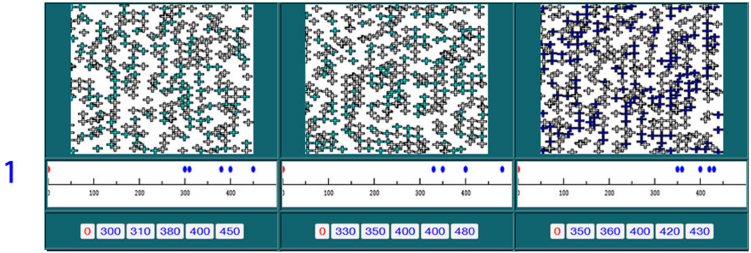
64 groups × 6 participants × 30 pictures × 3 rounds of estimations
71 groups × 6 participants × 30 pictures × 3 rounds of estimations

Incentives ⇒ Money $0.10 (+ ∼ $0.5) per 30min

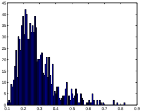
Experimental design



Example - Step 1, Round 2



Data analysis



Opinion dynamics models

$$x_i(t+1) = x_i(t) + \frac{1}{n}\sum_j a_{ij}(t)\ (x_j(t) - x_i(t)) + \eta_i(t)$$

Models

- **Null :** No influence : $a_{ij}(t) = 0$
- **Ours :** Infuencability Decay in time : $a_{ij}(t) = \alpha_i(t)$

Additive noise : $\eta_i(t)$

How to estimate the parameters ?

→ **Minimize the mean square error**

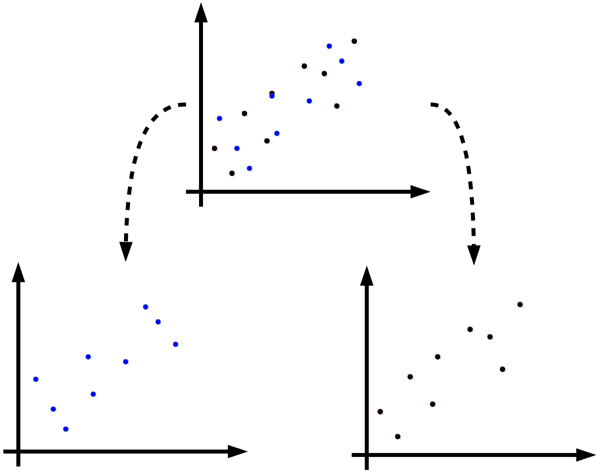
Mean-square error

$$MSE(\alpha, \sigma^2) = \sum_{g \in games} \|\tilde{x}(2) - x(2)\|^2 + \tilde{x}(3) - x(3)\|^2$$

$x(t)$: actual decision by the real participants
$\tilde{x}(t)$: prediction given $x(1)$ and α

How to assess the predictive power of the models ?

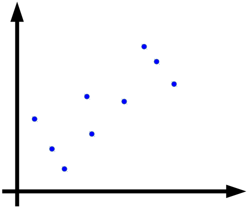
⇒ **Via crossvalidation :** Split population of 600 participants into 2



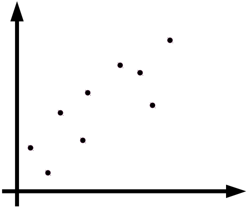
How to assess the predictive power of the models ?

⇒ **Via crossvalidation :** Split population of 600 participants into 2

Training data



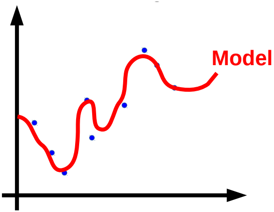
Validation data



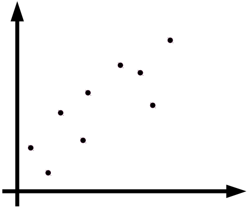
How to assess the predictive power of the models ?

⇒ **Via crossvalidation :** Split population of 600 participants into 2

Training data



Validation data

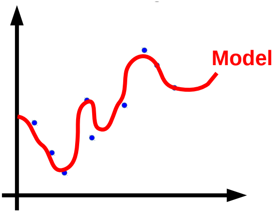


Learn $\alpha(1) \in \mathbb{R}$ and $\alpha(2) \in \mathbb{R}$ best predicting opinion evolution

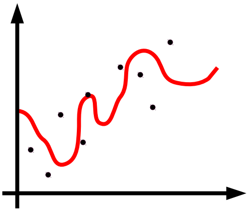
How to assess the predictive power of the models ?

