8 Days at ICA 4 First session in Paris

Oraee, Atrina [01]

[0] *international laboratory for Brain, Music, and Sound research (BRAMS)*

[1] *IAS PARIS*

ABSTRACT

This 4th edition of the ICA hosted by the Paris Institute for Advanced Studies (Paris IAS) from October 18 to 27, 2021, explores fundamental interdisciplinary issues at the intersection of cognitive science, neuroscience and artificial intelligence. Decisive advances have been made over the last decades in the analysis of brain activity and its behavioral counterparts, as well as in information processing sciences. The complementarities between neuroscience/cognitive science and artificial intelligence allow us to explore synergies and raise ethical questions between these disciplines, which present considerable challenges and opportunities for the progress of society.

Introduction

The Intercontinental Academia (ICA) creates a global network of future research leaders in which some of the very best young academics work together on paradigm-shifting, cross-disciplinary research, mentored by eminent researchers from across the globe.

The ICA was established in 2016 through the University-Based Institutes for Advanced Study (UBIAS) coalition which has 44 member institutes around the world (Cordelois, 2020).

During each edition of Intercontinental Academia, participants get together in three sessions over the course of one year. **This is my div**

Previous editions of ICA have focused on "Time", "Human Dignity" and "Laws: Rigidity and Dynamics".

The 4th edition of the ICA explores the complementarities between artificial intelligence and neuro/cognitive-science and the tremendous challenges and opportunities they raise for humanity. Fellows and mentors initially met online and in cyberspace, and now in presence in Paris, from October 18 to 27. They shall meet again in cyberspace in the next few months and then finally, in Belo Horizonte in Brazil next June.

The first session, hosted by Paris Institute for Advanced Study (Paris IAS), includes an intense 10-days of scientific sessions, discussion forums as well as scientific exchanges with ENS-Paris Saclay, Sorbonne Center for Artificial Intelligence and Ecole Normale Supérieure.

Each day at the [Paris IAS](#), [ICA4 Fellows](#) meet with their [Mentors](#) for a closed 3-hour seminar, during which two mentors launch the discussion with a presentation. Upon completion of the seminar, the Fellows then meet for 45 minutes to list the key takeaways and ideas that have emerged from the discussion, followed by a collective brainstorming session . This ensures that the output of collective intelligence is collected, formatted and capitalised.

The other half of the day is left free for participants to reflect on the scientific discussions in small groups. Such discussions are occasionally complemented by lectures from the [Chairs](#).

Day 1: "The future needs wisdom!"

The very first lecture of ICA4 - Session 1 came from [Robert Zatorre](#) who took us into the fascinating world of music while explaining the relationship between perception, predictions, and pleasure!

This was followed by another lecture that introduced a rather different perspective on AI and was presented by [Eliezer Rabinovici](#). The lecture mostly explored the complexities of scientific enquiries and methods in the context of AI.

Before leaving for cocktails and welcome speeches in the chambers of Paris IAS, a final lecture was given by [Helga Nowotny](#), who emphasised an urgent need for establishing a context-sensitive AI control system.

Perception, prediction, and pleasure: What can music teach us about neurocognition/intelligence?

Presented by [Robert Zatorre](#)

It was stated during the seminar that the brain represents the properties of the environment and guides behaviour through evaluation and reward. Aesthetic pleasure can be defined as the

phylogenetically older system that is centred on the striatum.

Moreover, results of the relation between connectivity of the auditory cortex with the striatum and several behavioural results were presented (e.g. related to amusia, music-specific anhedonia). Dynamic causal modelling and predictive coding frameworks have been presented as possible explanations of the relationship between learning and reward in music. Predictions make the rewards evolve from a biological event to the expectation of the event.

Through the post-seminar collective discussions, the relevance of affective experience (pleasure and fear) in learning was emphasised. Discussions concluded with a rather open-ended question, leaving ICA4 Fellows wondering about whether or not AI should have a similar system for learning, and how should the reward and punishment be differentiated?" Maybe AI does not need to understand or experience human emotions; it just needs to behave like a human by capturing the features of a dataset that correctly describes the behaviour...

