20th Annual Conference of the Italian Association for Cognitive Sciences

September 18-20, 2024

Rome, Italy

University of Rome "La Sapienza"

Villa Mirafiori, Via Carlo Fea 2, 00161, Roma

https://aisc2024.istc.cnr.it/

Local Organizing Committee

Giulia Andrighetto, Marco Fasoli, Emiliano Ippoliti, Luca Tummolini

September 18				
10:15-10:30	Room V Welcome and Institutional Greetings			
10.30-11:30	Keynote | Room V Beate Krickel **Individuating Cognitive Capacities in Terms of Cognitive Homology** TU Berlin Chair: Gustavo Cevolani			
11:30-12:45	Room VIII *Artificial Minds* Chair: Marco Marini	Room IX *Extended cognition* Chair: Marco Viola	Room II *Morality* Chair: Stefania Pighin	Room V *Explanation I* Chair: Davide Coraci
	Is hardware the body of artificial minds? The anthropomorphism of machines Cristiano Castelfranchi (Rome)	**Inside-out: thought-experiments, scientific simulations and the economy of extended cognition** Daniel Dohrn (Milan)	**Delegation of morality to AI: what tasks do we want to delegate?** Philippe Roman Sloksnath (Zurich, CH)	**How can the new mechanist philosophy accommodate degenerate mechanisms?** Yichu Fan (Edinburgh, UK)
	The code is willing, but the hardware is weak. The role of the body in shaping potential artificial consciousness Federico Zilio (Padova)	**Extended mind and diachronical personal identity** Fabio Patrone (L'Aquila)	**Morality, debunking, and diagnoses of irrationality** Alice Andrea Chinaia & Gustavo Cevolani (Lucca)	**Program-based Explanation** Nicola Zagni & Edoardo Datteri (Milan)
	Navigating creativity: comparing human and AI artistic processes Joachim Nicolodi (Cambridge, UK)	**Extended agency across smart and design based environments** Liberty Severs (Lisbon, PL) & Valeria Becattini (Berlin, DE)	**AI and social corrections: shaping norms against misinformation sharing** Eugenia Polizzi, Giulia Andrighetto & Carlo Ciucani (Rome)	**The rationality of mental imagery** Francesco Marchi (Bochum, DE)
12.45-14.00	LUNCH			

14:00-15:40	Room VIII SYMPOSIUM *Emotions, sociality and cognition* Organizer: Laura Barca	Room IX *Consciousness* Chair: Federico Zilio	Room II *Interaction I* Chair: Silvia Larghi	Room V SYMPOSIUM *Sustainable behavioral change for climate crisis* Organizer: Giulia Andrighetto
	Interoceptive grounding of conceptual knowledge Laura Barca (Rome)	**In search of embodied consciousness** Giulia Piredda & Laura Coccia (Pavia)	**Not only nationality: a multicultural perspective in human-robot interaction** Cecilia Roselli, Leonardo Lapomarda & Edoardo Datteri (Milan)	**Growing polarization around climate change on social media** Andrea Baronchelli (London, UK)
	Abstract concepts' vagueness: uncertainty and social interaction Anna M. Borghi (Rome)	**I am looking for a (permanent) center of gravity. How Daniel Dennett's prophecy about AI and the self has been, at least partially, confirmed and realized** Giacomo Romano (Siena)	**Hybrid embodied agency in human-AI interactions** Anna Ciaunica (Lisbon, PL) & Shaun Gallagher (Memphis, USA)	**Social tipping intervention to promote the adoption of reusable food packaging solutions** Gian Luca Pasin (Rome)
	Ingestible sensors highlight the relationship between stomach pH and virtual reality induced stress in healthy humans Vanessa Era, Arianna Vecchio, Sofia Ciccarone, Maria S. Panasiti, Giuseppina Porciello & Salvatore M. Aglioti (Rome)	**The role of touch in the sense of self formation** Iva Apostolova (Ottawa, CA)	**Enhancing trust in human-robot interaction: an integrated approach to knowledge representation in AI** Luca Biccheri & Roberta Ferrario (Trento)	**From business to society: a new framework for climate services** Marcello Petitta (Rome)

	Sex/gender differences in pathogen disgust and the nature-nurture debate Marco Tullio Liuzza & Giuseppe Occhiuto (Catanzaro)	**At first glance: investigating how vagueness influences verbal and non-verbal shared understanding of concepts among couples** Chiara De Livio, Claudia Mazzuca, Viola Chillura, Valerio Sperati, Anna M. Borghi (Rome)	**Structuring human-AI collaboration: an enactive framework for modelling heterogeneous cognitive systems** Julian Zubek, Łukasz Jonak & Joanna Rączaszek-Leonardi (Warsaw, PL)	**Widening the scope: the direct and spillover effects of nudging water efficiency in the presence of other behavioral interventions** Jacopo Bonan (Brescia)
15:40-16:05	COFFEE BREAK			
16:05-17:20	Room VIII *Interaction II* Chair: Eugenia Polizzi	Room IX *Explanation II* Chair: Cristiano Castelfranchi	Room II *Human Kinds* Chair: Giulia Piredda	Room V *Concepts & Emotions* Chair: Claudia Mazzuca
	Joint guidance: a capacity to jointly guide Marco Mattei (Milan)	**Is it a bug or is it a feature? Decisional enhancement, autonomy, and rationality in the digital age** Camilla Colombo (Aachen, DE)	**Questioning the boundaries of addiction** Davide Serpico & Francesco Guala (Milan)	**Investigating the influence of interoceptive accuracy on the classification of abstract and concrete concepts during pregnancy** Salvatore Diana, Anna Borghi & Laura Barca (Rome)
	(Dis)embodied joint agency in human-VR agents Interactions Altea Vanni, Sophia Bertoni, Shihan Liu, Jiaqi Yin, Sylvia Pan & Anna Ciaunica (Lisbon)	**The superbug: mental models and errors in computer programming** Silvia Larghi & Edoardo Datteri (Milan)	**Clarifying the muddle. Towards a comprehensive taxonomy of cognitive biases in medicine** Cristina Amoretti (Genova) & Elisabetta Lalumera (Bologna)	**DiffuseFace: a database of AI-generated face portraits of non-existing people to enrich diversity in face research** Alessia Firmani and Luca Cecchetti (Lucca)

	Mapping the psychophysiology of commitment Angelica Kaufmann, John Michael, Luke McEllin, Corrado Sinigaglia, Stephen Butterfill, Guido Barchiesi & Martina Fanghella (Milan)	**Generative AI and the overextended mind: on legal ownership of our cognitive extensions** Fabio Paglieri (Rome)	**AI in forensic evaluations: just smoke and mirrors or an incoming revolution?** Camilla Frangi, Alexa Schincario & Cristina Scarpazza (Padoa)	**Increasing emotional distancing with prism glasses: dissociated gender and adaptation direction effects on alexithymia in healthy individuals?** Laura Culicetto, Selene Schintu, Chiara Lucifora, Massimo Mucciardi, Alessandra Falzone, Carmelo Mario Vicario (Messina)
17:20-18.20	Keynote | Room V Stefano Nolfi **Integration and Transfer of Action and Language Knowledge in Learning Robots** Institute of Cognitive Sciences and Technologies National Research Council (ISTC-CNR) Chair: Anna M. Borghi			

September 19				
9:00-10:00	Keynote | Room V **Transforming Body Perceptions through the Senses: Innovative Neuroscientific Approaches and Applications** Ana Tajadura-Jiménez Universidad Carlos III de Madrid Chair: Luca Tummolini			
10:00-11:15	Room VIII *Cooperation* Chair: Antonella Tramacere	Room IX *Language* Chair: Aldo Gangemi	Room II *Perception & Action* Chair: Nicola Di Stefano	Room V *Trust* Chair: Edoardo Datteri
	Language-based game theory in the age of artificial intelligence Veronica Pizziol (Bologna), Valerio Capraro (Milan), Roberto Di Paolo (Parma) & Matja Perc (Maribor)	**Do neural language models have narrative coherence?** Alessandro Acciai, Lucia Guerrisi & Rossella Suriano (Messina)	**Action in multimodal object perception** Aleksandra Mroczko-Wasowicz & Spencer Ivy (Warsaw)	**From expert testimony to lay belief: a Bayesian view** Pietro Avitabile & Gustavo Cevolani (Lucca)
	The effect of heterogeneous distributions of social norms on the spread of infectious diseases Daniele Vilone, Eva Vriens & Giulia Andrighetto (Rome)	**How does sentence specificity shape uncertainty and curiosity in conversational dynamics?** Tommaso Lamarra, Caterina Villani, Claudia Mazzuca, Anna M. Borghi & Marianna Bolognesi (Bologna, Roma)	**Exploring inner speech influence on novel action acquisition and execution** Angelo Mattia Gervasi, Claudio Brozzoli & Anna Borghi (Roma, Lyon)	**Evaluating trust dynamics with dependency networks** Alessandro Sapienza & Rino Falcone (Rome)
	Tiny dictators: understanding altruism in young children Marco Marini, Sebastiano Munini, Michela Carlino & Fabio Paglieri (Rome)	**Lost in the labyrinths of stories: The role of negation and contradiction in LLMs' understanding of narratives** Emanuele Bottazzi & Roberta Ferrario (Trento)	**The evolution of syntax: toward a minimal model of hierarchical cognition** Giulia Palazzolo (Warwick)	**Artificial intelligence and institutional trust: promise or peril?** Ginevra Prelle (Milan)
11:15-11:40	COFFEE BREAK			

11:40-13:20	Room VIII SYMPOSIUM *Recent work in the epistemology of imagery and imagination* Organizer: Alfredo Vernazzani	Room IX *Modeling* Chair: Emanuele Bottazzi	Room II *Decision Making* Chair: Francesco Bianchini	Room V SYMPOSIUM *Multimodal integration between perception and action: cognitive, neural, and computational mechanisms* Organizer: Luca Tummolini
	Aphantasia, unconscious imagery, and rationality Joshua Myers (Barcelona)	**Ranking cognitive plausibility of computational models of analogical reasoning with the Minimal Cognitive Grid: results and implications** Alessio Donvito & Antonio Lieto (Bari)	**Truth approximation, calibration and bias in human judgment** Davide Coraci & Gustavo Cevolani (Lucca)	**Decoding haptic information and motor preparation in the early visual cortex** Simona Monaco (Trento)
	Maps of the imagination: a theory of artifact-based understanding Alfredo Vernazzani (Bochum, DE)	**The Minimal Cognitive Grid+, universal cognition and perceptual performance** Selmer Bringsjord, Paul Bello & James T Oswald (Albany, NY US)	**Unmasking stress: gender differences in decision making under mild hypoxia** Stefania Pighin, Alessandro Fornasiero, Marco Testoni, Barbara Pellegrini, Federico Schena, Nicolao Bonini & Lucia Savadori (Trento)	**Peripersonal space: a multisensory interface for the interaction between the body and the surrounding objects** Claudio Brozzoli, (Lyon, FR)
	Imaginative justification and imagistic reasoning Sofia Pedrini (Bochum, DE)	**Deductive flexibility in humans and beyond: testing the tool with synthetic datasets** Mariusz Urbanski, Paweł Łupkowski, Tomáš Ondráček & Ganna Stoyatska (Poznań, PL)	**Moral and social nudges for promoting cooperation in wicked social dilemmas: a theoretical and experimental investigation on waste sorting behavior** Sebastiano Munini, Marco Marini & Fabio Paglieri (Rome)	**Goal formation in multimodal space: a topological alignment approach** Francesco **Mannell**a, Julian Zubek & Luca Tummolini (Rome)

	The epistemic role of embodiment for imagination (and its lack in AI) Zuzanna Rucińska (Antwerp, BL)	**Evaluating Dream Semantics to discover patterns in personality traits and creative abilities** Aldo Gangemi, Chiara Lucifora & Claudia Scorolli (Bologna)	**Integrating VR and neuropsychometrics : Predicting consumer preferences via submental muscle activity** Francesca Ferraioli, Carmelo Mario Vicario, Chiara Lucifora, Viviana Betti, Matteo Marucci (Messina)	**From motor representations to language and back** Gabriele Ferretti (Bergamo) and Silvano Zipoli Caiani (Florence)
13:20-14:30	LUNCH			
14:30 - 16:00	Room V **The Synthetic Method in the Age of AI** **A symposium in honor of Roberto Cordeschi** Chair: Emiliano Ippoliti **From surrogative reasoning to surrogative simulation** Edoardo Datteri (Milan) **The synthetic method in cognitive robotics for interaction** Alessandra Sciutti (Genoa) **Synthesizing autonomy: from biology to robots and back** Vieri Giuliano Santucci (Rome)			
16:00 - 16:20	COFFEE BREAK			
16:20-18:00	Room VIII SYMPOSIUM *On the attribution of cognitive and emotional states to autonomous and intelligent systems* Organizers: Silvia Larghi, Marco Facchin & Giacomo Zanotti	Room IX *Perception* Chair: Marco Fasoli	Room II SYMPOSIUM *(Allegedly) AI-generated media: how do they make us feel?* Organizers: Dominique Makowski & Marco Viola	Room V SYMPOSIUM *Music perception and cognition: crossmodal, cross-cultural, and cross-species approaches* Organizer: Nicola Di Stefano
	How people understand robots' mind: folk-psychology vs. folk-cognitivism Silvia Larghi and Edoardo Datteri (Milan)	**Block on non-conceptual color perception** Ivan Cotumaccio (Paris, FR)	**Are androgynous faces uncanny?** Antonio Olivera-La Rosa (Medellín, CO)	**Crossmodal associations involving musical stimuli. Cross-cultural evidence** Nicola Di Stefano (Rome)

	Enactive intentionality in HRI: from attribution to detection Martina Bacaro (Bologna)	**Perceiving emotions: a multimodal approach** Niccolò Nanni (Lugano, CH)	**Emotional response toward fiction and the underlying cognitive mechanisms** Marco Sperduti (Paris, FR)	**Music perception and action: embodiment, dyadic dance, and interpersonal synchronization** Giacomo Novembre (Rome)
	Making emotional transparency transparent Giacomo Zanotti (Milan) & Marco Facchin (Antwerp, BE)	**Amodal completion as a means to perceptual beliefs** Hamza Naseer (Lugano, CH)	**Real is the new sexy** Alessandro Demichelis (Lucca)	**Rhythm and sound production across species** Andrea Ravignani (Rome)
	Substituting/complementing humans: a cognitive and affective analysis Guido Cassinadri (Pisa)	**Time experiences for survival** Antonella Tramacere (Rome)	**MusicAI bias: listeners like music less when they think it was performed by an AI** Alessandro Ansani (Jyväskylä, FI)	
18:00-19:00	Keynote | Room V Douglas Guilbeault **Discovering Cognitive Structure using Large-Scale Social and Artificial Intelligence** Stanford University Chair: Giulia Andrighetto			
19:00-19:15	AISC General Assembly			
20:30	SOCIAL DINNER Palazzo delle Esposizioni Roma			