$\Rightarrow$ **Via crossvalidation :** Split population of 600 participants into 2

Training data



Validation data



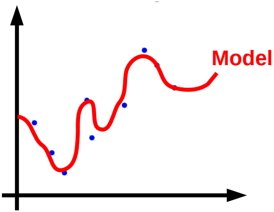
Learn $\alpha(1) \in \mathbb{R}$ and $\alpha(2) \in \mathbb{R}$ best predicting opinion evolution

Use $\alpha(1)$ and $\alpha(2)$ to predict opinion evolution

How to assess the predictive power of the models ?

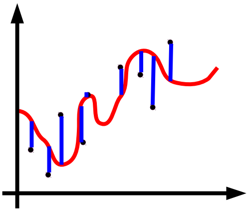
⇒ **Via crossvalidation :** Split population of 600 participants into 2

Training data



Learn $\alpha(1) \in \mathbb{R}$ and $\alpha(2) \in \mathbb{R}$ best predicting opinion evolution

Validation data



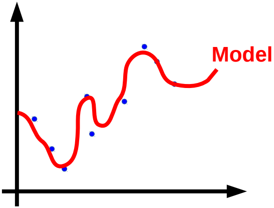
Use $\alpha(1)$ and $\alpha(2)$ to predict opinion evolution

→ Compute MSE on validation set

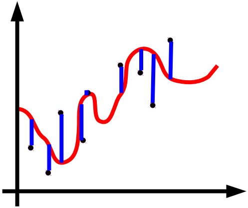
How to assess the predictive power of the models ?

⇒ **Via crossvalidation :** Split population of 600 participants into 2

Training data



Validation data



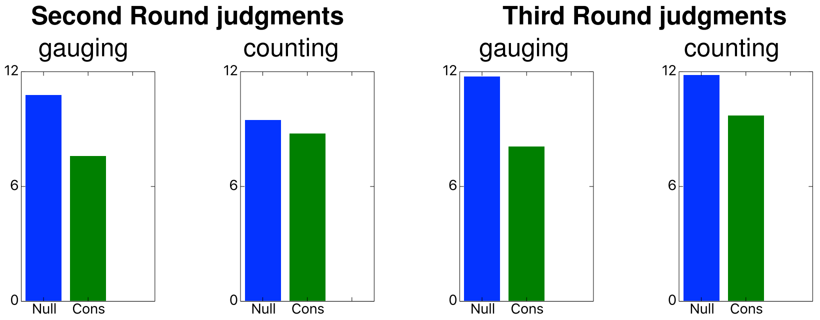
Learn $\alpha(1) \in \mathbb{R}$ and $\alpha(2) \in \mathbb{R}$ best predicting opinion evolution

