Model-Development in Neuroscience: Generalizability and Simplicity in Mechanistic Explanations

Final Symposium 26-28 October 2022

BOOK OF ABSTRACTS

1 - Integrating computation into the mechanistic hierarchy in the cognitive and neural sciences (in agenda)

Presenter: Oron Shagrir, Hebrew University of Jerusalem

Format: 45mins Presentation, discussions

It is generally accepted that, in the cognitive and neural sciences, there are both computational and mechanistic explanations. We ask how computational explanations can integrate into the mechanistic hierarchy. The problem stems from the fact that implementation and mechanistic relations have different forms. The implementation relation, from the states of an abstract computational system (e.g., an automaton) to the physical, implementing states is a homomorphism mapping relation. The mechanistic relation, however, is that of part/whole; the explaining features in a mechanistic explanation are the components of the explanandum phenomenon and their causal organization. Moreover, each component in one level of mechanism is constituted and explained by components of an underlying level of mechanism. Hence, it seems, computational variables and functions cannot be mechanistically explained by the medium-dependent states and properties that implement them. How then, do the computational and the implementational integrate to create the mechanistic hierarchy? After explicating the general problem, we examine two possible solutions. On one solution, the mechanistic hierarchy embeds at the same levels computational and implementational properties. This picture fits with the view that computational explanations are mechanistic sketches. On the other solution, there are two separate hierarchies, one computational and another implementational, which are related by the implementation relation. This picture fits with the view that computational explanations are functional and autonomous explanations. It is less clear how these solutions fit with the view that computational explanations are full-fledged mechanistic explanations.

2 - More Levels of Explanation (in agenda)

Presenter: Jäkel Frank, Technical University of Darmstadt

Format: Presentation and discussions

David Marr's levels of explanation are the de-facto standard for computational explanations in cognitive science. However, there is a lot of confusion about where to draw the boundaries between the computational, the algorithmic, and the implementational levels. I will compare Marr's levels to the more detailed levels that Alan Newell uses for computational systems. Newell's levels allow us to subdivide Marr's algorithmic level into hardware-dependent and hardware-independent levels. They are also necessary if cognition is to be described as a physical symbol system. I will present why I prefer Newell's levels over Marr's. The main reason is that Newell's levels are a better match for universal computers. I would like to discuss with you whether the human cognitive system is indeed a universal computer and what this means for mechanistic explanations.

3 - The “Better” Model: How Epistemic Goals Influence Model Choice in (Neuro)Economics

Presenter: Carla Nassisi, alumna Witten/Herdecke University

Format: lightning talk

Ever since the publication of Adam Smiths’ “Wealth of Nations”, many have tried to bring a “psychological” element back into the study of Economics, in an effort to make the infamous “Homo Oeconomicus” more human, and to ground economic models in psychological realism rather than in mathematical abstraction. Such attempts have culminated in the birth of Neuroeconomics, a discipline lying at the border between neurosciences, behavioral sciences and microeconomics. But what if this whole debate has been ill-construed? What if both neuroeconomists and traditional economists were right: their models are good enough… for a specific epistemic goal?

In this lightning talk, we will look into a concrete example (the application of Drift Diffusion Models - DDMs - to Neuroeconomics) to discuss how the concept of epistemic goals or epistemic virtues can shed some light on the feud between different schools. We will further discuss how the epistemic goals can be conceptualized as a way to concilize Popper's stances with a Kuhnian view of the evolution of social sciences.

4 - Neural Hardware for the Language of Thought (in agenda)

Presenter: Gualtiero Piccinini, Professor of Philosophy, University of Missouri–St. Louis

Format: Presentation and discussions

Time: 26th, 4pm

5- Predictive coding models at the intersection of deep learning and neuroscience (in agenda)

Presenter: André Ofner, Otto von Guericke University of Magdeburg

Format: 20 mins Presentation + discussions of a work-in-progress paper

Predictive coding is a promising candidate for an integrative computational theory of human and artificial cognition. Recent successes in training deep neural networks have enabled large scale implementations of information-theoretic objectives, such as the evidence lower bound (ELBO). While the ELBO also serves as the central objective for predictive coding networks (PCNs), the similarities between PCNs and deep neural network (DNN) based architectures like VAEs lack in-depth comparison. Focusing on on-going research, this talk will relate PCNs with their DNN based counterparts in terms of performance and biological plausibility.

