



Anthropo-Lab

ETHICS - EA 7446

Wednesday, march 16, 2022 - 10am-11:30am

Seminar

Delegation to artificial agents promotes  
pro-social behavior in collective risk dilemmas



Credit photo : stockadobe

**Venue**



Université Catholique de Lille  
Maison des chercheurs  
1er étage - salle 103  
14 boulevard Vauban  
59000 Lille



**Information & Registration**

[julien.navarro@univ-catholille.fr](mailto:julien.navarro@univ-catholille.fr)

/ LES FACULTÉS DE L'UNIVERSITÉ CATHOLIQUE DE LILLE /

## Speaker



Elias Fernández Domingos holds a PhD in Information and Communication Systems from the Vrije Universiteit Brussels (Belgium) and the University of Vigo (Spain). He is currently a post-doctoral researcher at the Machine Learning Group (ULB - Belgium). In his research, he combines behavioral economics, machine learning and evolutionary game theory to investigate how to foster cooperation in solving important human problems, such as climate change, and how artificial intelligence affects this process.

## Abstract

Home assistant chat-bots, self-driving cars, drones or automated negotiations are some of the several examples of autonomous (artificial) agents that have pervaded our society. These agents enable the automation of multiple tasks, saving time and (human) effort. However, their presence in social settings raises the need for a better understanding of their effect on social interactions and how they may be used to enhance cooperation towards the public good, instead of hindering it. To this end, we present an experimental study of human delegation to autonomous agents and hybrid human-agent interactions centred on a non-linear public goods dilemma with uncertain returns in which participants face a collective risk. Our aim is to understand experimentally whether the presence of autonomous agents has a positive or negative impact on social behaviour, fairness and cooperation in such a dilemma. Our results show that cooperation increases when participants delegate their actions to an artificial agent that plays on their behalf. Yet, this positive effect is reduced when humans interact in hybrid human-agent groups. Also, we show that humans have biases towards agent behaviour, assuming that they will contribute less to the collective effort. In general, we find that delegation to autonomous agents might act as commitment devices, which prevent both the temptation to deviate to an alternate (less collectively good) course of action, as well as limiting responses based on betrayal aversion.