Use $\alpha(1)$ and $\alpha(2)$ to predict opinion evolution

→ Compute MSE on validation set

→ Repeat many times + Compute average MSE

Model prediction errors

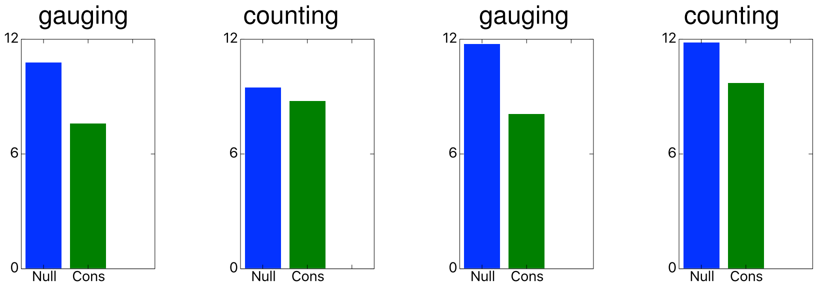


Situations

- Predictions assuming constant opinions
- Predictions via consensus model

Model prediction errors

Second Round judgments **Third Round judgments**

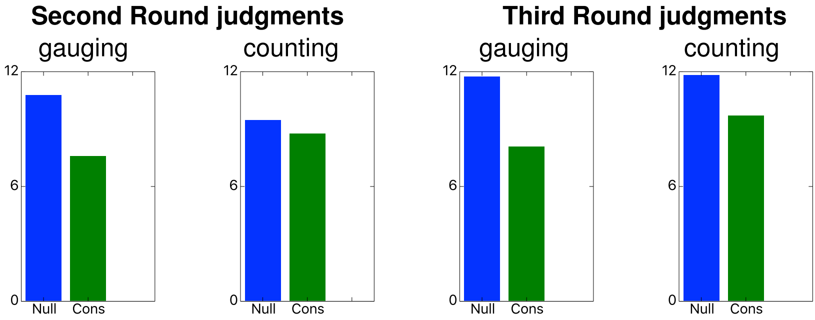


HOW GOOD IS THIS ?

Situations

- Predictions assuming constant opinions
- Predictions via consensus model

Model prediction errors



HOW GOOD IS THIS ? ... compared to BEST POSSIBLE PREDICTIONS ?

Situations

- Predictions assuming constant opinions
- Predictions via consensus model

Assume we know the ideal predictive model

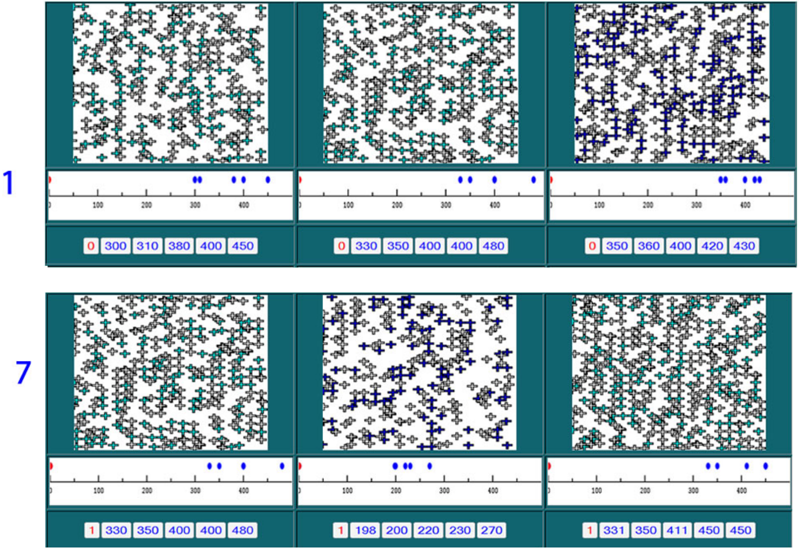
$$x_i(2) = f_i(x_i(1), x_{others}(1), picture) + \eta,$$

$\rightarrow$ std(η) intrinsic variation of a participant : lower bound for prediction error

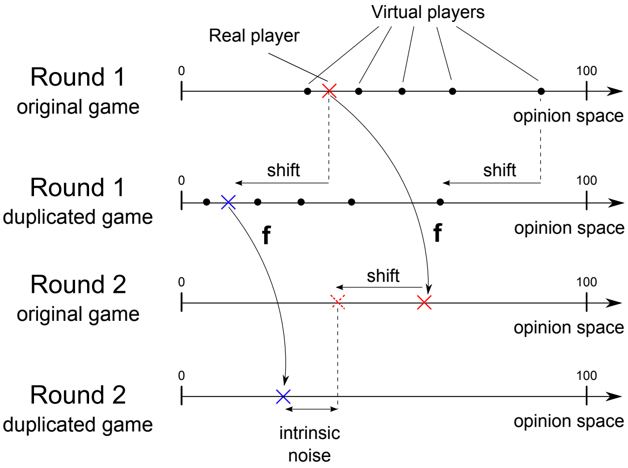
How can we measure std(η)?