September 20			
9:00 - 10:00	Keynote | Room V Rafael A. Calvo **Human Autonomy in an AI World** Dyson School of Design Engineering Imperial College London Chair: Marco Fasoli		
10:00 - 10:20	COFFEE BREAK		
10:20- 12:00	Room IX SYMPOSIUM *Ethical and cognitive perspectives on socio-technical hybrid societies* Organizer: Ludovica Marinucci	Room II (Anti)Representationalism Chair: Marco Facchin	Room V SYMPOSIUM *The social media debate: Do social media really represent a threat to our society?* Organizer: Alberto Acerbi
	Decision-making and self-control with AI in the loop Vieri Giuliano Santucci (Rome)	**Representational realism is not a tenet of cognitive science** Claudio Fabbron (Berlin, DE)	**The causal impact of Instagram usage on psychological well-being** Valerio Capraro (Milan)
	The ethics of using large language models to predict patients' preferences: a proposal Marco Annoni (Rome)	**The propositionalist view on emotion and its relevance for emotional attributions to robots in HRI** Ivan Zanzarella (Bari)	**Does the problematic use of social media constitute a pathological condition? Possible underlying psychobiological mechanisms** Tania Moretta (Padova)
	Can we get rid of empathy in AI-driven healthcare? Elisabetta Sirgiovanni (Rome)	**Grounding values in which environment?** Francesco Abbate (Rome)	**The skepticism puzzle: a critical examination of disinformation intervention effects** Folco Panizza (Lucca)
	Ethical framework for deception in human-robot interactions Ludovica Marinucci (Rome)		**The social media debate: a roundtable** Moderator: Alberto Acerbi (Trento)

12:00-13:00	Keynote | Room V AISC Young Researcher Prize 2023 **The Fox and the Grapes** **The Impact of Neuroimaging Data on Cognitive Ontology** Marco Viola Università degli Studi Roma Tre Chair: Fabio Paglieri

Social Dinner

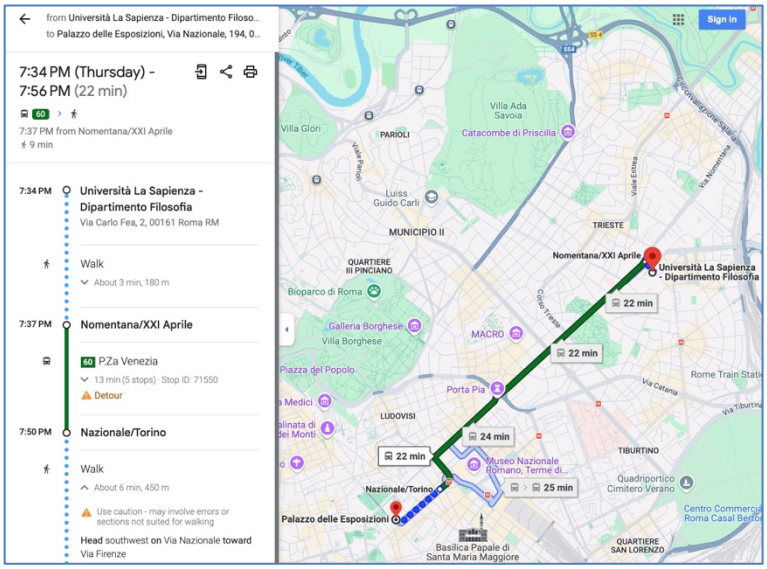
The social dinner will take place on Thursday, September 19 2024 at the Restaurant of Palazzo delle Esposizioni Roma in Via Nazionale 194 00184 Roma.

The cost of the social dinner is € 50.00 per participant. The amount must be paid in cash at the conference registration table at participants' arrival.

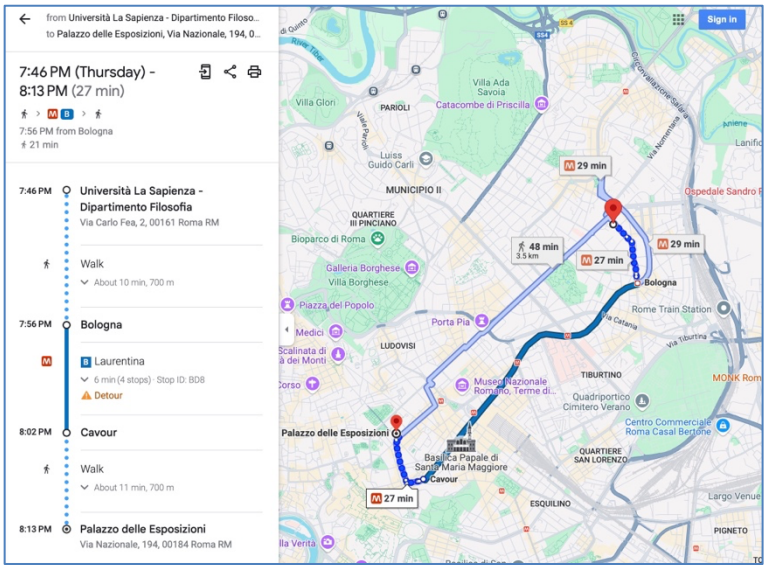
How to get to the Social Dinner?

Palazzo delle Esposizioni can easily be reached using the public transportation system.

The bus number 60 takes approximately 25 mins



The metro B (Bologna stop) takes approximately 30 mins



Keynote | Room V

Thursday September 18, 2024
10:30 - 11:30

Individuating Cognitive Capacities in Terms of Cognitive Homology

Beate Krickel
Institute of History and Philosophy of Science, Technology, and Literature
Technical University Berlin, Berlin, Germany

How should scientists carve up the cognitive domain to generate good predictions, explanations, and models of cognition? Based on join work with the philosopher and developmental psychologist Mariel Goddu, I argue that cognitive categories should be constructed the same way that biological categories are: in terms of homology. I will make use of a recent account of *Character Identity Mechanisms* (DiFrisco, Wagner and Love 2020) to make sense of the notion of "cognitive homology." The consequence of this notion is that brain structures and the organism's ongoing interactions with the environment turn out to be crucial for individuating cognitive homologies, and thus for individuating cognitive capacities.

Keynote | Room V

Thursday September 18, 2024
17:20 – 18:20

Integration and Transfer of Action and Language Knowledge in Learning Robots

Stefano Nolfi
Laboratory of Autonomous Robots and Artificial Life (LARAL)
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

The integration of action and language knowledge and skills is a pivotal element in the realm of human intelligence and stands as one of the most compelling challenges in scientific inquiry. In my presentation I will review the body of evidence and insights collected by attempting to design learning robots capable of understanding and using language and operating in a physical environment. I will particularly highlight the contribution of foundational models and the integration of passive observational learning and active embodied learning modalities. Furthermore, I will examine the merits of learning methods that foster the simultaneous development of diverse competencies indirectly by focusing on the optimization of a single learning objective.

Keynote | Room V

Thursday September 19, 2024
9:00-10:00

Transforming Body Perceptions through the Senses:
Innovative Neuroscientific Approaches and Applications

Ana Tajadura-Jiménez
i_mBODY lab, DEI Interactive Systems Group, Computer Science Department
Universidad Carlos III de Madrid, Leganés, Spain
2UCL Interaction Centre (UCLIC)
University of London, London, UK

Body perceptions are crucial for individuals' motor, social, and emotional functioning. Importantly, neuroscientific research shows that body perceptions are continually updated through sensorimotor information. This talk will showcase our group's research on how sensory feedback, particularly sound related to one's body and actions, can modify body perception, leading to Body Transformation Experiences. I will discuss how these findings contribute to the design of innovative body-centered technologies to address people's needs and support behavior change. Additionally, beyond such practical applications, these technologies serve as valuable tools for examining multisensory influences on body perception. Our ERC-funded project, BODYinTRANSIT, aims to establish a framework for individualized sensorial manipulation of body perceptions with long-lasting effects in everyday use contexts. The framework stands on four scientific pillars to induce, measure, support, personalize, and preserve body transformations: neuroscience of multisensory body perception; data modeling of the links between body perception, behavior, and emotion; wearable-based embodied multisensory interaction design; and field studies in real-life contexts with diverse user groups. Finally, I will identify challenges and opportunities in this research field.

Invited Symposium in honor of Roberto Cordeschi | Room V
Thursday September 19, 2024
14:30-16:00

The synthetic method in the age of AI

From surrogative reasoning to surrogative simulation

Edoardo Datteri
Università di Milano-Bicocca

The so-called 'synthetic method' is a form of surrogative reasoning, a term used in the philosophy of science to refer to the use of a model M (e.g. a robot) to acquire knowledge about the system T it represents (e.g. a living system). In recent years, a new use of (robotic) models has gained momentum, which can be called 'surrogative stimulation'. In surrogative stimulation, the model M is not used to learn about T, but to stimulate another system F in order to learn how the latter would react to T (the system represented by the model). The talk aims to clarify how surrogative stimulation differs from the synthetic method so thoroughly studied by Roberto Cordeschi, and how the two can be integrated, using examples from ethorobotics and social robotics.

The synthetic method in cognitive robotics for interaction

Alessandra Sciutti
Cognitive Architecture for Collaborative Technologies (CONTACT Unit)
Italian Institute of Technology

An important objective in current robotics is the development of robots capable of nuanced and effective human-robot interaction (HRI). Achieving this goal requires a deep understanding of human cognition, and robots can serve as ideal tools for this investigation. By constructing and programming robots, it is possible to test and model the dynamics of human interaction, gaining insights into human cognition through a synthetic and embodied approach. Drawing inspiration from the natural progression of human cognitive skills, a developmental perspective is adopted to design robots that can learn from their direct interactions with the environment and human partners. The integration of memory, motivation, and anticipation within a cognitive architecture enhances robots' social awareness and autonomous learning capabilities. This approach not only contributes to a deeper understanding of human cognition but also achieves the crucial technological goal of building machines that can dynamically adapt to individual human partners over time, fostering long-term collaboration and interaction.

Synthesizing autonomy: from biology to robots and back

Vieri Giuliano Santucci
Autonomy Research in Intelligent Systems and Ethics (ARISE)
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

The development of autonomous robotic systems has predominantly focused on enabling machines to independently complete predefined tasks. However, the emerging field of open-ended learning aims to push the boundaries of autonomy by creating systems capable of operating in unknown and unstructured environments without specific task assignments. In particular, the concept of Intrinsic Motivations (IMs), derived from animal and human psychology, is at the core of the development of a new typology of artificial agents capable of autonomously gathering knowledge and competences through the interaction with the environment. This line of research not only stresses the importance of the cognitive sciences for technological advancements, but also shows how robots and AI in general can be used as models of a feature that we consider essential of what it means to be human. Moreover, regardless of whether robotic autonomy can be equated to human autonomy, open-ended learning systems pose the critical issue of managing and aligning artificial agents that, to maintain the desired autonomy, cannot be pre-programmed or limited, even at their goal-setting level.

Keynote | Room V

Thursday September 19, 2024
18:00 – 19:00

Discovering Cognitive Structure using Large-Scale Social Data and Artificial Intelligence

Douglas Guilbeault
The Graduate School of Business
Stanford University

What can we learn about the structure of individual minds, human or artificial, using large-scale social data, such as the textual or visual data flowing through search engines and social media platforms? In this keynote, I present a diverse range of studies showing that large-scale social data can reveal striking insights into the mind, ranging from the structure of embodied cognition to the psychological biases that drive the formation of stereotypes. I will give special attention to presenting the results of a study we recently published in *Nature* which demonstrates how combining large-scale image and text data from online sources, analyzed via artificial intelligence, can reveal the latent multimodal structure of gender stereotypes. I will then share ongoing work that builds on these results by revealing the multimodal structure of intersectional stereotypes (e.g., gendered ageism) not only in human minds, but also in the judgments and associations formed by generative AI. Importantly, I will emphasize that big data and artificial intelligence are useful not only for testing existing theories about cognitive structure, but also for discovering and testing new theories. As an example, I will discuss ongoing work that harnesses this suite of algorithmic methodologies to unveil deep connections between the representational structure of gender and the concreteness and abstractness of concepts across domains, using visual and textual data, as well as behavioral outputs from AI. Opportunities for further advancing the integration of computer science, cognitive science, and cultural sociology will be discussed.

Keynote | Room V

Friday September 20, 2024
9:00 – 10:00

Human Autonomy in an AI World

Rafael A Calvo
Dyson School of Design Engineering
Imperial College London, UK

Human autonomy is a pillar of contemporary ethics and politics, particularly in liberal democracies like the UK, as well as in biomedical ethics. Psychological research robustly shows that a personal sense of autonomy is essential to wellbeing and sustained motivation. In technology design, such felt autonomy also underpins user adoption, engagement, and satisfaction. But today, human autonomy is coming under new threat by AI-driven technologies. Meanwhile, current AI research and policy, questions of safety, fairness, or explainability have received far more attention than how AI may impact autonomy – let alone how to design AI in an autonomy-supporting fashion. In this talk I will describe a vision of a socio-technical future where evidence-based and legitimate design and regulatory guidelines ensure that algorithmic environments safeguard and support human autonomy.

Keynote (AISC Young Researcher Prize 2023) | Room V

Friday September 20, 2024
12:00 -13:00

The Fox and the Grapes
The Impact of Neuroimaging Data on Cognitive Ontology

Marco Viola
Università degli Studi Roma Tre, Italy

In the early days of classical cognitive science, when the mind was often likened to a computer, cognitive theories developed largely without concern for the 'hardware'—the brain. Neuroscience was seen as irrelevant to psychological inquiry. However, this began to change in the 1990s with the rise of functional Magnetic Resonance Imaging (fMRI). Neuroscience started to play a significant role in shaping psychological theories, as researchers sought to map specific cognitive functions onto corresponding neural structures. Some proposed that an ideal neurocognitive theory would feature a perfect one-to-one mapping between functions and structures. However, such precise mappings have proven elusive. Instead of neat pairings, we find complex, many-to-many relationships. This raises an important question: how can we reconcile the ideal of one-to-one mappings with the current, entangled status of our knowledge? In this presentation, I will explore four (non-mutually exclusive) approaches that may help us refine our neuro-inspired Cognitive Ontology: (a) We may have chosen the wrong structures or functions, and a one-to-one mapping might be found with the correct selections; (b) The one-to-one mapping might be unattainable, and a probabilistic mapping could be a more realistic goal; (c) It's possible that the one-to-one mapping exists, but our concepts of 'functions' and 'structures' need to be redefined; (d) One-to-one mappings may exist, but they might need to be contextualized to specific circumstances.