High Energy Physics: Successes, Challenges and Magic

Presented by *Eliezer Rabinovici*

It was discussed that observing natural phenomena can motivate scientific enquiry and drive us to understand the unknown. Moreover, equations are a way to increase predictability. However, a single, compact and reductionist explanation for all phenomena in the universe may not necessarily exist. The scientific method requires that results are reproducible. The correspondence principle requires that new theories can explain all phenomena for which a preceding theory was valid. To understand a phenomenon, one has to identify the relevant players and determine the correct explanation scale.

In AI We Trust: Power, illusion and control of predictive algorithms

Presented by **Helga Nowotny**

The session began with introducing the concept of singularity and defining it as a tipping point: a change of state that can lead to the collapse of a system. In an attempt to define Ethical AI, examples such as Transhumanism (ideas of transcending the limitations of a mortal body through information sharing) were discussed. Furthermore, the illusion that AI knows humans better than humans know themselves was elaborated, ultimately concluding by mentioning the existence of a possibility for human beings to both profit or suffer from an AI system depending on how it is applied.

"The future needs wisdom": an urgent need to institutionalise context-sensitivity, rather than creating a standardised system to control AI, was discussed and collectively developed. This lead to further debates regarding the concentration of technology advancement and its deteriorating impact on inequality. Thus, a global agreement is necessary to control AI, although it is currently almost impossible to obtain! Therefore, we should educate AI as a child of humanity that can grow to

contribute to society. AI research is undertaken at such a massive scale that it requires global efforts which go beyond a single country. Scientists are paid by society, and their curiosity-led work should return to society as a whole…

Day 2: "In AI we trust"...or not!**

The ICA4 continued onto the second day, through which three seminars took place with mentors who had joined the first session in Paris from around the world!

The first lecture was by *Robert Aumann*, a Nobel prize laureate, who focused on the convoluted concept of consciousness and its counterparts.

This was followed by a lecture from *Karen Yeung*, who offered a rather critical point of view on the prevalence of AI, as well as some of its surrounding myths and misconceptions. She then went on to explain how responsibility should be re-defined to consider the unintended impact(s) of AI in human societies.

Finally, *Raouf Boucekkine* took the fellows for an exploration into the world of economics and finance, using the concept of equilibrium as an example to illustrate the difference between disciplines: mainstream economics VS. statistical physics!

- Why Consciousness?

  Presented by Robert Aumann

  Essentially, the seminar was focused on the purpose which consciousness serves. Consciousness was defined as the ability to do the following:

- Perceive

- Feel (emotions)

- Think/intend

- Carry out intentions (volition)

  Of all the above, perceiving, thinking/intending, and carrying out intentions may be done by machines. However, feelings and emotions belong exclusively to human beings. In such a context, it may be argued that the evolutionary function of consciousness is to enable the operation of emotions. This being said, we currently have no idea about how does consciousness work. Although considerable progress has been made in AI, Artificial Emotions (AE) has remained rather untouched.

Myths and misunderstandings about responsibility for the unintended impact of AI

Presented by [Karen Yeung](#)

The talk mostly focused on responsibility for the unintended impact of Artificial Intelligence, based on the presenter's Council of Europe study. It was argued that Machine Learning's (ML) capacity to enable task automation and machine "autonomy" raise important questions about responsibility. Thus, responsibility-relevant attributes of ML were identified, for which an illustration is the data-driven profiling of individuals, and other ML applications, which may hold adverse impacts on human rights, on both individual and collective levels.

While responsibility is important for human beings, who are considered as moral agents, to maintain peaceful social co-operation within the community, only a few studies have focused on tackling the fundamental role of responsibility for individuals, as well as the society.

The impacts produced by complex socio-technical systems using ML technologies have generated a range of concerns that fall under the heading of "algorithmic responsibility". While existing laws have an important role to play in ensuring the accountability of algorithmic systems, the implications of these technologies for their interference with human rights need to be studied further. This has been the primary focus of Karen Yeung's research.