6 - Distinguishing agent-specific and decision-specific channels of attention in risky decision-making (in agenda)

Presenter: Jan Engelmann, University of Amsterdam

Format: tba

Economists have become increasingly interested in using attention to explain behavioral patterns both on the micro and macro level. This has resulted in several disparate theoretical approaches. Some, like rational inattention, assume a “top-down” model of executive optimization. Others, like salience theory, assume a “bottom-up” influence where attention is driven by contextual factors. This distinction is fundamental for the economic implications of attention, but so far there is little understanding of their relative importance. We propose a multi-attribute random utility model that unifies prior theoretical approaches by distinguishing between the impact of agent-specific attention and decision-specific variation in attention, which parallel the distinction between top-down and bottom-up attention that is commonly made in cognitive science. We verify our framework in a number of experimental contexts: (1) an eye-tracking experiment on risky choice; (2) a follow-up experiment that determined exogenously how information about risky lotteries can be sampled; and (3) in patients with gambling disorder (GD). Jointly, our results underline the utility of differentiating between agent-specific attention and decision-specific variation in attention in identifying the underlying cognitive mechanisms involved in economic choice across different contexts.

7 - Title: Toward a Cognitive Science of AI: Methods and Norms

Presenter: Carlos Zednik, Assistant Professor of Philosophy, Eindhoven University of Technology

Format: Lightning talk, brainstorming sessions, discussion

What can and should explainable artificial intelligence (XAI) learn from efforts to explain biological intelligence? In a short introductory talk I will briefly distinguish between three dimensions of a "cognitive science of AI": explanatory norms; ("top-down") behavioral experiments; and ("bottom-up") neuroscientific investigation. After introducing some recent examples along with possible AI application areas, the audience will be invited to collectively brainstorm other potentially relevant cog sci norms and methods, as well as to critically evaluate their value for AI.

8 - Cross-talk and competition between levels of analysis in neuroeconomics (in agenda)

Presenter: Sebastian Gluth,University of Hamburg

Format: Presentation, questionnaire, online voting, discussions

Time: 27th

Neuroeconomics seeks to better understand and predict economically relevant behavior by studying the neural and cognitive principles of psychological processes such as value-based decision making and reward-based learning. Although the long-term goal is to achieve a comprehensive understanding of the biological and computational mechanisms to allow specific and precise predictions of individual as well as collective behavior, the field faces certain frictions across the different levels of analysis (i.e., from processes at single neurons to large-scale phenomena in society) and disagreements about their relative importance. In this interactive talk and discussion, I will first provide a short overview of the different sub-fields within neuroeconomics and then illustrate difficulties in balancing the different levels of analysis, using examples from my own research. Interactive elements such as online voting will be used to stimulate the discussion throughout the presentation.

9 - Blackbox AI: How to Discover What Happens Inside (in agenda)

Presenter: Lena Kästner, University of Bayreuth

Format: TBD

Modern artificial intelligence (AI) systems are often complex and opaque. At the same time, they are becoming increasingly prevalent in our lives. As a result, there is an increasing demand to make AI systems explainable and their behaviour intelligible. Recent work primarily approaches this problem by employing specific explainability methods to aid in-context understanding. We think, however, that important desiderata such as safety and reliability might be best satisfied through the increase of expert understanding with respect to how AI systems as a whole work.

We suggest this requires research undertaken from the scientific perspective. Our approach starts from the premise that once AI systems become sufficiently complex, they are best investigated and explained through the same lens as biological organisms. Accordingly, this work seeks to characterise the functional structure that emerges in AI systems through training. As we will describe, researchers pursuing this approach adopt strategies for discovery that have proved successful in the life sciences, such as pattern recognition, functional decomposition, localization, and systematic manipulation.

10 - What is a Neural Approximation? (in agenda)

Presenter: Arnon Levy, Hebrew University of Jerusalem

Format: Presentation, discussions

Many computational neuroscience models attribute to the brain computations that are highly demanding, resource-wise. In these cases, it is often assumed that while the brain does not, strictly speaking, perform the computation in question, it can be seen as *approximating* it. But it is often unclear what is assumed when the brain is modeled in terms of such approximation algorithms. What does it mean to say that the mind/brain approximates some algorithm? Under what conditions can the brain treated as approximating a specified computation? When are approximation-based models legitimate and what do they tell us about how the brain computes?

My presentation will not attempt to settle these questions. Instead, it will largely consist of presenting cases (primarily from recent Bayesian modeling) and raising questions. I will outline several possible interpretations of claims of the form “the brain approximates algorithm *A*“, and will discuss their pros and cons with a view to generating a group discussion. The main aim is to bring the notion of a mental/neural approximation into clearer view and to problematize it.

11 - Why Social Robots Need Self-Preservation to Be Objects of Moral Consideration

Presenter: Mohamed Hassan, MA Alumnus, Witten/Herdecke University

Format: lightning talk

Can social robots (SRs) feel? Should we grant SRs legal rights and on what basis? While there is a wide agreement today on their lack of technological advancement to be conscious or sentient, two possible dangers are important to address today. Firstly, SRs could be conscious in the near future or even today in some way that we are not able to understand or verify. Secondly, if we decide to err on the side of caution and grant them legal protection anyway, we could be infringing on personal and intellectual freedom by restricting particular uses of robots (e.g. sexual acts) or the research and development of said robots. This brings the question that is central to this paper: Where can we draw the line? Put in another way, how can we know if SRs are objects of moral consideration (OMC: such as dogs, bees, or trees) or an object of human desire (OHD: such as toasters or toys)? This paper presents the condition of self-preservation as a necessary and sufficient condition to draw the line between OMCs and OHDs.