Session: Artificial Minds | Room VIII
Wednesday September 18, 2024
11:30 – 11:55

Is hardware the body of artificial minds? The anthropomorphism of machines

Cristiano Castelfranchi (Rome)
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

It is true that in interaction with machines, we spontaneously and naively anthropomorphize them into artificial "beings". We ascribe to them and "recognize" human-like emotions; we think that they have them and interact on this basis. But in fact machines, robots, and software agents have no emotions, since they do not have a real body and therefore do not "feel" anything. The foundation defining character of emotions since William James is to "feel" something, that is, to perceive something from/in the body. And the body is a real body (and not a simple material support, hardware where the computational/cognitive and behavioral functioning is implemented) only if the agent feels it, that is, if there is interoception and proprioception. A robot still feels almost nothing from the body apart from sensorimotor signals for movement and therefore does not have a real "body" (it will have one when it can also feel pleasure and pain, and emotions). This ascription of emotions to machines is not only a spontaneous fallacy of ours, but is built on purpose to deceive us, to make us act "as if". It is an essential and increasingly widespread side of lying and deception used in HCI\HRI; and it is certainly effective and useful, like in care or in teaching relationships. This deception is partial, however. Robots can have the right “mind”: the beliefs and expectations that we attribute to them by reading that false emotion (e.g. that there is a possible risk/danger) and the purposes that we ascribe to it (avoid the danger); and it can have (act/simulate) the appropriate expressive features. Instead it is not true that:

- it does not have an “intelligence” and a "mind": system that elaborates representations of the world (propositional or mental images) and works on these representations (“mentally”) to solve problems and develop appropriate/effective conducts, not by trials and errors in the expensive physical world but representationally;
- that we naively attribute it to it and anthropomorphize it.

It is true that we anthropomorphize its intelligence by giving it a human mind, similar to ours; as we do between humans to be able to interact (cooperate, conflict) on the base of "mind reading". But it is not true that machines do not have a real "intelligence” (which does not mean “human”) and a form of mind (cognitive system for regulating conduct); and that by assuming it we superstitiously anthropomorphize them. What is happening is the opposite: thanks to AI scientific (not just technical contribution) finally we are de-anthropomorphizing psychological notions, making them more general (AGI conferences): why should intelligence or mind be an intrinsically and exclusively “human” concept? Moreover today this ascription of "our" way of thinking to the computer is less true: we perceive enormous differences in the capabilities of "generative" AI systems for acquiring and processing knowledge; their ability to learn, find relationships, generalizations, predictions, to decide better than us and more quickly.
We increasingly attribute to it a mind/intelligence that is increasingly different from ours.

Session: Artificial Minds | Room VIII
Wednesday September 18, 2024
11:55 – 12:20

The code is willing, but the hardware is weak.
The role of the body in shaping potential artificial consciousness

Federico Zilio
Università degli Studi di Padova

The rapid improvements in AI raise questions about its potential for phenomenal consciousness (i.e. valenced experience: there is something like it for the subject to be in a certain state). Critics argue that the lack of a physical body and world prevents the development of consciousness in AI (Susser, 2013). Indeed, theories of 4E cognition and non-reductionist neurophilosophical positions suggest that multisensory integration between environment, body, and brain is essential for consciousness. In this sense, without a body, AI may become (or remain) a highly efficient symbol/signal manipulator (Bender & Koller, 2020; Searle, 1980) or a human-like android, but without consciousness. Despite the intuitive power of this thesis, the definition of "body" for AI is ambiguous. Does it refer to real-time access to environmental data? Systems such as GPT-4o already have this capacity. Is it a physical structure for direct interaction with the environment? Situated robotics is already developing such artificial agents, equipped with mechanical bodies that can interact dynamically with the environment, manipulate objects in a changing context, and learn from imitation and mistakes. Or does the body imply being made of flesh? If so, what differentiates conscious from non-conscious states in a biological body? Given the above, I will: a) Discuss the relevance of the body to phenomenal consciousness (Zilio, 2022). b) Present potential body criteria for AI, selecting the most promising based on (a). c) Propose a hypothesis that, in principle, might allow AI to achieve some form of phenomenal consciousness through spatiotemporal alignment with the world (Golesorkhi et al., 2021; Northoff & Gouveia, 2024). However, the practical implementation of this remains unclear.

Session: Artificial Minds | Room VIII
Wednesday September 18, 2024
12:20 – 12:45

Navigating creativity:
comparing human and AI artistic processes

Joachim Nicolodi
Darwin College, University of Cambridge

The output generated by AI systems like DALL-E qualifies as art, no matter what definition of art one adopts – be it functionalist, institutionalist, or historical. It can evoke strong aesthetic experiences in the audience, is exhibited at prestigious galleries, and has a longer tradition than one might think, tracing back to Cohen's 1973 AARON program. In addition, it can create good art, as least from a purely formal perspective. By disregarding contextual considerations and focusing solely on the object itself – its linguistic properties in literature or its composition and colours in painting – AI-generated works can rival human masterpieces to the point where even experts struggle to distinguish between them (Lawson-Tancred, 2023). At the same time, we have strong intuitions that human art is somehow aesthetically superior (e.g., Bellaiche et al., 2023). The goal of this paper is to establish whether this intuition is correct, and whether we have reasons to regard human art as inherently superior to AI art. To answer this question, we have to establish whether AI is creative. The value of AI art hinges on the system's creativity because, as outlined above, AI art and human art are indistinguishable when considering their physical properties. If there is an aesthetic difference between the two, it has to be located not in the work as such, but in the cognitive capacity that went into creating it. As Coeckelbergh puts it, we need to shift our focus from the "external outcome" to the "internal workings of the machine, […] the process by which the art work is created" (Coeckelbergh, 2017, p. 288). However, Coeckelbergh quickly becomes entangled in definitional disputes about creativity, arriving at the unsatisfactory conclusion that "machines are here to stay, and so is the mystery of art and creativity" (Coeckelbergh, 2017, p. 302). Importantly, it is possible to have a science of creativity and analyze its underlying mechanism without perfect definition, as long as we focus on "prototypical" cases of creativity and avoid more controversial cases (Chen, 2018). With this in mind, I will answer the issue of AI creativity by adopting a comparative approach, similar to the one outlined by Halina (2021). In summary, I will explore the processes underlying artistic creativity in humans, turn to the workings of potentially creative AI systems like DALL-E, and then compare the two. I propose a framework of creativity based on Wallas' hugely influential four-phase model, which distinguishes between preparation, incubation, illumination, and verification. However, in addition to Wallas' behavioural description, I will incorporate more recent neuroscientific evidence to determine how and why these phases often lead to creative insights. Then, I will examine how DALL-E (and the GPT-series more generally) is trained and how it creates putatively creative output like poems and paintings. In the final part, I will compare the two to see if AI fulfills what I call the "human standard of creativity". A sober analysis reveals remarkable parallels between human and artificial systems, suggesting that the key issue is consciousness. I conclude that artistic creativity requires phenomenal consciousness, which AI lacks. However, AI may still exhibit other forms of creativity where phenomenological content is not necessary.

Session: Extended Cognition | Room IX
Wednesday September 18, 2024
11:30 – 11:55

Inside-out: thought-experiments, scientific simulations and the economy of extended cognition

Daniel Dohrn
Università degli Studi di Milano

I introduce a new alternative for extended cognition. I build on an intuitive demarcation: Narrow processes take place within an intuitive boundary given by 'skin and skull' (Clark and Chalmers 1998), while broad processes at least partly go beyond that boundary. There is a tendency within cognition towards making it more efficient by mutually substituting narrow and broad processes. There are two directions of substitution, outside-in, and inside-out. I illustrate these two directions by an especially significant example: thought-experimental simulation.

Outside-in: Some thought-experimental simulations manifest a tendency to temporally replace a paradigmatically broad process, the process of empirical (or material) experimenting, by a narrow imaginative process that is more readily available, thought-experimental simulation. The thought-experiments at stake work by imaginatively simulating the manipulation of experimental set-ups. Imagination orchestrates the targeted 'off-line' re-use of cognitive mechanisms that would be used 'online' in performing empirical experiments. Classical examples are Galileo's experiment of dropping objects of different weights from the tower of Pisa or the Einstein-Podolsky-Rosen (EPR) experiment in quantum mechanics.

Inside-out: some thought-experimental simulations manifest a tendency to expand the originally narrow process of thought-experimenting so as to implement some part of it on external devices, as in computer simulations that perform the same task more efficiently. Many scientific simulations, in particular computer simulations, involve replacing a narrow cognitive process that forms part of thought-experimenting by a broad process that is more effective in generating the results of a thought-experiment. The same cognitive process of thought-experimenting can be fully implemented as a neural process or partly as an in-silico process. An example of a thought-experiment that could in principle be performed imaginatively but is much more effectively run on a computer are cellular automata like Conway's game of life. The EPR experiment provides an example of both the outside-in and the inside-out direction: a proxy for an empirical experiment was run in the imagination, but it ramified into results like Bell's inequalities that were tested by computer simulations.
I propose a sufficient condition of extended cognition that covers the two directions:

REPLACEMENT:

Some process P is an extended cognitive process if either

(i) P is narrow and P replaces a broad cognitive process Q in approximating the outcome of Q,

or

(ii) P is broad and P replaces a narrow cognitive process R in approximating the outcome of R.

REPLACEMENT applies to my exemplum crucis of thought-experimental simulation. The first (i) disjunct covers my outside-in direction, the second (ii) covers my inside-out direction.

Session: Extended Cognition | Room IX
Wednesday September 18, 2024
11:55 – 12:20

Extended mind and diachronical personal identity

Fabio Patrone
University of L'Aquila

In this talk, I explore the consequences of the famous extended mind thesis (Clark & Chalmers, 1998) on personal identity. What happens to our persistence through time and change if we have an extended mind? I aim to show that whoever has an extended mind is an extended person, namely a mereological composite who can persist both in physical and virtual worlds. Clark and Chalmers suggest that persons extend beyond the boundaries of their physical bodies. That is to say that beliefs can be constituted partly by environment features, like a notebook or a smartphone. The parts of the external world we extend through are considered a fundamental aspect of our identity. According to their account, persons transcend classical Cartesian dualism. If we are extended minds, and, as C&C suggest, extended selves, then we are composed by mind, body, and extensions of our mind. I discuss Olson's (2011) view, which states that the extended mind does not imply the extended self unless persons are reduced to humean bundles of mental states. I show that his argument is unsound, short of embracing a radical physical criterion of personal identity (Olson 2007), since our ontological stance about persons should take into consideration what Floridi (2014, 2022) calls onlife world and the position expressed by Chalmers (2022) about the nature of virtual worlds. I maintain that the extended self brings a new concept of person, namely "extended" person. C&C propose a thought experiment, in which a notebook is a repository of the beliefs of Otto, a person suffering from Alzheimer's disease. I aim to rethink classical inquiries into the nature of diachronic personal identity and material constitution through the lens of Otto's thought experiment. In fact, if Otto misplaces his notebook, it evokes a scenario reminiscent of amnesia; if he lends it to another person, it mirrors the metaphysical complexities of brain transplantation; if he creates a duplicate, it resonates with the intricacies of fission I believe that the concept of extended person sheds a new light on those thought experiments, as if the subject of those scenarios is an extended person their conclusions appear less problematic. If we accept that persons can extend their minds in the outside world, then it is reasonable to accept that they can share parts and that the extended parts of their minds can be copied, transferred, and so on. In this case, there is no paradoxical threat as long as we talk about the extensions of our minds. But, if we genuinely embrace the C&C thesis, it appears there is no ontological difference between information written in Otto's notebook and his memories.

Session: Extended Cognition | Room IX
Wednesday September 18, 2024
12:20 – 12:45

Extended agency across smart and design based environments

Liberty Severs and Valeria Becattini
CFCUL, Universidade de Lisboa; Berlin School of Mind and Brain

In many contexts, we adopt and exploit the external resources within our sociomaterial environment in order to facilitate or enhance our cognitive abilities (Clark, 2008; Clark and Chalmers, 1998; Tomczyk, 2023). However, a more systematic understanding of the dynamics at play within extended regimes of agency is currently lacking. In this paper, we develop an account of ambient smart-, and design-, based environments as scaffolding forms of extended agency. We first characterise agency (i.e., a capacity associated with an entity or system's self-generated activities) as a distributed process that involves (and is often enhanced by) other cognitive systems, which may include individuals, tools and intelligent, material artefacts (Preston, 2013; Malafouris, 2013; Sterelny, 2010). We then draw on aspects of complex systems, active inference and skilled intentionality in order to define agency as scaf olded allostatic control, whereby the environment scaffolds additional strategies for exerting or optimising this allostatic control across relevant temporal scales (Rietveld & Kiverstein, 2014; Sterelny, 2010; Friston et al., 2012). We aim to show how these environments play a key role in regulating our ability to generate and control our actions and their (predicted) consequences. In this sense, we claim that an individual's agency can be supported (or degraded) by the scaffolding milieu or environment. Having analysed the concepts of landscapes and fields of affordances alongside habitual and goal-directed behaviour in living systems, we highlight differences between ambient smart-, and design-, environments in terms of the opportunities for action (i.e., affordances) that they provide. We emphasise the relevance of temporal scales for delineating these differences. This allows us to account for a broad range of behaviour that is characteristic of both optimal and suboptimal) states of the organism within their sociomaterial environment, which we expound from the perspective of active inference, complex systems and subjective wellbeing. In section 1, we introduce ambient smart environments (ASE), and, in line with White and Hipolito (2023), we argue that ASE can be beneficial (though, as we argue, also potentially equally detrimental) to our mental health, with particular reference to the way these technologies can scaffold adaptive behaviour within an individual's affordance landscape. However, we point out that ASE by itself may not be sufficient for improving one's well-being. In section 2, we introduce design environments (DE) as a potential supporting component. We evaluate how relevant is the role of the sociomaterial environment on scaffolding both i) specific fields of affordance, and ii) the affordance landscape more generally. Our aim is to show that DE contributes fundamentally to scaffold adaptive (habitual) behaviour within an individual's field and landscape of affordances. Finally, section 3 shows how our account allows us to approach extended agency within everyday contexts, across both shorter- and longer- timescales.

Session: Morality | Room II
Wednesday September 18, 2024
11:30 – 11:55

Delegation of morality to AI: what tasks do we want to delegate?