Control experiment

→ Over 30 pictures, 20 were couples of replicates



Control experiment



Computing the intrinsic variation

Theorem

Assume it exists $\lambda \in [0, 1]$, function g_i and h_i such that

$$f_i(x_i(1), x_{others}(1), picture) = \lambda g_i(x_i(1), x_{others}(1)) + (1 - \lambda) h_i(picture)$$

and

$$g_i(x_i(1) + s, x_{others}(1) + s) = g_i(x_i(1), x_{others}(1)) + s,$$

Then,

$$std(\eta) = \sqrt{mean\left(\frac{1}{2}(x_i'(2) - x_i(2) - \lambda(x_i'(1) - x_i(1)))^2\right)}$$

where mean is taken over all repeated games and all participants and where the prime notation is taken for judgments from the second replicated game in the control experiment.

Proof : Judgments in two replicated games by a same participants :

$$\begin{aligned} x_i(2) &= f_i(x_i(1), x_{others}(1), picture) + \eta \\ x_i'(2) &= f_i(x_i'(1), x_{others}'(1), picture) + \eta' \end{aligned}$$

η, η' : 2 independent draws. Judgments are shifted by same known constant $s = x_i'(1) - x_i(1)$:

$$\begin{aligned} x_i'(1) &= x_i(1) + s, \\ x_{others}'(1) &= x_{others}(1) + s, \end{aligned}$$

With the assumption on f_i :

$$\begin{aligned} x_i(2) &= \lambda g_i(x_i(1), x_{others}(1)) + (1 - \lambda) h_i(picture) + \eta, \\ x_i'(2) &= \lambda g_i(x_i'(1), x_{others}'(1)) + (1 - \lambda) h_i(picture) + \eta' \end{aligned}$$

Then, with assumption on g_i :

$$x_i'(2) = \lambda \left(g_i(x_i(1), x_{others}(1)) \ + \ s \right) + (1 - \lambda) h_i(picture) + \eta'$$

and

$$x_i'(2) - x_i(2) = \lambda s + \eta' - \eta.$$

Recall

$$x_i'(2) - x_i(2) = \lambda s + \eta' - \eta.$$

Notice η, η' independent with zero mean, *i.e.* , $\mathbb{E}(\eta) = \mathbb{E}(\eta') = 0$, the theoretical variance of η is

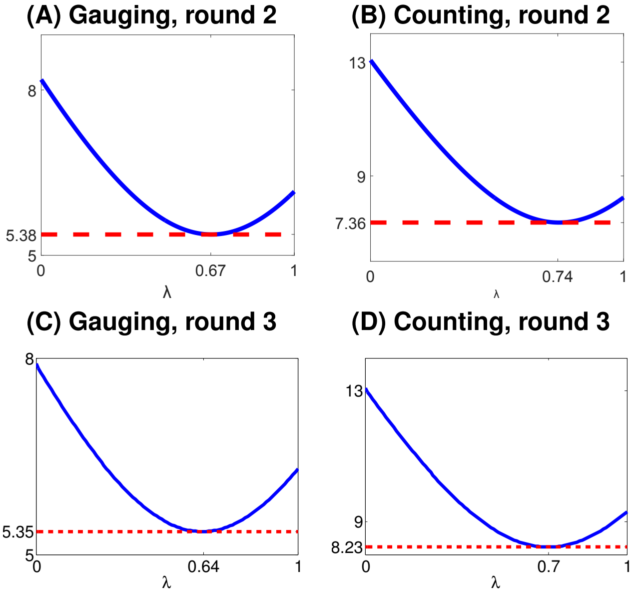
$$\begin{aligned}
\mathbb{E}(\eta^2) &= \frac{1}{2}\mathbb{E}(\eta^2) + \mathbb{E}(\eta'^2) \\
&= \frac{1}{2}(\mathbb{E}(\eta'^2) - 2\mathbb{E}(\eta')\mathbb{E}(\eta) + \mathbb{E}(\eta^2)) \\
&= \frac{1}{2}(\mathbb{E}(\eta'^2 - 2\eta'\eta + \eta^2)) \\
&= \frac{1}{2}\mathbb{E}((\eta' - \eta)^2).
\end{aligned}$$