In a nutshell, two dimensions of responsibility are required:

- Historic or retrospective responsibility: responsibility for conduct and events that occurred in the past
- Prospective responsibility: roles and tasks that look to the future

Finally, five common myths and misunderstandings concerning responsibility for the unintended adverse impacts of AI were identified:

- Need for effective and legitimate mechanisms to protect human rights from AI applications.
- Identifications of the appropriate responsibility model for allocating, distributing and preventing the various threats and risks.
- Responsibility of states to ensure that these policy choices are made in a transparent and democratic manner, in order to effectively protect human rights.
- Need for more interdisciplinary research
- Application of the fundamental principle of reciprocity so as not to allow those who develop and run our advanced digital technologies and systems to increase and exercise their power without responsibility.

Data science and deep learning vs theory: two examples from economics and finance

Presented by *Raouf Boucekkine*

The session included discussions on Data Science, Machine Learning (ML), and some relevant theories in the field of economics and finance that share common disciplines. Certain examples from macroeconomics, in which characteristics of the underlying mechanisms for complex systems are of great interest, were then discussed in more detail. In such context, a misunderstanding between different disciplines was highlighted: the concept of equilibrium is of great significance in mainstream macroeconomics, whereas this is not the case for statistical physics (e.g., the "equilibrium" bias outside the econ area, discussed by Bonneuil & Boucekkine (2020)). Finally, the use of various methods and approaches, such as DSGE (Dynamic Stochastic General Equilibrium), ABM (Agent-Based Modeling), and Neural Network-Based methods, in the field of macroeconomics were discussed.

Day 3: "What you do FOR people, you do TO people, so do it WITH people!"

Day 3 of the first session of ICA4 continued in Paris IAS, where the Fellows sat through three more scientific seminars, followed by discussions and brainstorming sessions.

The day kicked off by a framework proposed by *Saadi Lahlou*, called "Installation Theory", which enables scientists to analyse and regulate human behaviour. This was complemented by a new technique to capture the subjective perception of action, ultimately bringing psychological and behavioural sciences one step closer to what was once considered a technically impossible task: introspection!

William Hopkins then joined the discussions with some stimulating videos from research done on apes, while exploring self-recognition and social cognition in animals.

Finally, Toshio Fukuda revealed the Moonshot project: a society where humans and robots live together in 2050 !

Distributed Intelligence & Distributed Agency

Presented by *Saadi Lahlou*

We want intelligence to perform relevantly adapted actions that change the situation in which we are for the better. To design intelligence, we must first understand the nature of the actual activity. In this sense, the behaviour was defined as what people do, seen from the outside. In other words, behaviour remains an external description of objective phenomena. This is while activity is how people subjectively perceive their action and how they see it from their own perspective.

Installations consist of components that simultaneously support and control. In other words, they are specific, local, and societal settings where humans are expected to behave predictably, e.g., airport, metro, cash machine, etc. Installations consist of three layers: affordances of the physical environment, embodied competencies and social regulations. Intelligence is thus distributed over these three layers.

The question now is: why do we have these installations? Because installations channel many of our behaviours and consequently make us very efficient, although our short memory and cognitive processing are very limited compared to animals. Installations are redundant, and redundancy produces resilience and learning.

Moreover, certain questions on designing trade-off issues were raised: which agency for whom? AI agent? what kind of competence for the AI? What affordances? What rules? What degree of awareness? To whom does the agent report? How is it evaluated? He also added the "privacy dilemma." In other words, for better service, one must disclose information. Is there an "agency dilemma"? Can we make it explicit? Because the agency is distributed, so the responsibility is shared. It means that we now have the "many hands" problem. Thus, credentials for AI were suggested, which include values (what does it try to optimise), ownership (who takes responsibility for its actions), principles of action (rules, algorithms, domain of awareness and action), track record (list of transactions executed, includes initial training).