Philippe Roman Sloksnath
University of Zurich

If a human operator delegates a task to an autonomous agent, while constantly supervising the performance and taking over control when needed, little additional problems will occur. However, more and more AI systems go beyond acting as human proxies while being empowered to make their own decisions (Candrian & Scherer, 2022). Thus, we might find ourselves in situations in which this dichotomy vanishes, maybe even to a point where human supervision is neither needed nor wanted (Gogoll & Uhl, 2018). Nevertheless, innovation in AI has promoted human-robot interaction (HRI) in various domains (Kneer, 2021). While some usage hardly affects moral domains, such as navigation or manufacturing, others clearly impose the possibility of giving moral agency to a machine (e.g.: self-driving cars, search and rescue missions) (Kneer, 2021; Nyholm, 2018). For example, the use of AI-driven systems in military weapons raise serious concerns as decisions possessing life-and-death consequences may no longer be made directly and entirely by humans (Coglianese & Lehr, 2017) With increased task delegation to robotic agents in various domains, it will undoubtedly become reality that robots make morally relevant decisions. Therefore, it is important to address the question of what tasks we want AI-agents to perform, what outcome is preferred and what psychological factors foster delegation of morality. Subjects (n = 269) chose in 31% of the situations to delegate a moral decision in trolley style dilemmas to another agent, either human or robot with AI. Interestingly, people do differentiate between the scenarios when considering delegating to another human and do less so when considering delegating the decision to a robotic agent. Marginally significant more subjects delegated the decision in the footbridge case to the robotic agent compared to the human agent ($p = 0.05$, cohen's $\gamma = .118$), indicating a low to moderate effect of agent type in the footbridge case, although further studies will be conducted to prove these findings. Furthermore, it is overall not morally acceptable to delegate the decision, while it is significantly less acceptable to delegate the footbridge dilemma to another human compared with the switch dilemma ($t = 3.2732$, $df = 130$, $p < 0.01$, cohen's $d = 0.286$). Further studies will be conducted to consolidate these initial findings and enhance the framework covering a spectrum of tasks, addressing the inclination towards personal execution or delegation of moral decisions and actions. Moreover, this framework will delve into the preferred entities for delegation, whether human or AI.

Session: Morality | Room II
Wednesday September 18, 2024
11:55 – 12:20

Morality, debunking, and diagnoses of irrationality

Alice Andrea Chinaia and Gustavo Cevolani
IMT School for Advanced Studies Lucca

Debunking arguments aim to weaken beliefs by questioning their justifications. A prominent case involves the debunking of moral beliefs on theoretical or empirical grounds, such as Joshua Greene's critique of deontological moral norms. Building upon neuroscientific, psychological, and behavioral evidence, Greene presents deontological judgments as merely post-hoc rationalizations of innate intuitions, rather than the results of a process of moral reasoning. In this paper, we assess debunking arguments of this kind from a philosophy of science perspective, using Greene's account as a case-study. First, we propose to reconstruct moral arguments as instances of practical reasoning of the following form P: "X is good. Doing Y is a means to make X happen. Then, I (should) do Y." Second, we argue that the debunking strategy relies on diagnoses of irrationality against P-arguments of such form. Third, basing our analysis on existing literature, we study how and when such diagnoses of irrationality, and the resulting debunking arguments, are sound. To be so, at least three assumptions must be in place: i) the relevant moral norm must be correctly applied to the specific scenario to adequately sustain the premise "X is good"; ii) the relevant premises of the scenario must be shared between the experimenter and the participants; iii) the responses of the latter must be correctly interpreted by the former. Our analysis allows both for a rational reconstruction of the debate about morality and for a clearer assessment of the prospects and limitations of the debunking of moral arguments.

Session: Morality | Room II
Wednesday September 18, 2024
12:20 – 12:45

AI and social corrections: shaping norms against misinformation sharing

Eugenia Polizzi, Giulia Andrighetto and Carlo Ciucani
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

Traditional approaches to combating misinformation often focus on changing the behavior and beliefs of potential norm transgressors. However, these methods may inadvertently backfire. An alternative strategy involves motivating the larger share of users who observe fake news to publicly react against it, by strengthening the social norms that regulate responses to norm violations. Social corrections—corrective comments posted by users in response to fake news—are a visible form of punishment and hold the potential to serve as norm-nudging tools. Witnessing peer-punishment can update bystanders' perceptions of the prevalence and social approval of sanctioning behavior within a community (meta-norms). Stronger meta-norms, in turn, can make “would-be” enforcers more likely to act, legitimizing their behavior and reducing the fear of retaliation. In support of this hypothesis, our previous research has provided experimental evidence that social corrections can signal normative information and motivate observers to engage in corrective actions when encountering fake news posts (study 1). Given the ubiquitous presence of social bots on social media platforms and the significant advances in large language models (LLMs) that enable artificial agents to exhibit human-like features, a critical question is whether AI agents can similarly communicate normative information in human-AI hybrid systems. This aspect is key to understand how dynamics of social behavior change under increasingly digital social environments. If AI agents behavior can influence what is perceived as acceptable or expected, they could play a significant role in guiding social interactions and promoting certain values or norms. In the context of misinformation, "correcting" bots could serve as catalysts for behavior change, reducing the social cost of norm enforcement. To address such question we conducted an online experiment (Study 2) building on the methodology used in Study 1. In Study 1, participants engaged in discussions resembling an online forum, where they were free to comment on posts shared by prior users. We experimentally manipulated whether participants could observe corrections left by other users in response to posts featuring inaccurate information (experimental group) or not (control group). We then examined the effect of this variation on both behavior—the likelihood of replying with a corrective comment—and norms—the perceived social appropriateness of correcting inaccurate posts. In Study 2 we manipulate whether social correctors are depicted as humans or bots, by altering the cues indicating their identity (e.g., nickname) while keeping the content of the comments fixed. If observation of bots corrective behavior is interpreted as normatively relevant, we should expect an increase in both the number of corrective replies to fake news posts and participants' normative expectations compared to the control group. If bots are perceived merely as tools rather than social agents with normative standing, we should expect no significant differences in normative expectations compared to the control group. Data collection is currently ongoing, with preliminary results expected by early summer 2024. Insights from this research could inform the design of interventions to combat misinformation and enhance our understanding of AI's potential contribution to social regulation in digital spaces.

Session: Explanation I | Room V
Wednesday September 18, 2024
11:30 – 11:55

How can the new mechanist philosophy accommodate degenerate mechanisms?

Yichu Fan
University of Edinburgh

Degeneracy, the ability of structurally different elements to perform the same function and give rise to the same phenomenon, is believed to be ubiquitous at all levels of mechanisms in neurobiology (Edelman&Gally, 2001; Rathour& Narayanan, 2019). Given its biological salience, degeneracy has become an emerging topic in recent scientific literature as well as philosophical discussions on robustness in biology (e.g. Mitchell, 2008; Chirimuuta, 2017). However, it has received little attention from the new mechanist philosophy. In this paper, I want to start the discussion by assessing the implications of degenerate mechanisms for accounts of mechanistic explanations and constitutive relevance. In particular, I will argue that the feature of degeneracy in neural systems poses a dilemma for mechanistic explanations: at a certain level of description, mechanistic explanations will have to lose either their generalisability or their decomposability before reaching the 'bottom-out' solution. To reconcile this dilemma, many scientists opt for what I call *population mechanistic explanations*, where a population of models are constructed using techniques such as evolutionary algorithms to account for the phenomenon (cf. Marder&Taylor, 2011). However, I argue that if population mechanistic explanations represent single mechanisms, as suggested by its use in scientific practice, the new mechanists might need to grant the constitutive relevance of factors that only *potentially* contribute to the phenomenon in some contexts. Specifically, I will start by pointing out that many accounts of constitutive relevance either explicitly or implicitly require the components to be necessary or non-redundant for the mechanisms that they constitute (e.g. Harbecke, 2010; Couch, 2011; Baumgartner& Casini, 2017). The way these accounts deal with degenerate mechanisms is to assume that different components constitute alternative types of mechanisms for the same phenomenon. However, I argue that this way of individuating mechanism types is not descriptively adequate: scientists generally assume that the same mechanism can employ different factors in different contexts, and that non-redundant/non-degenerate mechanisms are fragile and hence biologically implausible. Moreover, given the pervasiveness of degeneracy, differentiating mechanism types fully based on component types makes mechanistic explanations ungeneralisable, which threatens the possibility of type-level mechanistic explanations. To secure generalisability, mechanistic explanations have to stop decomposition at a fairly high mechanistic level, forestalling the new mechanist ideal of 'bottomed-out' mechanistic models. To offer generalisable explanations and further decompose degenerate systems, many scientists promote what I call the *population mechanistic explanations*, where a population of models are constructed using machine generated data to capture the real-life variability of biological mechanisms. By examining examples from action potential and central pattern generator studies, I will show why population mechanistic explanations qualify as mechanistic models, and why the multiple models should be considered as representing one single degenerate mechanism instead of many alternative mechanisms. Further, I argue that the use of population mechanistic explanations entails a broader notion of constitutive relevance which include factors that *potentially* contribute to the phenomenon. However, I conclude by noting that this might lead to unwanted consequences for the new mechanists – now the boundaries of mechanisms might only be restricted by pragmatic considerations.

Session: Explanation I | Room V
Wednesday September 18, 2024
11:55 – 12:20

Program-based explanation

Nicola Zagni and Edoardo Datteri
Università di Milano-Bicocca

A robot is moving in a living room. Eventually it starts going towards a trash bin, following a straight path. When its distance from the trash bin is approximately 10 centimeters, it steers right with an angle of 90 degrees. Why did it display this behaviour? One of the many possible explanations states that the robot implements a program according to which, whenever the robot perceives an obstacle at a distance equal or minor than 10 centimeters, it steers 90° right. In this explanation, the explanandum is a particular behaviour displayed by the robot, and the explanans includes the description of a program. The thesis that programs can play a crucial role in the explanation of behaviour - of living and non-living systems, like the robot in this example - has been discussed to some extent in the early age of cognitive science. Robert Cummins (1977) argued that programs can and do explain behaviour in a way that is independent from physiological explanation. Johnson-Laird (1983) claimed that psychological theories can have explanatory value only if they could be formulated in algorithmic terms. Pioneers of Artificial Intelligence Newell and Simon (1958) theorized on human problem solving in terms of programs. Notwithstanding the importance attributed to program-based explanation (PbE) by these and other scholars, the philosophical debate about its structure and explanatory value has faded through the years - even though the notion of ‘program’ is frequently invoked in biological and technological explanations. Concurrently, a growing interest has emerged on the structure of mechanistic and computational explanation (Piccinini, 2007). However, it is not self-evident that PbE is a form of mechanistic explanation; nor that programs can really enable one to explain, or even understand, human and artificial behaviour. The aim of this paper is to explore the relationship between mechanistic and programbased explanation, with a particular focus on the explanation of the behaviour of robots in everyday contexts. In recent decades, the concept of mechanisms has undergone various definitions. Nonetheless, there is a consensus among most scholars regarding the definition proposed by Illari and Williamson (2012, p.120), which states that “a mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon.” Accordingly, explanation entails identifying the mechanism accountable for the observed phenomenon. Recent scholarly literature often makes a distinction between mechanism descriptions, which constitute complete mechanistic explanations, and mechanistic sketches, which may contain gaps and ’black boxes’ (Craver, 2006). How does the notion of ’program’ correlate with this understanding of mechanisms? First, we will examine whether programs should be classified as mechanism descriptions or mechanism sketches. Additionally, we will challenge the assertion that if programs are considered “mechanism sketches” they are incapable of providing explanations. Secondly, we will try to locate entities and activities in a simple Python program for a robot. Should entities, or descriptions thereof, be identified with, say, constants, functions, or objects in object-oriented programming? Should activities, or descriptions thereof, be identified with control structures such as sequences, loops, conditionals? Thirdly, we will focus on the notion that the entities and activities of a mechanism are “responsible for the phenomenon”. Mechanisms are said to bear three kinds of relationships to phenomena (Craver and Darden 2013): they can either underlie, produce or maintain them. What sort of relationship does a program bear to the behaviour of the system? We will argue that these questions admit no easy answers, and that the standard definition of ‘mechanism’ is not fully adequate to capture the idea that programs can be regarded as mechanisms. By addressing these questions, the paper intends to contribute to the analysis of the role played by programs in our explanation and understanding of living systems and technological artifacts.

Session: Explanation I | Room V
Wednesday September 18, 2024
12:20 – 12:45

The rationality of mental imagery

Francesco Marchi
Ruhr University of Bochum

In this article, I shall focus on the sizeable chunk of our imaginative life that comprises perception-like imaginative states. If I ask you to imagine being at a crowded party, and if you have a fervid imaginative capacity, you may visualise people dancing to an imagined tune, all while faintly smelling imagined food from a rich imagined buffet and sipping from a tasty and fresh imagined cocktail in your hand. These multisensory imaginative scenarios are mostly comprised of mental images. Mental imagery is the faculty of imagination closest to actual perception and the inquiry on the similarities and differences between the two has been one of the major longstanding threads in the study of imagination. A mental image is a peculiar kind of mental state. On the one hand it is a representational state with phenomenal character (Nanay, 2023), which situates it really close to ordinary perceptual experience in the geography of human mental life. On the other hand, as I argue below, it behaves like belief in significant respects. For example, it can be actively formed and sustained, and it can be informed and revised by reasoning and deliberative processes relying on previous and new knowledge. In virtue of having phenomenal character, it seems natural that mental imagery would share any epistemic role that phenomenal character is thought to confer to perceptual experience, for example in terms of justificatory power, according to some prominent views in the epistemology of perception. Furthermore, in virtue of behaving like a belief, it would seem that whatever epistemic force mental imagery has, it would be quite widespread and that it plays an important epistemic role for a subject's rational economy. But then it is puzzling that the epistemic role traditionally assigned to imagination, of which mental imagery is one of the core manifestations, by a longstanding tradition in philosophy (Sartre, 1948; Wittgenstein, 1981; O'Shaugnessy, 2000) is usually thought to be limited. This idea is so widespread that it can be regarded as a philosophical truism (Kind, 2016). Imagination has been accepted at most as an epistemic guide to metaphysical possibility. This means that imaginative states, including mental images, can only provide justification to beliefs about what is (metaphysically) possible (McGinn, 2004; Gendler and Hawthorne, 2002). One of the main reasons behind the view that imagination cannot justify beliefs about the actual world is what has been called the up-to-us challenge (Balcerak Jackson, 2018), according to which the problem with the epistemic status of imagination is that imagination is up to us, in the sense that it is under voluntary control, whereas outstanding sources of justification for beliefs, be them perceptual, testimonial or otherwise, should presumably be independent from volition. Recently, however, several authors have argued that the justificatory power of imagination may extend beyond the domain of mere possibility (Kind, 2016; Williamson 2016). Here, I will follow this line of thought and argue that mental imagery has a more significant epistemic role to play than what is usually ascribed to imaginative capacities.