12 - Generalizability and Simplicity in Boolean Inferential Methods for the Establishment of Constitutive-Mechanistic Models in the Cognitive and Biological Sciences (in agenda)

Presenters: Jens Harbecke & Johannes Mierau, Witten/Herdecke University

Format: project talk

According to the “mechanistic approach” to the cognitive and biological sciences, scientific explanations succeed by analyzing the mechanisms that underlie a phenomenon or “constitute” it on several levels. In this paper, we are concerned with the formal strategies to establish such multi-level causal-mechanistic models. Our goal is threefold: On the one hand (A), to offer a novel algorithm that transforms data tables obtained from tests on multi-level systems into causal-mechanistic models compatible with these tables. On the other hand (B), we would like to offer some philosophical insights suggested by, and associated with, the solutions produced by this script. The Python script “mLCA” developed by us is able to generate adequate and highly complex causal models and causal-mechanistic models. Based on some prototypical examples of models produced by this script/algorithm we are also able to defend two philosophical claims: 1. Inference methods generating causal-constitutive models require information about level assignments of the variables listed in the data tables produced by multilevel structures. Multi-level mechanistic models are not retrievable from coincidence data without such additional qualitative information. 2. The number of solutions generated by the mLCA script grows steeply relative to the number of relevant variables listed in a data table. Further reductions inevitably involve a pragmatic dimension, which has sharp consequences for the realistic ambitions and generalizability of mechanistic explanatory projects. Finally, (C) we want to offer some insights into the problem of simplicity in the context of mechanistic explanations. The topic has been introduced already by Harbecke, Grunau and Samanek (unpublished manuscript) without offering a workable solution. We now want to take up the challenge and offer two tentative solutions to the problem of simplicity of mechanistic explanations. One will propose a minimization of model structure in relation to the real-world system represented. The other will employ the notion of “understanding” as a crucial criterion for simple model selection. Both will be declared to involve an non-eliminable pragmatic ingredient.

13 - The neuroconnectionist research programme (in agenda)

Presenter: Tim Kietzmann, Osnabrück University

Format: Presentation and discussions

Time: 27th, afternoon

Deep learning is now the dominant paradigm in artificial intelligence (AI), but its roots lie in the connectionist movement of cognitive science. Now, deep learning is influencing another discipline, as deep neural networks (DNNs) are beginning to be used more widely in the computational neuroscience community, where they serve as powerful models to enrich our mechanistic understanding of brain function. This talk will present this emerging approach, which we call neuroconnectionism, as a cohesive large-scale research programme centered around ANNs as a computational language for expressing falsifiable theories about brain computation. It will describe the core of the programme, the underlying computational framework and its tools for testing specific neuroscientific hypotheses. Taking a longitudinal view, I will review past and present neuroconnectionist projects in the domain of vision and argue that the research programme is highly progressive (from a Lakatosian perspective), generating new and otherwise unreachable insights into the workings of the brain.

14 - Can Mechanistic Explanations of Cognition be Computational or Representational?

Presenter: Beate Krickel, Technical University of Berlin

Format: Presentation and discussions

Cognitive neuroscientists offer different types of explanations of cognition. This talk focuses on (1) mechanistic, (2) computational, and (3) representational explanations of cognition. Mechanistic explanations describe how neurons, networks of neurons, or brain regions interact to produce a particular cognitive or behavioral phenomenon. Computational explanations rely on abstract mathematical models of cognitive processes that specify the rules by which neural processing (may) operate. Representational explanations detail which neural activity patterns represent which items in the world. Often, these types of explanation appear in one and the same explanatory text (e.g., in research publications). This raises the question of how these types of explanation are related. According to proponents of the new mechanistic approach to explanation, the three types of explanation are not really different. Rather, computational and representational explanations are in fact special versions of mechanistic explanations. I will discuss a challenge for this view—the Compatibility Challenge—which seems to show that mechanistic explanations cannot be computational or representational. I will discuss strategies for dealing with the challenge and potential consequences for cognitive neuroscience.

15 - CogXAI: Cognitive Neuroscience Inspired Techniques For Explainable AI

Presenter: Sebastian Stober, Otto von Guericke University of Magdeburg

Format: presentation and discussion

Artificial Intelligence (AI) systems are very successful in various fields of application. In particular, Deep Learning (DL) models show impressive performance in, for example, computer vision and audio processing tasks. Their success, however, is mostly achieved by increasing the model complexity in terms of types of architectures or the number of neurons. This makes the models more opaque and severely complicates to understand how they make their decisions. Explainable Artificial Intelligence (XAI) aims to counteract this problem by developing methods for explaining decisions and internals of opaque models. In this talk, I will introduce our novel and more generally applicable neuroscience-inspired XAI techniques.