To conclude, ICA4 Fellows were left with some questions as food for thought. For each activity, do we want to augment existing agents with more agency? If so, which ones? Humans? Material objects? Social system? New agents? Who learns what? What values do we want to foster? What do we want AI for must be addressed for each activity, starting from the activity and discussed?

Perspective on Artificial Intelligence research from studies on Agency, self-recognition and social cognition in animals

Presented by **_William Hopkins_**

The session began by discussing humans constructed concepts to reflect intellectual abilities in various domains of cognitive functions. In this sense, we use tests like the WAIS or Stanford Binet to quantify and scale performance to standards for specific age classes. These tests rely heavily on language. There are many approaches to developing fair tests of cognition between species with different sensory and motor capabilities. It began with Darwin and Furness. Then, George Romanes (1884) focused on animal intelligence and later on, Kohler (1925) on insight learning. Within the same field, Robert Yerkes (1916) worked on "The mental life of monkeys and apes: a study of ideational behavior". Yerkes later developed the IQ test used by the army in WW1 (army alpha test).

Upon drawing on some of the literature, several videos were played, in which apes carries out various tasks including retrieving a peanut in the bottom of a tube followed by a video from an ape imitating a human, and so on. Several animals passed the mark test. E.g., magpies with a yellow

stick in their neck can identify it and try to remove using the mirror. Cortical parcellation of chimpanzee brain - compared to humans, the ones that passed the test show differences in some cortices. Grey matter differences between MSR+ vs. MSR-. They also analysed the anterior cingulate since such neurones are rather long and connect the anterior cingulate with the insula.

Moreover, results from studies that showed that human children outperformed apes on the social, but not physical, cognition tasks were presented and discussed. Much like the research in AI, Most early comparative studies of cognition and intelligence were strongly rooted in associative learning theory. However, associative or operant theories of learning were and are notoriously anti-cognitive. In the 1960s, there were attempts to reach apes alternative communication systems. The goal of the ape language studies was to determine whether language is uniquely human. The answer depends on how we define language.

However, is it language? There is very little evidence for declarative production (e.g., turn on the TV, give me an onion) in communication signals by primates and other animals. The other question is: are social stimuli rewarding? For chimpanzees, yes. Experiment: touch one button to see other chimpanzees or another button to see random animals. The chimp chose to press the button to see other chimps. Thus, the role of reward guides the learning and behaviour of animals. Although animal cognition is often used to explain animal behaviour, most can be explained by an associative learning mechanism.

AI and Robots for Future: The Moon Shot Project

Presented by *__Toshio Fukuda__*

Robots are avatars that pop up to help when humans need them. There is an information and physical interaction between robots and humans. Toshio showed several multi-scale robots, e.g., monkey-type robots, multi-locomotion, intelligent cane, etc. One of these robots is the Brachiator I-III. Brachiation is a form of long-armed ape locomotion. It uses dynamics of the pendulum, under-actuated mechanical system, variable constraint system, machine learning, AI, reinforce learning, soft computing (fuzzy, genetic algorithm). Regarding multi-locomotion types, in many cases, one creature has multiple types of locomotion in order to improve its mobility. The motivation of their study is to develop a robot mechanism and a control architecture that can achieve multiple locomotions. Hybrid computational intelligence, i.e., AI and brain interface were also commented upon by the speaker while showing a series of related videos. An example of such videos illustrated the Boston dynamics atlas and others: three robots dancing and jumping which was quite impressive!

Moreover, AI+Robot+IoT (Internet of Things), the use of robots in mega-trend (energy, urbanisation, food, ageing, global warming, robot, and AI) were discussed. This was followed by further discussions on autonomous cars, which may be safer than human drivers, mixed reality, the Eve project (a transparent body that simulates the human body), cyborg technology (fusion of robot and animal), and multiple robots (communication among robots).