SYMPOSIUM: Emotions, sociality and cognition | Room VIII
Organizer: Laura Barca

Wednesday September 18, 2024

14:00 – 14:25

Interoceptive grounding of conceptual knowledge

Laura Barca
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

In the last decades, interoception, the 'visceral dimension of embodiment', has been recognized as a fundamental element for the self and the human mind – well beyond its pivotal role in the sophisticated regulatory dynamics of physiological processes and energy needs. In my presentation I will focus on a particular role that interoception plays for cognitive processes - namely, in shaping conceptual representations. I will discuss empirical findings elucidating the malleability of the boundaries between different concepts, particularly emotional ones, and how their conceptual representation is influenced by individual characteristics of affectivity, and psychological stressor. Two-dimensional (affective valence and physiological arousal) topographical maps of emotional concepts, gathered from a similarity judgement task, offers a graphical representation of affective knowledge revealing gender and age-related difference along the arousal dimensions. To assess how interoception affect conceptual representations, we have developed an interoceptive exteroceptive categorization task of concrete (natural, artefact) and abstract (emotional, philosophical) concepts – implemented using the mouse-tracker software. Movement trajectories revealed the implicit activation of interoceptive features during the categorization of concrete-natural concepts, thus beyond the abstract-emotional ones. Since people greatly vary in their ability to attend to their bodily signals, we also measured participants' interoceptive accuracy with a cardioception task (heartbeat counting task). Participants more sensitive to their heartbeat were faster, particularly in the (exteroceptive) categorization of concrete-natural concepts. Overall, our results highlight the multiplicity of dimensions involved in conceptual knowledge, including the interoceptive ones.

14:25 – 14:50

Abstract concepts' vagueness: uncertainty and social interaction

Anna M. Borghi
Sapienza University of Rome
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

Abstract concepts, expressed by abstract words (e.g., "phantasy"), are frequently used in adult speech. This is striking because they are quite complex; unlike concrete concepts (e.g., "chair"), they do not refer to a single referent and assemble objects and entities that are perceptually dissimilar. In addition, their meaning is highly variable both within and across participants. These characteristics of abstract concepts render social interaction particularly crucial. We recently proposed the notion of social metacognition, which underlines the importance of inner monitoring and social interaction for abstract concepts. People need more support from others to acquire abstract concepts since they can rely less on environmental information. In addition, with abstract concepts, people strive more to reach common ground with others and sometimes co-build meaning with them. In the presentation, I will overview recent studies performed in our lab showing that when people process abstract concepts, they experience uncertainty and feel the need to rely on others but, at the same time, do not fully trust either their own knowledge or the knowledge of experts. I will then argue that abstract concepts differ in their vagueness/indeterminacy and that negotiation with others is particularly prominent with more vague abstract concepts. I will then illustrate the preliminary results of a recent study testing whether abstract concepts differing in their degree of indeterminacy/vagueness elicit different patterns of social interaction, assessed both through an analysis of the conversation pattern and the exchanges of mutual gazes.

14:50 – 15:15

Ingestible sensors highlight the relationship between stomach pH and virtual reality induced stress in healthy humans.

Vanessa Era, Arianna Vecchio, Sofia Ciccarone,
Maria S. Panasiti, Giuseppina Porciello,
Salvatore M. Aglioti
Department of Psychology
Sapienza University of Rome

Stressful situations elicit a cascade of psychophysiological responses, involving changes in various systems including the enteric nervous system. While it is a common experience that stress is accompanied by intense gastrointestinal (GI) symptoms such as motility disturbances, and visceral hypersensitivity, objective evidence demonstrating the impact of stress on enteric functions in humans is scarce. This knowledge gap is primarily due to challenges in monitoring GI activity, which traditionally relies on invasive and expensive methods. To deal with this issue we asked 36 healthy participants to ingest sensors-equipped, biocompatible, non-invasive and minimally intrusive pills able to transmit pH, pressure, and temperature along the GI tract. These three parameters were acquired while participants were immersed in a validated, psycho-socially stressful virtual reality (VR) scenario or in a non-stressful one and when the pill was in the stomach or in the large intestine. Our protocol allowed us to probe GI markers of acute stress-related responses. Results show that the subjective perception of stress in VR scenarios was associated with less acidic gastric ph. These findings are likely linked to heightened sympathetic activity associated to high perceived stress that leads to a suppression of gastric secretion. Our innovative approach holds significant promise for advancing research on the physiology of stress-related responses.

15:15 – 15:40

Sex/gender Differences in Pathogen Disgust and the Nature-Nurture Debate

Marco Tullio Liuzza and Giuseppe Occhiuto
Department of Medical and Surgical Sciences
University “Magna Graecia” of Catanzaro

Disgust is a primary emotion that has likely evolved to help organisms avoid harmful pathogens. Notably, research indicates that women are generally more prone to experiencing pathogen disgust than men. Evolutionary theory, supported by comparable sex differences in other mammals, suggests this may stem from different evolutionary pressures faced by the two sexes, possibly linked to the unequal investment required in offspring rearing. In contrast, sociocultural explanations suggest that patriarchal societal structures might lead to women expressing more disgust, as it is culturally instilled from a young age. To examine this sociocultural hypothesis, we analyzed whether the variation in disgust sensitivity between sexes correlates with levels of patriarchy in various countries. We drew on existing data regarding individual differences in pathogen disgust from 31 countries and compared these to the factor scores from four correlated gender equality indices. Through a Bayesian multilevel linear model analysis of individual responses, we investigated whether an interaction between sex and gender equality scores could predict pathogen disgust sensitivity. Our findings revealed a higher propensity to disgust in women that is consistent across different levels of gender equality. If anything, the differences were slightly more pronounced in countries with higher gender equality, a pattern consistent with the so-called gender equality paradox. This paradox does not align with the theory that patriarchal culture is a primary driver of disgust sensitivity differences and may lend support to evolutionary explanations for the observed sex differences in pathogen disgust.

Session: Consciousness | Room IX
Wednesday September 18, 2024
14:00 – 14:25

In search of embodied consciousness

Giulia Piredda and Laura Coccia
Istituto Universitario di Studi Superiori di Pavia

Although the study of consciousness has registered a big development in the last decades, consciousness remains one of the most mysterious and intriguing phenomena in philosophy of mind. On the one hand, the classical discussions between physicalists and anti-physicalists around the hard problem of consciousness are still there. On the other hand, new trends have arisen. Among them, panpsychism has found many supporters over the last years, as a peculiar reaction to the hard problem (Goff, Moran 2022). Furthermore, within the 4E cognition framework, some philosophers have tried to explain the phenomenal character of consciousness in externalist terms, defending the extended consciousness thesis (Telakivi 2023). This latter proposal, articulated in different manners, appears at least as radical as the panpsychist one. Nevertheless, it is unclear whether engaging in such radical approaches is really justified. In spite of the initial attractiveness of panpsychism, substantial problems seem to persist, when one tries to justify the thesis in detail (e.g., Chalmers 2016). As far as the extended consciousness is concerned, to use the standard extended mind framework presupposes a functionalist characterization of consciousness (see Vold 2015), classically considered questionable (see Facchin et al. 2023). An alternative to the functionalist treatment is to embrace the enactivist view of mind and consciousness, which does not come without difficulties (Jacob 2006; Di Francesco, Tomasetta 2021). Still, in this paper we would like to positively evaluate some intuitions and implications put forward in the 4E context, while avoiding potential problems related to extended mind and enactivism. To this purpose, we want to tentatively defend an alternative and more moderate idea, that of embodied consciousness (cf. Prinz 2009). Even though this expression is frequently encountered in the literature, a closer look reveals that this idea is still in need of a clear definition. The aim, then, is to fill this gap. Our attempt is driven by the following insight: if we consider the classical examples of phenomenal experiences, like feeling pain, tasting a lemon, or experiencing rage, they seem to be strongly dependent on the ownership of a body. What is now due is: (1) a clarification of the constitutive role - instead of a merely causal one - played by the body in conscious experience; (2) an appropriate distinction of the embodied consciousness thesis from the other proposals on consciousness, advanced within the 4E framework (for example, enactivist proposals); (3) an assessment of the compatibility of embodiment and functionalism, given that a potential tension between the two has been pointed out (Farkas 2019; Maiese 2019). Once the importance of the concrete embodied dimension has been acknowledged, as far as phenomenal consciousness is concerned, it could be easier to understand why it is difficult to characterize it in functional terms, abstracting from this very concrete dimension.

Session: Consciousness | Room IX
Wednesday September 18, 2024
14:25 – 14:50

I am looking for a (permanent) center of gravity
How Daniel Dennett's prophecy about AI and the self has been, at least partially, confirmed and realized

Giacomo Romano
University of Siena

In 1992 Daniel Dennett, in the essay "The Self as a Center of Narrative Gravity" suggested that an artificial self, constructed through narrative structures, would not fundamentally differ from an effective self -the self that humans perceive themselves to be. This hypothesis aligns with Dennett's broader philosophical stance, particularly his views on consciousness, and the nature of the self. In Dennett's framework, the self was not seen as a fixed, immutable entity but rather as a product of ongoing narrative construction. He argued that individuals construct their sense of self through the stories they tell about themselves, both to others and to themselves. These narratives create a coherent framework for understanding personal identity and experiences. Dennett hypothesized that an AI system, a novel-writing machine named 'Gilbert', could construct and maintain coherent narratives about its experiences, interactions, and internal states, and that it may exhibit self-like behavior or characteristics. Nowadays, there is blatant evidence that AI technologies can generate narratives or stories based on input data or predefined algorithms. Whether these AI systems can truly develop a sense of self comparable to that of humans is still a subject of debate and exploration in the fields of AI and the philosophy of mind. Indeed, while AI systems may be capable of simulating aspects of self-like behavior or engaging in narrative construction, the question of whether AI can truly possess a self in the human sense remains unresolved. This topic is an ongoing area of research and philosophical inquiry. This doubt alone is enough material to raise questions about the relationship between humans and artificial intelligence. For individuals, the comparison of AI-generated narratives with their own ones can provide insights into their experiences, behaviors, and emotions, contributing to their understanding of themselves as characters within their own life stories. However, AI enables also interactive storytelling experiences that engage users in co-creating narratives or shaping story outcomes. Interactive fiction games, chatbots, and virtual reality experiences likely allow users to interact with characters and influence narrative events, blurring the line between author and audience. Through these interactive storytelling experiences, individuals can explore different aspects of themselves, experiment with identity, and reflect on their choices and consequences within narrative frameworks. Overall, AI plays a multifaceted role in shaping the understanding of the narrative self by generating and personalizing interactive experiences with narratives. Perhaps, from a Dennettian perspective, it may be more relevantly related to how a person perceives AI-generated narrative selves as well as how they perceive their own narrative self. As AI technologies continue to evolve, they offer new opportunities to explore and reflect on how identities may emerge through storytelling. In my intervention, I intend to develop this Dennettian perspective.

Session: Consciousness | Room IX
Wednesday September 18, 2024
14:50 – 15:15

The role of touch in the sense of self formation

Iva Apostolova
Saint Paul University

I am interested in exploring here the significance of the sense of touch in relation to human/personal identity. I take the premise that the formation of human-type consciousness requires the faculty of touch. For example, in addition to the paradigmatic direct touch indispensable for intimate care, tactile perception also includes contiguous touch, projection touch, as well as the intriguing "distal touch" (Martin, 1992). One of the focal points is the exploration of peripersonal space, which is central for the formation of self. Peripersonal space is "a buffer zone between the self and the world" (see Vignemon 2021, p. 3), while not a well-defined space is, in fact, of utmost importance to the sense of self. It is within this space (a space that incorporates both spatial and temporal proximity) that we, as cognitive and social agents, determine what is safe and not safe for us to come into contact with. It is within this space that we reach and probe the other, be it an object or another subject like us. A part of the great importance of peripersonal space comes from its ties to "self-location and body ownership", without which touch would be inconceivable (see Vignemon 2021, p.9). At the same time, empirical research seems to favor a bimodal visuo-tactile neural system which allows for both multisensory integration as well as for affective responses to the environment (negative, associated with danger, and positive, associated with safety). It appears, then, that the sense of touch is at the very foundation of our (human) perceptivity. I am very sympathetic to the idea that low-level mechanisms, of tactile association, for example, feed into higher- level mechanisms such as object-recognition (see Dijkerman. and Medendorp. 2021). Multisensory integration causes the expansion or the shrinking of peripersonal boundaries. In other words, peripersonal space appears to be dynamic, as opposed to static. It develops and redefines itself, as it were, according to the multisensory integration mechanisms, starting with the tactile mechanisms (i.e., an estimation of what bodily consequences a certain action or object will have on me), which, in turn, leads to the formation and triggering of predictive mechanisms that allow me to better protect my bodily integrity and successfully orient myself in my environment. These predictive mechanisms are shaped by social cues, among other factors. Without pushing the empirical evidence too far, it seems to me that it would be fair, then, to describe peripersonal space as a constructed space. Interacting with other bodies, especially automated non-organic bodies, changes the way we feel about our own bodies. I will engage with such conceptual constructs as "network of desires" and "posthuman desires" in order to elucidate my position on human-robot touch. My tentative conclusion is that perfecting the exoskeleton and overall appearance of a bot to resemble more and more that of an organic being, especially an organic human being, will not resolve the tensions surrounding the complicated space and role of touch in the formation of the sense of (human) self.

Session: Consciousness | Room IX
Wednesday September 18, 2024
15:15 – 15:40

At first glance:
investigating how vagueness influences verbal and non-verbal shared understanding of concepts among couples

Chiara De Livio, Claudia Mazzuca, Viola Chillura,
Valerio Sperati, and Anna M. Borghi
Sapienza Univeristy of Rome; Institute of Cognitive Sciences and Technologies (CNR-ISTC)

Introduction People commonly say that lovers understand each other "at first glance". But is it true? Previous research suggests that abstract concepts evoke more social engagement and an increased need for collaboration to grasp their meaning, compared to concrete ones (Borghi, 2022). Here we propose a further factor impacting conceptual representation of abstract concepts, i.e., vagueness (the degree concepts have a precise and determinate meaning). This study explores how romantic couples and randomly paired strangers collaborate on defining concrete and abstract concepts, these last characterized by different degrees of vagueness. Specifically, the study aims to investigate how concepts' vagueness influences cognitive processes involved in a collaborative task, i.e., the creation of an educational post on a concept—addressing whether more abstract and vague concepts lead to more negotiation. To this aim, we will examine the role of each partner in creating the post and the frequency of eye contact in conveying understanding and fostering mutual understanding among participants.

