

The Research Group  
**Artificial Intelligence Lab**

has the honor to invite you to the public defense of the PhD thesis of

## **Elias Fernández Domingos**

to obtain the degree of Doctor of Sciences

Joint PhD with University of Vigo

Title of the PhD thesis:  
**Coordinating Human and Agent Behavior in Collective Risk Scenarios**

Promotor:

**Prof. dr. Tom Lenaerts**

**Prof. dr. Juan C. Burguillo**

Co-promotor:

**Prof. dr. Jelena Grujić**

The defense will take place on

**Thursday, December 17, 2020 at 11h00**

The defense can be followed through a live stream. Contact

[Elias.Fernandez.Domingos@vub.be](mailto:Elias.Fernandez.Domingos@vub.be) for more information

### **Members of the jury in Vigo**

Prof. dr. Ann Nowé (VUB, Chair)

Prof. dr. Cristina López (University of Vigo, Spain, Secretary)

Prof. dr. Francisco Santos (Universidade de Lisboa, Portugal)

Prof. dr. Aleksander Byrski (AGH, Poland)

Prof. dr. Bipin Indurkha (Jagiellonian University, Poland)

### **Curriculum vitae**

Elias Fernández Domingos (1991) obtained the M.Sc. degree in Telecommunication Engineering on the double specialization of Electronics and Telematics in 2015 at the University of Vigo, Spain. He began his joint-PhD research under the supervision of Tom Lenaerts and Juan C. Burguillo at the Artificial Intelligence Lab in Brussels (VUB) and the University of Vigo. In 2017 he was awarded a F.W.O. Strategic Basic Research (SB) Ph.D. fellowship co-supervised by Jelena Grujić. He has collaborated internationally with researches at Instituto Superior Técnico, University of Lisbon (Portugal).

### **Abstract of the PhD research**

Climate action, vaccine hesitancy or control of pandemics are some of the many collective endeavours that put individual and collective interests under the shadow of a future collective disaster. These complex scenarios correspond to social dilemmas that are not only subject to many uncertainties. Additionally, increasingly often, these social situations are constituted of human-agent interactions through socio-technical systems already deployed at different levels of society.

In this thesis, we performed and analysed a series of behavioural economic experiments in a collective risk dilemma (CRD), which abstract the previously listed scenarios. Firstly, we designed an experimental study which evaluates how humans respond to uncertain deadlines in the CRD. We found that low levels of this timing uncertainty do not affect group success in avoiding the disaster, but increases earlier contributions. Moreover, in both timing uncertainty treatments, successful players adopted a reciprocal response. These results indicate that under timing uncertainty peers should respond early and in kind to avoid the consequences of a collective disaster. This outcome, as well as the results for other forms of uncertainty, are confirmed through a population-based learning model.

Secondly, we studied how the presence of autonomous agents influenced participants' behaviour in a CRD with a fixed number of rounds. We designed three experiments in which, respectively, participants have to delegate, customize or interact in a hybrid human-agent group. Delegation and customization considerably increased group success. Yet, participants were reticent to give away agency and would still prefer to make their own contributions throughout the experiment. Nevertheless, the ability to customize or personalize the agent, increased the willingness to delegate. Finally, while group success did not increase, in the hybrid experiment, most human participants considered that agents acted more selfishly, contributing less than humans, even though this was not true. These results highlight the relevant interplay between customization and social expectations towards autonomous agents and should be taken into account when designing hybrid socio-technical systems.