Day 4: A visit to The University of Paris-Saclay**

![[](https://lh6.googleusercontent.com/NlgQpGHaBLRystLQsRXAbmAklsdIPADYo2m2139yVcZ4-zzpizb3wV2JZt231Ei1_LKcE10ENUndEKh6G8APDMxkjlHUFiQ_eN9eDjj-Pw6qEoONXtNSNyGxe1sEsn6APg =602x451)

The Fellows embarked on their first scientific trip for ICA4 and were hosted by **Ecole Normale Superieure de Paris-Saclay** throughout Day 4. The sessions at Saclay included two thought-provoking talks by *Xiao-Jing Wang* and *Jay McClelland*, both of which touched upon the principles underlying cognitive behaviours, as well as the difference between human and machine intelligence. These were followed by a symposium on AI at University Paris-Saclay. The symposium was followed by a half-day event with multiple workshops in which ICA4 mentors discussed major advances and issues surrounding AI with other world-class researchers such as *Stanislas Dehaene*. Finally, the intellectually intense day came to an end with a talk in which *Zaven Paré* raised important questions regarding how we will interact with AI algorithms and intelligent robotics in the decades to come…

Efforts to understand the computational principles underlying cognition

Presented by *Xiao-Jing Wang*.

Deep neural networks, despite their recent success, differ from human cognition because they have no internal mental life - instead, they act as complex, nonlinear input-output functions. In humans, the prefrontal cortex (PFC) is known to be crucial for cognitive functions such as working memory, decision making, and executive function. An early avenue of this research involved understanding how persistent neural activity may underlie working memory by sustaining stimulus information in the brain after the sensory cue has disappeared. Such persistence is linked to recurrent connectivity, which is lacking in most deep networks. Wang described his previous research using spiking networks and tools from dynamical systems to understand the attractor dynamics behind this form of memory. In the second half of the talk, he showcased his more recent work which uses recurrent neural networks (RNNs) as a form of a model organism to probe how the PFC may perform multiple cognitive tasks simultaneously. These RNNs can then be used to address questions such as whether the PFC encodes cognitive building blocks in a compositional manner, similar to the psychological concepts of schema.

A different distinction between human intelligence and AI

Presented by *James McClelland*

While the latter (in particular machine learning algorithms) learns from statistics on large-scale input data, humans learn to learn from explanations structured by culturally invented systems. Indeed, humans fail to perform in systematic ways, which we would expect if the structure were built into our cognitive functionality. But, McClelland points out that simply building in structure, as proposed by the pioneers of GOFAI, limits flexibility. This structure, McClelland argued, is built by culture. For example, he described a classic study by Scribner and Cole in 1973 which showed that non-Western cultures often lack a concept of absolute number and tend to classify objects based on concrete situations rather than abstract category membership. These authors proposed that Western education creates a context in which certain abstract relational concepts are learned, consistent with McClelland's later work correlating sudoku puzzle performance to mathematical education level. McClelland closed by reiterating that AI learns by examples but humans learn by explanations and that his explanation-based learning (rather than built-in structure) may underlie our propensity for one-shot learning.

Upon completion of the talks by ICA4 Mentors, Paris-Saclay hosted a half-day event with multiple workshops in which ICA4 mentors and Paris-Saclay researchers discussed major advances and issues surrounding AI. *Stanislas Dehaene* presented a series of fMRI, MEG, and behavioural evidence that humans use symbolic and recursive strategies on prediction tasks with complex sequences, as compared with monkeys which seem to use a picture-based strategy. In a session focusing on AI and ethics, *Paola Tubaro* revealed the hidden human workers who provide the hand-labelled training data for products such as Siri. This is due to companies and corporations needing a cheap workforce in the same language, ultimately reproducing historic colonial patterns.

Finally, the intellectually intense day came to an end with a talk in which *Zaven Paré* discussed his artistic works based on electronic marionettes and his collaborations with robotics specialists in Japan. Paré's conception of automaton-centred theatre enchants audiences while challenging our tendency towards anthropomorphisation. This raises important questions regarding how we will interact with AI algorithms and intelligent robotics in the decades to come...

Bibliography

Cordelois, A. (2020). How hard can it be to make an ideal publication website? *Nature*, *222*.