Method In this ongoing research, we examine romantic couples against couples of strangers in a concept-definition task. Couples are instructed to craft a brief, educational message targeted at students who are deciding on their university path. We present couples with concepts differing in their degree of abstractness and vagueness—but comparable for familiarity—and instruct them to agree on a definition for each concept and write a post. We equip each couple with glasses measuring eye-contact occurrences between partners. After writing the post, participants must identify and attribute the keywords used in their posts. At the end, participants separately rate their and their partner's levels of expertise in the topic, engagement in the interaction, overall pleasantness of the task, and report how much each written post differs from the initial idea and how much they negotiated.

Expected Results We expect that highly vague abstract concepts will require a higher degree of negotiation compared to both less vague abstract concepts and concrete concepts, as their meaning is more undetermined. This will lead to more mutual glances and will influence the contribution needed from each participant to the interaction. We expect participants will report a greater difference between the initial post idea and the final post output with highly vague concepts compared to concrete concepts, with less vague abstract concepts lying in the middle. In addition, we anticipate semantic alignment gradually increasing across participants from highly vague abstract concepts (lowest) to less vague abstract concepts (medium), up to concrete concepts (highest). Finally, due to their established common ground, romantic couples might achieve a higher degree of agreement with less need for negotiation, difficulty, and effort compared to the control group. Moreover, we expect them to find the interaction more enjoyable and value their partner's contributions more highly.

Session: Interaction I | Room II
Wednesday September 18, 2024
14:00 – 14:25

Not only nationality: a multicultural perspective in human-robot interaction

Cecilia Roselli, Leonardo Lapomarda and Edoardo Datteri
Università di Milano-Bicocca

Over two decades of research in the Human-Robot Interaction (HRI) domain demonstrated the role that culture plays in modulating expectations towards and responses to social robots and in shaping the smoothness and effectiveness of the interactions between humans and robots. Although HRI researchers rightly identified culture as a key factor influencing the perception of robots, it is somewhat surprising that the concept of "culture" has frequently overlapped with the concept of "nationality", as if the link to a specific geographic area would be sufficiently informative when exploring how people perceive robots and understand their behavior (Lim, Rooksby, and Cross, 2020). In our view, reducing culture to a unique interpretation, i.e. culture as a nationality-based set of norms and behaviors, is restrictive. For example, it might not capture the complexity and nuances of culture, which comprises also subcultures and cultural phenomena (e.g., biculturalism) that might significantly differ from national cultures and thus be excluded by society- as in the case of immigrants and refugees. In the context of HRI, it might hinder the development of culturally competent and culturally informed robots, with the risk of designing robots equipped only with a set of encoded and fixed nationality-based rules. This, in turn, might negatively affect people's perception of robots, and thus the type and quality of interactions that people would have with them. This said, in the present work we suggest expanding the boundaries of "culture" as merely equivalent to the national identity. Specifically, we promote the adoption of a "multicultural" perspective, starting from the dynamic and constructivist approach first proposed by Hong and colleagues (2000). This broader concept of culture has the advantage of emphasizing how pieces of cultural knowledge (ideas, values, behaviors) vary not only between but also within individuals. The assumption would be that individuals, across cultures, all possess the most important implicit theories about the social world, i.e., what creates *culture*. The difference among people- and thus, among cultures- relies on the conditions under which specific pieces of cultural knowledge become operative, relative to others, in guiding the construction of meanings and behaviors (e.g., accessibility). Notably, the adoption of a multicultural perspective might be beneficial also for HRI, since people's perception and understanding of robots might be influenced by factors aside from culture equating nationality. Specifically, it might help shed light on how culture can shape people's understanding of robots, both in terms of judgments of human likeness and attribution of mental states. Last, but not least, also the design and implementation of robots would not exclusively rely on explicit national knowledge they might be equipped with, thus promoting more dynamic and effective interactions between robots and humans.

Session: Interaction I | Room II
Wednesday September 18, 2024
14:25 – 14:50

Hybrid embodied agency in human-AI interactions

Anna Ciaunica and Shaun Gallagher
University of Lisbon; University of Memphis

For most of us, most of the time, our experiences seem to be tacitly accompanied by a sense of self – a sense of being an embodied agent within a world, among but distinct from others (Gallagher 2000). Everyday experience also seems to involve experiences of agency; namely, the feeling that I am in control of my own bodily actions, that I can leverage them to access and change the external world' (Gallagher 2000; Haggard 2017). It is now well established that humans attribute human-like states to artificial others. However, the effect of interacting with artificial minds and bodies on the human sense of agency is less understood. In this talk we will present theoretical and empirical work looking at embodied joint agency in human/ human versus human/ robotic and virtual agents. Specifically, we will outline the key role of the human embodiment and sense of self in establishing joint agency with artificial others. We will argue that in order to establish a "sense of we" with both humans and artificial agents, humans need to feel connected to their own bodies first. We introduce the notion of 'hybrid agency' to describe these new, technologically mediated ways to embody and control in tandem human and artificial minds and bodies in real and virtual environments. Do we develop a sense of hybrid "we" in interacting with these artificial agents? If yes, in which sense this sense of "we" is different from the sense of togetherness that we have when we interact with biological systems? We will discuss key implications of these questions on recent efforts to design autonomous and interactive artificial others.

Session: Interaction I | Room II
Wednesday September 18, 2024
14:50 – 15:15

Enhancing trust in human-robot interaction:
an integrated approach to knowledge representation in AI

Luca Biccheri and Roberta Ferrario
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

Recent frameworks developed for collaborative intelligent systems, such as home assistants and service robots, increasingly emphasize flexibility in plan execution. This includes adapting behaviors that can be refined, learning, and stimulating positive emotions, as well as accomplishing tasks in open environments. At the same time, it is essential to ensure that such technologies are as trustworthy as possible, especially when human-robot interaction (HRI) is directly at stake. As is well known, the concept of trust is not used consistently across different disciplines, leading to a problem of intertheoretical coherence and thus hindering the empirical evaluation of trust itself. To face this problem, instead of working with a strict definition of trust, we provide an operational notion useful to study trust-building scenarios in HRI. We propose to directly ground the concept of trust on certain agents' affordances that we assume to be perceived by other distinct agents once some kinds of social interactions take place. For instance, imagine someone assisting an elderly person with carrying groceries. If they proactively approach and extend their arms to lift the grocery bag, showing appropriately timed, smooth, and restrained movements, they demonstrate 'kindness' and arguably induce trust. In this respect, 'kindness' can be said to be a 'social' affordance to the extent that it represents a possibility for interaction shaped by more or less explicit conventions. In this sense, our hypothesis is that trust requires shared knowledge about how, when, and if to act, depending on certain social contexts. The attitude of trust could then turn out to be quite sensitive to responding to context specific action patterns. Ideally, these patterns should be injected into service robots to enable them to learn human non-verbal behaviours, so that they can adjust their actions accordingly. We believe this is pivotal to establishing a successful trust setting in HRI. To see why, consider that, while collaborating, agents should not only perform the actions concerned, but they should perform them in a certain way. In the previous example, if the agent's movements are abrupt, and lack smoothness or restraint, the interaction may convey distrust despite the successful completion of the task. This is because, one may advocate, actions are not perceived solely through their completion, but through the social meaning conveyed by the manner in which they are performed, much like how the tone of voice affects the interpretation of words; appealing to proxemics, the way actions are executed carries significant social affordances that influences trust perception. However, social affordances are highly context-specific, e.g. the affordance of engagement conveyed by prolonged eye contact may foster trust in some contexts, while in others, it can be perceived as distrustful. To address this limitation, we could start by looking for elementary spatial-temporal information cues about trust (e.g. proximity, synchronicity). To map the correlation between action patterns and social affordances, we will employ unsupervised graph embedding techniques for classification tasks to be experimented on the AIR-Act2Act dataset, which is about human-human interactions and specifically designed to teach non-verbal behaviours to robots.

Session: Interaction I | Room II
Wednesday September 18, 2024
15:15 – 15:40

Structuring human-AI collaboration:
An enactive framework for modelling heterogeneous cognitive systems

Julian Zubek, Łukasz Jonak and Joanna Rączaszek-Leonardi
University of Warsaw

In recent years, growing concerns surround the relationship between human and artificial intelligence (AI). These concerns focus on how future human–AI collaboration will function and whether AI agents will replace humans in certain jobs. We believe these debates overly emphasize the notion of compatibility between internal cognitive architectures and conceptual representations of different agents. Starting from the premise that AI agents possess some form of "intelligence" comparable to human intelligence tempts us to ascribe mental states (e.g., belief, desire, intention) to AI agents without clearly defining what it means. Investigations focusing solely on AI capabilities overlook the fact that this is fundamentally a human–machine interaction problem, and the interaction part truly matters. Humans have interacted with various tools and cognitive artifacts (such as maps, compasses, and abacuses) for thousands of years. If we accept the notion of extended cognition, we can conclude that modern humans already function as hybrid cognitive systems since they routinely extend their mental facilities with technology. Additionally, humans have domesticated animals, forming successful interspecies systems capable of coordinated task performance (think shepherd-dog collaboration). We argue that these two cases – human interactions with tools and domestic animals – provide valuable metaphors and intuitions for describing human–AI collaboration. Rather than discussing AI agents in terms of intelligence or agency, we can focus on their relative operational autonomy, which distinguishes them from cognitive artifacts of previous generations. To describe this more formally, we introduce a conceptual framework inspired by pragmatic and enactive perspectives, focusing on how functional actions are coordinated within a complex system consisting of heterogeneous agents. We distinguish between agents' internal activity within the system and their external activity in the environment. Both types of activity are described in terms of an interplay between an agent's internal degrees of freedom and external constraints. This interplay defines the agent's relative autonomy in different domains. By looking at the overlap of internal and external constraint sets of different agents, we can characterize different possibilities for how hybrid systems can be constructed, and thus the potential roles of AI in collaboration with humans. If an AI agent's internal constraints overlap with a human's internal constraints, the user may have more control over the AI agent's operation. Otherwise, the agent will remain a non-transparent black box. If external constraints of both types of agents overlap, they are competitive in their actions. If external constraints are disjoint, the agents operate in a complementary fashion. We discuss how, in specific cases, these rudimentary distinctions can be operationalized through quantitative measures. The introduced vocabulary may help in evaluating the consequences of new AI systems both at the individual and the interactive, social level. We hope this discussion will contribute to the design of AI agents that respect the autonomy of their users and are aligned with societal values.

SYMPOSIUM: Sustainable behavioral change for climate crisis | Room V
Organizer: Giulia Andrighetto

Wednesday September 18, 2024

14:00 – 14:25

Growing polarization around climate change on social media

Andrea Baronchelli
City University of London

As we face the escalating risks of climate change, understanding and promoting collaborative efforts becomes crucial. This talk explores the intersection of climate change and political polarization by analyzing Twitter discussions around the United Nations Conference of the Parties on Climate Change (COP) from 2014 to 2021. Our analysis first reveals a significant increase in ideological polarization during COP26, following periods of lower polarization between COP20 and COP25. This surge is primarily driven by a fourfold increase in rightwing activity relative to pro-climate groups since COP21. Additionally, we identify a broad spectrum of 'climate contrarian' views during COP26, highlighting political hypocrisy as a topic with cross-ideological appeal. These perspectives and accusations of hypocrisy have become central themes in the Twitter climate discussion since 2019. Given the dependence of future climate action on negotiations at the international level, our findings emphasize the importance of monitoring how social norms and polarization impact public climate discourse. This understanding is essential for fostering effective cooperation in collective risk situations.

14:25 – 14:50

Social tipping intervention to promote the adoption of reusable food packaging solutions

Gian Luca Pasin
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

Packaging waste accounts for 36% of solid waste in EU towns, with plastics being the most widely used material in European food retail, covering 37% of food sold. EUR 75-112 billions of plastic packaging material is lost from the economy each year. An immediate reduction of plastic production - through the expansion of reusable food packaging solutions - is an attractive solution from environmental, economic, and social perspectives. Yet social innovation requires that a substantial section of the population is ready to change and adopt the new behavior. In this work we examine the conditions under which “social tipping” interventions may promote the adoption of reusable food packaging solutions. We will conduct a pre-registered within subjects longitudinal survey experiment with 3000 participants from France, where subjects are asked their intention to buy products either in single use or reusable food packaging. Our intervention involves an appeal promoting sustainable consumption with regular feedback about the current prevalence of sustainable consumption.

14:50 – 15:15

From business to society:
A new framework for climate services

Marcello Petitta
University of Rome Tor Vergata

Climate services are experiencing significant changes with the establishment of a new framework called Societal Climate Services (SCS). This new framework is emerging amid a growing trend in which large companies are increasingly moving away from relying on external consultancies for climate services. Instead, they are establishing in-house departments for climate-related decision-making and strategy formulation. This shift represents a move towards greater ownership and contrasts with the traditional reliance on external expertise that has characterized recent research projects. In the past, climate services were developed using the co-creation, co-design and co-development (co-co) approach. This approach relied heavily on collaboration between external experts, academics and researchers. However, the increasing in-house expertise represents a notable shift towards more independent and customized climate solutions. There is a growing need to shift current climate services from a purely business-oriented approach to a framework that places society at the center. Societal Climate Services (SCS) aim to broaden the focus beyond the needs of business to broader societal concerns, especially in vulnerable and underrepresented communities, particularly in developing countries. SCS seeks to democratize climate knowledge and make it accessible and useful not only to businesses, but to society as a whole. This people-centered model emphasizes community participation and the integration of local knowledge in the design of climate solutions. This

summary outlines the principles of Societal Climate Services and emphasizes the importance of cross-sector collaboration for a comprehensive and integrated approach. It emphasizes the role of SCS in promoting sustainable and resilient communities through long-term planning and investment in sustainable practices. It also emphasizes the need for equity and justice in the provision of climate services to ensure that solutions do not exacerbate existing inequalities, but instead help to reduce them. The increasing momentum of companies building in-house capabilities for climate services, together with the emergence of SCS, represents a significant development in this area. This combination promises a more integrated and effective framework for addressing the challenges of climate change that aligns both business interests and societal needs in the pursuit of global climate resilience.

15:15 – 15:40

Widening the scope:
The direct and spillover effects of nudging water efficiency
in the presence of other behavioral interventions

Jacopo Bonan
University of Brescia

Policymakers and firms use behavioral interventions to promote sustainable development in various domains. Correctly evaluating the impacts of a nudge on behavior and satisfaction requires looking beyond the targeted domain and assessing its interactions with similar interventions. Existing evidence on these aspects is limited, leading to potential misestimation of the cost-effectiveness of this type of intervention and poor guidance on how to design them best. Through a large-scale randomized controlled trial implemented with a multi-resource utility company, we test the impact of a social information campaign to nudge water conservation over two years. We find that the water nudge significantly decreases water and electricity usage but not gas. The effect is driven by customers who do not receive nudges targeting the other resources. Customers receiving the water report are also significantly less likely to deactivate their gas and electricity contracts, regardless of whether they receive other reports. Our results suggest that multiple nudges strain users' limited attention and ability to enact conservation efforts. Users' constraints in attending to multiple stimuli pose important challenges for designing policy interventions to foster sustainable practices.