Consequently, the variance of η is empirically measured as the average of

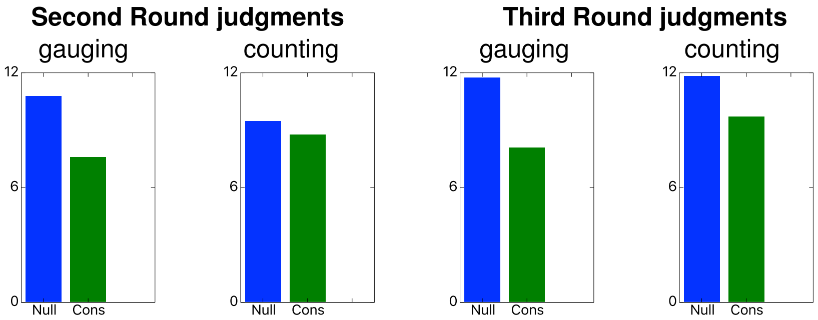
$$\frac{1}{2}(x_i'(2) - x_i(2) - \lambda s)^2$$

over all repeated games and all participants.

Take the most conservative std(η) over all λ



Model prediction errors



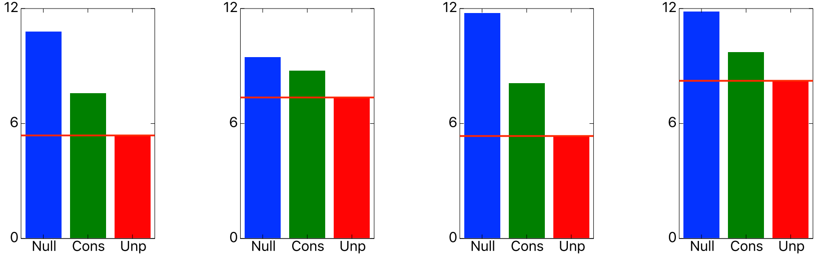
Situations

- Predictions assuming constant opinions
- Predictions via consensus model

Model prediction errors

Second Round judgments **Third Round judgments**

gauging counting gauging counting



Situations

- Predictions assuming constant opinions
- Predictions via consensus model
- Level of unpredictability

Prediction improvement

→ Use past information on each participant

- distinct $\alpha_i(1), \alpha_i(2)$ or
- 2 classes : **stubborn** $(\alpha^S(1), \alpha^S(2))$ or **compliant** $(\alpha^C(1), \alpha^C(2))$

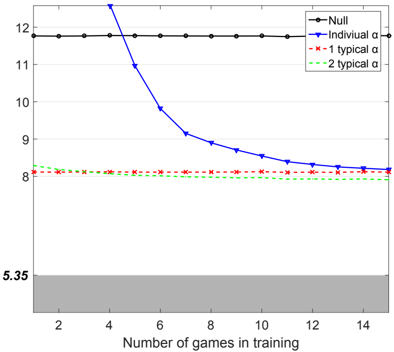
Prediction improvement

→ Use past information on each participant

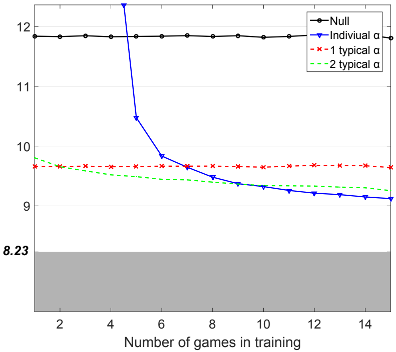
- distinct $\alpha_i(1), \alpha_i(2)$ or
- 2 classes : **stubborn** $(\alpha^S(1), \alpha^S(2))$ or **compliant** $(\alpha^C(1), \alpha^C(2))$

Third Round judgments

gauging

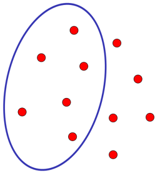


counting



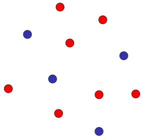
Challenges

Network



$$x(t+1) = \mathbf{A}_k\, x(t)$$

Bots



$$x(t+1) = \mathbf{A}\, x(t) + \mathbf{B}\, u(t)$$

Modelling influence and opinion evolution in online collective behaviour

Samuel Martin
Joint work with **Corentin Vande Kerckhove, Pascal Gend, Julien Hendrickx, Jason Rentfrow, Vincent Blondel**
Centre de Recherche en Automatique, NANCY, CNRS-Uni Lorraine
UCLouvain, UCambdridge

2016