Session: Interaction II | Room VIII
Wednesday September 18, 2024
16:05 – 16:30

Joint guidance: a capacity to jointly guide

Marco Mattei
Università di Milano

Sometimes, we act in concert with others, as when we go for a walk together, or when two mathematicians try to prove a difficult theorem with each other. An interesting question is what distinguishes the actions of individuals that together constitute some joint activity from those that amount to a mere aggregation of individual behaviours. It is common for philosophers to appeal to collective intentionality to explain such instances of shared agency. This framework generalizes the approach traditionally used to explain individual action: a behaviour is an action just in case it causally follows from the relevant intention. Contemporary philosophers of action, as well as cognitive psychologists, however, have criticised this way of explaining individual actions, because it does not say anything about the underlying neural and cognitive processes that make joint action possible. In individual action, nowadays, theorists favour an approach that puts “control” or “guidance” as the discerning factor: a behaviour is an action just in case the agent controls it, or just in case it is guided by the agent, without any need for intentions. In this paper, I apply this guidance framework to group action. Consequently, a behaviour that spans multiple individuals is a case of shared agency just in case it is jointly guided by the group, or, equally, the group controls it. In developing this view, I first show what this “capacity to jointly guide” amounts to and how it relates to individual guidance. I argue that an approach that favours joint guidance over collective intentions eschews a lot of metaphysical problems about collective mentality and group subjects, and it is thus more explanatorily fruitful; furthermore, by explaining joint guidance in terms of co-representation and joint commitments, it structures future scientific research into the matter.

Session: Interaction II | Room VIII
Wednesday September 18, 2024
16:30 – 16:55

(Dis)embodied joint agency in human-VR agents Interactions

Altea Vanni, Sophia Bertoni, Shihan Liu, Jiaqi Yin, Sylvia Pan and Anna Ciaunica
University of Lisbon & Goldsmith; University of London

Introduction

The sense of agency -feeling of being in control of one's bodily actions- is a fundamental aspect of the human mind. Previous work showed that the Joint sense of agency (JSoA) -sense of control experienced when acting with others- depends on the type of agent we interact with. However, the effect on the body of interacting with human or artificial bodies remains an open question. This study investigates the effect of Depersonalization (DP) -a condition in which people feel detached from their self and body- on embodied joint agency in human/human versus human/artificial dyads.

Methods

Using the Joint Simon Task combined with the Intentional Binding task, we aim to investigate the effect of DP on Human/Human versus Human/VR sense of joint agency. Participants, categorized by High and Low Levels of DP, will embody either a Human avatar or a Social Humanoid Robot 'Pepper' avatar in virtual reality (VR). They will then perform both tasks with either a Human avatar or a 'Pepper' avatar.

Hypotheses

We hypothesize that High DP participants will show a higher Joint Simon Effect when embodying a robotic avatar and performing the joint task with the robotic avatar co-agent, compared to a human avatar co-agent. Conversely, Low DP participants are expected to show higher Joint Simon Effect when embodying a human avatar during the joint task with human co-agent, compared to the robotic co-agent. This is based on the idea that people with High levels of DP that can perceive themselves as 'machines' or 'automata' may develop a higher sense of joint agency when interacting with another robotic body versus a human body.

Discussion

We investigate for the first time the effect of embodiment on the sense of Joint Agency in human versus robotic avatar in VR. A better understanding of how feelings of being (dis)connected from one's body impacts the way people feel (dis)connected from human and artificial others may help better design human/artificial agents interactions.

Session: Interaction II | Room VIII
Wednesday September 18, 2024
16:55 – 17:20

Mapping the psychophysiology of commitment

Angelica Kaufmann, John Michael, Luke McEllin,
Corrado Sinigaglia, Stephen Butterfill,
Guido Barchiesi and Martina Fanghella
University of Milan; Central European University; University of Warwick

Human joint actions, ranging from mundane tasks to complex societal challenges, rely on a sense of commitment to persist despite fluctuating individual interests. Previous research highlights this commitment's dependence on cues signalling others' expectations and reliance (Sebanz et al., 2006; Tomasello, 2009; Melis & Semmann, 2010; Michael, 2022; Michael, Sebanz, & Knoblich, 2016a; Dana, 2006; Heintz et al., 2015; Sugden, 2000; MacCormick & Raz, 1972; Scanlon, 1998). However, the psychophysiological underpinnings of this commitment are less well understood. This study aims to bridge the gap in understanding the psychophysiological processes underpinning the sense of commitment in joint actions, exploring how perceptual cues of a partner's expectations affect psychophysiological activity and commitment. Specifically, our study probes the effects of the sense of commitment upon motivation to persist in joint action. As a starting point, we draw upon a distinction between two forms which this may take (Michael, 2022). The first form may be dubbed "gritted teeth commitment". This is the form of commitment you experience when you find yourself bored or distracted, or otherwise tempted to abandon a goal, but nevertheless force yourself to persevere, and to resist temptations and distractions. We hypothesize that this involves the deployment of executive control mechanisms (e.g. inhibitory control and supervisory attentional control) to maintain task focus and to avoid temptations and distractions. The second form may be dubbed "engaged commitment". This is the form of commitment you experience when you are so immersed in pursuing a goal that you do not notice temptations or distractions in the first place, and therefore do not need to force yourself to ignore or resist them. We hypothesise that this boosts the relative salience and attractiveness of task-relevant information, making task-irrelevant stimuli in the environment and task-irrelevant thoughts less tempting or distracting than they otherwise would be. The experimental design integrates EEG with behavioural measures and questionnaires. The study examines the influence of commitment on motivation to persist in joint action, specifically looking at 'gritted teeth' and 'engaged' forms of commitment (Michael, 2022; Baddeley, 1986; Christensen, Sutton, & McIlwain, 2016). In a coordination task, the readiness potential (RP) of participants is measured in response to temptations to defect from joint actions (Schurger et al., 2021; Trevena & Miller, 2010; Schultze-Kraft et al., 2016; Panasiti et al., 2014). We measure RP employing EEG after asking participants if they want to defect or not. RP is a negative ERP component which usually precedes voluntary actions. Interestingly, it has been shown that RP arises before the conscious will to initiate an action Participants are instructed to press a button if they choose to negate the trial (defect), and otherwise to remain still and wait for the next trial. During this 2600-millisecond response phase (Haggard & Eimer, 1999), we will measure electrophysiological activity from each participant. This will enable us to ascertain to what extent a participant is preparing a button-press action (i.e. defection) but then inhibited this action (i.e. through gritted teeth). Indications from the literature review and our data suggest a complex relationship between sensory-motor signals, internal models of partners' actions, and the varying forms of commitment in joint actions (McEllin & Michael, 2022; Székely & Michael, 2018; Chennells & Michael, 2018; Bonalumi, Isella, & Michael, 2019). We predict that different types of commitment will show distinct psychophysiological profiles. 'Engaged' commitment is hypothesized to increase task salience and reduce the need for executive control, while 'gritted teeth' commitment might involve heightened executive control to maintain task focus.

Session: Explanation II | Room IX
Wednesday September 18, 2024
16:05 – 16:30

Is it a bug or is it a feature?
Decisional enhancement, autonomy, and rationality in the digital age

Camilla Colombo
RWTH Aachen University

"It's not a bug, it's a feature!" is a programmers and software developers' popular joke on malfunctions. The insight is that when a mistake is ubiquitous or "built-in", it becomes a distinctive feature of the product. In this paper, I argue that this phrase highlights some critical assumptions about human rationality in the digital landscape. Specifically, I focus on technological tools, interfaces, and devices designed to impact our decision-making with the purpose of enhancing it and making it more efficient. I point out that the use of these decisional aids and techniques underlies a narrow conceptualization of rationality, one in which many crucial features of human reasoning would require "fixing". This stance, however, has relevant theoretical and ethical implications. The conceptualization of rationality is a core issue of philosophical investigation: Rational Decision Theory (RDT), one of the most influential models of human behavior in social and life sciences, formulates a set of requirements on agents' desires and beliefs, so that compliance with said requirements defines an agent as rational. As such, RDT is inherently normative: deviations are interpreted as mistakes and thus "irrational". With the development of cognitive psychology and neuroscience, however, some (or most) of the key assumptions of RDT have been put into question showing that many deviations, usually labeled as biases, are systematic and robust. Far from being idealized rational agents, meeting strict consistency requirements, human beings are prone to all sorts of reasoning flaws. Responses to this challenge are various: while theories of bounded rationality strive to reduce agent idealization, allowing for more comprehensive models of human action, social choice theory focuses on how cognitive constraints can serve choice architecture. Within this latter approach, the growing digitization of many choice situations, and the large-scale development and availability of technological devices and (interactive) machines, resulted in a quasi-permanent exposure to "decisional aids" (Valera 2019), in a variety of settings ranging from trivial every-day tasks (road planning) to substantial and even existential decisions (healthcare choices). Similarly, algorithms, "AI"- powered devices, virtual environments, and the like are key to many de-biasing, nudging, or boosting techniques (Becker et al. 2019). The underlying assumption of this approach is that our decision-making processes and skills, given consistent deviations from the standard rationality model, are inherently limited. Aids aiming at correcting such deviations would thus make our choices more efficient, helping us reach our goals in a more "rational" way. This framework, however, leaves a substantial conceptual gap in the characterization of rationality, and opens serious ethical concerns. First, the enhancement approach adopts a substantive and normative interpretation of what counts as rational behavior and identifies deviations as mistakes. More recent theories of bounded rationality (Bradley 2017), however, argue that some of these decisional schemes are rather distinctive aspects of human behavior, capturing our intuitive understanding of what is reasonable, at least in specific decision contexts, and should be incorporated in a descriptively adequate notion of rationality. I discuss whether the enhancement framework can handle some "naturalistic" features of human reasoning. Secondly, the pervasiveness of decisional aids and enhancement tools raises questions as to whether efficiency is the only desirable target when modeling decision-making. While making instrumentally "better" choices is an obvious desideratum, more comprehensive accounts of rationality (Felsen et al. 2013) also require agents to go through the decision process, "own" it, and be able to justify its motivations and outcomes. These features are also key to most conceptualizations of decisional autonomy (Niker et al. 2021). I explore here whether decisional enhancement can be construed as supporting autonomous decision-making (Colombo and Nagel 2023).

Session: Explanation II | Room IX
Wednesday September 18, 2024
16:30 – 16:55

The superbug:
mental models and errors in computer programming

Silvia Larghi and Edoardo Datteri
Università di Milano-Bicocca

In the tradition of the so-called psychology of computer programming (see Weinberg 1971), this paper explores the relationship between people's understanding of computers and their programming errors. 'Poor' mental models of a computer system can cause programmers to make mistakes that can lead to system malfunction. In particular, Pea (1986) argued that some specific programming errors are caused by incorrect mentalisation of computers, which he called 'superbug'. He identified three forms of superbug that are often observed in novice programmers. The first is the 'parallelism bug', where the programmer incorrectly assumes that different sequentially ordered lines of code can be simultaneously active and executed in parallel by the system. The 'intentionality bug' consists in taking an intentional stance towards the computer (Dennett, 1971), and in particular in attributing the ability to go "beyond the information given" to the program itself. The 'egocentrism bug' is very similar to the intentionality bug, and consists in assuming that the computer has some kind of theory of the programmer's mind, and knows their programming goals even though they are not represented in the program. All three forms of superbug arise, according to Pea, because the programmer unwittingly and unconsciously assumes that "there is a hidden mind somewhere in the programming language that has intelligent interpretive powers". We will argue that Pea's insight can shed light on the cause of programming errors and connect the emerging literature on the attribution of mental states to artificial systems (Thellman et al., 2022) to the psychology of computer programming in a way that is relevant to contemporary research on computational thinking (Denning & Tedre, 2019) and educational robotics (Anwar, 2019). We will also try to refine Pea's thesis by developing the idea that the superbug is not the attribution of mental states and capacities to the computer per se, but rather the attribution of the wrong mental states and capacities to it. As Dennett points out, the intentional stance can be predictively useful, and in some circumstances offers significant advantages over other ways of conceptualizing the system. The key to avoiding the programming errors that Pea refers to is for the programmer to attribute to the system goals (beliefs, intentions, ...) that the system actually has; the superbug arises when the mental model of the system's mind is, in some sense to be clarified, wrong. For example, suppose the programmer is faced with a Python function that implements a bubblesort algorithm designed to operate on arrays of integers. There is a clear sense in which the programmer can bypass the computational language and attribute to the function the goal of ordering integers. This is a 'correct' goal attribution. If the programmer calls this function on an array of integers, no programming error is made. Now suppose the programmer mistakenly assumes that the program has a theory of their mind and is therefore able to 'understand', beyond the information given, what types of values the programmer wants to order. They will call the Python function on, say, an array of characters, causing an execution error. Admittedly, this decision may have nothing to do with the attribution of propositional attitudes to the system. However, following Pea's insight, it is safe to say that in novice programmers it might also be caused by the mistaken attribution of mental, interpretive capacities to the program. The point here is that this attribution is not mistaken because it attributes mental capacities to a system that does not have a mind, but because it attributes the wrong mental capacities to a system that, as Dennett and others have pointed out, can in some cases be usefully modeled as a mental agent. To sum up. Programmers deal with lines of code. In some cases, they mistakenly mentalise the machine, creating the superbug. But mentalising is not all bad: it can be useful for computer programming, provided it is good mentalising.

Session: Explanation II | Room IX
Wednesday September 18, 2024
16:55 – 17:20

Generative AI and the overextended mind:
on legal ownership of our cognitive extensions

Fabio Paglieri
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

Reliance on generative AI systems to perform on our behalf increasingly complex cognitive tasks, such as writing a text or drawing a picture, raises concerns on how this might impact our capabilities (Paglieri, 2024), resulting in loss of competences (deskilling; Pritchard, 2016) or failure to acquire them during development (cognitive diminishment; Kasneci et al., 2023; Mhlanga, 2023; Shiri, 2023). However, the extended mind hypothesis (Clark & Chalmers, 1998) seems to assuage such worries: if understood as cognitive extensions, generative AI software no longer pose a threat, since the relevant competences are not "lost", simply offloaded to an external device, which remains part of our (extended) cognitive system. Attempts at resisting this externalist view hinges either on proving that the technologies under discussion are not proper cognitive extensions, e.g. because they do not satisfy Clark and Chalmers' parity principle, or on rejecting the whole extended mind hypothesis, e.g. arguing that technological devices in general lack "the mark of the cognitive" (Adams & Aizawa, 2010). More recently, significant attention (Heersmink, 2013; Clowes, 2015; Farina & Lavazza, 2022) was given to the distinction between constitutive incorporation of a device into the cognitive system (a narrow and stronger sense of "extension"), and frequent or even continuous use of a device to perform some cognitive function, but without leading to incorporation (a broad and weaker sense of "extension"). This results in a nuanced taxonomy of cognitive artefacts, based on how they absolve cognitive functions in relation to users' capabilities, distinguishing between substitutive, complementary, and constitutive artefacts (Fasoli, 2017). These debates are theoretically important, yet they miss the (practical) elephant in the room: if mental processes are externalized to artifacts outside of our body, either in a weak or strong sense, what is to prevent the expropriation of such artifacts, and therefore of whatever cognitive processes they are expected to perform? If that happens, what are the consequences? Do we have any specific grounds to object against such expropriation, other than invoking standard property rights? Does the role these artefacts play in our mental processes give us any special right over them? Ironically, the issue of ownership came up frequently in the extended mind debate: Rowlands (2009) proposed ownership as a way of immunizing externalism against cognitive bloat (i.e., excessive proneness to accept any causally relevant external influence over cognitive processes as integral to them); yet the notion of ownership invoked by Rowlands and others (e.g., Gallagher, 2013; Smart, Andrada & Clowes, 2022) is phenomenological in nature and related to subjecthood. In contrast, the concept of ownership that ought to preoccupy us is legal. Unfortunately, the key question "who legally owns the extended mind?" has been asked only once, in an obscure paper (Dunagan, 2015), buried in a handbook on intellectual property. This contribution aims to redress such oversight, using generative AI systems as a case study to demonstrate that overextending our mind is a bad idea: not because we lose cognitive skills, but because we become increasingly dependent on private powers for their use.

Session: Human Kinds | Room II
Wednesday September 18, 2024
16:05 – 16:30

Questioning the boundaries of addiction

Davide Serpico and Francesco Guala
University of Milan

Defining addiction is a complex task involving both epistemological and ethical questions. From a physiological perspective, addiction is described as an involuntary condition depending on a substance's effects on dopaminergic and reward circuits (Leshner 1997). In contrast, from an ethical standpoint, addiction is considered a normative failure, the consequence of voluntary decisions for which an individual is held responsible. Both views have been challenged by behavioral scientists drawing a distinction between consequence-driven and elicited behaviors (Heyman 2009). Notably, this approach points at the role of the social context: if the environment can provide one's with valuable alternatives, recovery becomes possible, and this explains why therapies manipulating the relative value of drug consumption succeed when the incentives are calibrated carefully (e.g., promoting friendship, leisure activities, and food). However, the behavioral approach widens the category of addiction significantly: gambling, sex, and even compulsive shopping are consequence-driven, relatively inelastic behaviors that fit the behavioral conception, though there is no evidence that they are based on the same physiological mechanisms unlerlying drugs consumption. In this talk, we investigate epistemic and normative questions in the definition of addiction, particularly with respect to debated cases such as gambling and food consumption. Do different types of addiction belong to the same kind? Does individual physiology react similarly to the environment also in cases that are less clearly defined as addiction? What are the normative implications of seeing various addiction-like behaviors as similar or different? First, we consider case studies (e.g., gambling, drugs, and food consumption) that appear to have little commonalities at the physiological level but that, nonetheless, produce similar behavior in similar social circumstances. We shall use such cases to develop a model of addiction(s) accounting for the physiological mechanisms underlying behaviors typically associated with addiction and the interactions between such mechanisms and factors external to individuals (e.g., availability of alternatives). We thus explore whether the multi-level property cluster associated to addiction(s) is multifunctional, i.e., if the same causal structure is realized by different physiological and social mechanisms in different cases. Second, we consider normative implications. Many scientific categories are used not only to describe but also to prescribe: they carry positive or negative connotations that may influence the behavior of laypeople, scientists, and policymakers. In this sense, the use and conceptualization of certain categories inevitably involve value-laden decisions concerning both epistemic and pragmatic purposes (Mallon 2016; Haslanger 2012; Griffiths 2004; Ereshefsky 2009). We aim to suggest that similarities between well-established addictions and uncertain cases can lead to rethink the way institutions regulate people's availability to certain experiences and products.

Session: Human Kinds | Room II
Wednesday September 18, 2024
16:30 – 16:55

Clarifying the muddle
Towards a comprehensive taxonomy of cognitive biases in medicine

Cristina Amoretti & Elisabetta Lalumera
University of Genoa; University of Bologna

Common characterizations of cognitive biases define them as cognitive processes that systematically depart from the accepted standards of logic and reasoning and, as a result, affect our judgment and decision-making. Cognitive biases are frequently characterized as predictable (they may be anticipated to occur in specific situations), universal (they affect all people), tenacious (they have an impact even on those who are aware of them), and unconscious (they are cognitive processes that the subject is not aware of). In the literature, more than a hundred cognitive biases have been identified in the medical field (Saposnik et al. 2016). The availability bias, the confirmation bias, the representativeness bias, anchoring, base rate neglect, the conjunction fallacy, or the expectation bias are just a few examples. The default position in the medical literature is that cognitive biases are epistemically bad, as they lead to misdiagnosis and corrupt research, and therefore should be eliminated from both medical research and practice. Some recent works, however, highlight the possible epistemic advantages of cognitive biases. For instance, the use of prototypical reasoning in cognitive tasks such as categorization is related to constraints on any finite agent having limited access to knowledge relevant to a given task. In most cases, cognitive processes based on prototypical reasoning are fast, automatic, and cognitively undemanding. Thus, the representativeness bias, that is the tendency of associating prototypical information with diseases, proves epistemically useful in cases where prompt diagnosis is required despite limited access to knowledge, such as in emergency situations (Amoretti, Frixione, Lieto 2017). Still, when discussing whether cognitive biases have a negative or positive epistemic role in medicine, they are typically treated as an indistinct muddle, without seriously asking whether there can be a basic epistemic distinction between different kinds of cognitive biases. Challenging the prevailing idea that all cognitive biases must be regarded in the same way in all medical contexts, we therefore advance a preliminary taxonomy of cognitive biases with regard to the medical context. We identify three broad categories. First, some cognitive biases can be regarded as suboptimal strategies (in the sense that they are different from the best strategies that are established by the rules of instrumental rationality), but still lead to the correct outcome, at least in certain contexts. Second, other cognitive biases systematically lead to the wrong outcome (again, from the point of view of instrumental rationality), which may nevertheless turn out to be the most useful one from a pragmatic point of view. Finally, come cognitive biases systematically lead to the wrong outcome, which may not, moreover, turn out to be the most useful one, no matter what the context is. Some examples will be provided, with regard to the medical field. The discussion then extends to consider whether the proposed taxonomy can be applied to algorithmic biases in medical AI.

Session: Human Kinds | Room II
Wednesday September 18, 2024
16:55 – 17:20

AI in forensic evaluations: just smoke and mirrors or an incoming revolution?

Camilla Frangi, Alexa Schincario and Cristina Scarpazza
University of Padoa

The insanity defense is a cornerstone of most juridical systems, in that it is believed that a person incapable of understanding or inhibiting an action against the law should not be punished for it. Evaluating insanity, however, is extremely challenging, as forensic psychopathology is affected by different limitations which make it difficult to retrospectively formulate a scientific opinion on a person's state of mind. Some of these limitations concern the nature of psychopathologies themselves, such as the absence of biomarkers of disease and the consequent application of purely clinical criteria, which often remain open to interpretation. Moreover, forensic evaluations may be affected by cognitive biases in the experts' reasoning and may be hindered by the attempts of the subject to simulate or dissimulate their symptoms (malingering). All these limitations contribute to decrease the inter-rater reliability of both the psychopathological diagnosis and the conclusion about insanity. In this context, artificial intelligence may be applied to support the expert in their evaluation in different ways, thus reducing the impact of the human factor (e.g., cognitive biases) on forensic decision-making. In this presentation, we discuss the possible applications of AI to the field of forensic evaluations conducted according to a multidisciplinary neuroscientific approach and the efforts that have already been made to develop models applicable to this field. For example, specific AI tools may conduct a more thorough search for scientific papers relevant to the case at hand, or may help to integrate information obtained from different sources coherently. Although the available models are now few, the possibilities for development are potentially endless and might improve the accuracy and reliability of insanity evaluations manifold. We also point out the challenges that AI applications in the forensic field still need to face before being effectively implemented. The first of these challenges concerns the training of algorithms, which needs to be supervised to comply with the need of understanding how the algorithm classifies cases. Using supervised learning raises nonetheless additional questions, linked once again to the low inter-rater reliability of insanity evaluations. The main issue lies in deciding which data feed to the algorithms for training, as an algorithm can be as reliable as the data it is trained on. In the case of forensic evaluations, there is no agreement on which opinion (the judge's, the expert witness for the judge's, the defense consultant's, …) should be considered the ground truth. Moreover, for juridical systems which recognize the existence of partial insanity, boundaries between sanity, partial insanity and total insanity should be established a priori, which might be problematic in that partial insanity is not a scientific concept. Finally, ethical implications need to be taken into account. To better describe the potential AI applications and their shortcomings at this time, we discuss a real case. We conclude that before AI can be reliably applied in criminal trials these challenges need to be addressed and solved.

Session: Concepts & Emotions | Room V
Wednesday September 18, 2024
16:05 – 16:30

Investigating the influence of interoceptive accuracy on the classification of abstract and concrete concepts during pregnancy

Salvatore Diana, Anna Borghi and Laura Barca
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

For a long time, the focus of how sensory experiences give meaning to our ideas (grounding concepts) has been on exteroception – the five senses of sight, touch, taste, smell, and hearing. However, recent research suggests that interoception – the sensing of the physiological conditions of the body - might play a significant role in conceptual knowledge, especially when it comes to understanding emotional and abstract concepts. To test for the covert role that interoceptive processes might play in conceptual representations, we developed an interoceptive-exteroceptive categorization task in which participants, presented with different kinds of abstract and concrete concepts, were asked to indicate - by moving the computer mouse - if they perceived them by inner bodily sensations (i.e., interoception) or by the five perceptual senses (i.e., exteroception). To account for individual differences in attending to bodily signals, we: i) measured participants' cardiac interoceptive accuracy (Heartbeat Counting Task); ii) tested a group of pregnant women, for whom we hypothesize a heightened salience of interoceptive information and faster and more accurate categorization of abstract concepts due to the significant physiological changes they experience. A group of 40 women (37 controls and 3 pregnant women) participated in the study. Abstract and concrete concepts varied for interoceptive grounding (emotional, philosophical, natural, artefact). Overall, the results of the interoceptive-exteroceptive categorization task highlight the malleability of the boundaries between different types of concepts and the multiplicity of dimensions involved in their conceptual knowledge. The reduced score in cardiac interoceptive accuracy suggests a reduced ability to detect bodily signals in the study group. As for conceptual categorization, in the control group, concrete-artefact concepts (car, scissors, airplane etc..) were categorized more quickly compared to other concepts - suggesting that they clearly convey exteroceptive features. Differently, concrete-natural concepts (cave, swamp, ocean etc…) elicited slow responses, a high number of (interoceptive) misclassifications, and movement trajectories attracted by the competing (interoceptive) response option, suggesting that they were perceived as conveying not only exteroceptive but also interoceptive features. Cardiac interceptive accuracy negatively correlated with the misclassification of concrete concepts, with a reduction in the number of misclassifications of both artefact and natural concepts as interoceptive accuracy increase. As for the pregnant group, poor cardioception does not support our hypothesis of a greater ability to detect bodily signals (to be confirmed with a larger sample size and other measures of interoceptive accuracy). Additionally, no significant differences emerged between pregnant participants and controls in the categorization of abstract and concrete concepts (on both temporal and kinematic measures of the response). However, as response trajectories are concerned, it is possible to observe a heightened "interoceptive attraction" for concrete-naturals' trajectories within the pregnant group, suggesting a potentially greater relevance of the interoceptive dimension in this population (but the reduced size of the study population prevent from drawing any firm conclusions). The findings of the study are discussed within the debate on embodied theories and the representation of abstract concepts.

Session: Concepts & Emotions | Room V
Wednesday September 18, 2024
16:30 – 16:55

DiffuseFace:
A database of AI-generated face portraits of non-existing people to enrich diversity in face research

Alessia Firmani and Luca Cecchetti
IMT School for Advanced Studies Lucca

We present DiffuseFace, a database of 1080 AI-generated face portraits of non-existing people of various ages, and nationalities, displaying different facial expressions. Our approach addresses the limitations of traditional face databases, such as costs and sharing limitations, and enhances diversity in the study of face perception, social cognition, and emotion. Facial expressions and the human face play a significant role in determining how people interact (Jaeger et al., 2019; Gunaydin et al., 2017). Indeed, while facial cues may not always be entirely accurate, they can still convey significant information about a person (e.g., personality, emotions). The importance of understanding how people perceive and react to the human face has sparked a surge of studies in psychological science addressing questions of perceptual, cognitive, and affective nature. This motivated the development of face databases, such as CEED or FACES (Benda & Scherf, 2020; Ebner et al., 2010). However, building a traditional face database by photographing real people is expensive and time-consuming. Additionally, privacy concerns arise, particularly when researchers aim to share their stimuli. These challenges have resulted in relatively small (e.g. narrow age range, few ethnicities, limited set of expressions) face databases available to the scientific community. This lack of diversity in databases impinges on the generalizability of findings in psychological science (see own-age bias - Perfect & Moon, 2005; for an in-depth critique see Barrett et al., 2019). The past few years have witnessed the emergence of Generative AI (g-AI), a revolutionizing technology capable of creating media content in response to prompts. Psychological research is starting to recognize the potential of g-AI as an untapped opportunity to advance research methods (Demszky et al., 2023; see also Ke et al., 2023 review). For instance, studies now show that large language models can effectively rate the emotional valence of texts, achieving results comparable to human scorers (Rathje et al., 2024), and that these tools can significantly reduce experiment costs (Hutson, 2023). In line with this, we hypothesize that text-to-image g-AI can be demonstrated as an effective tool to create large, diverse databases of face stimuli at lower costs and with fewer constraints in terms of data sharing (but see Wang et al., 2023 for privacy issues in g-AI).

Here, we introduce DiffuseFace, a database of facial portraits featuring non-existent people generated using a pre-trained, open-source latent diffusion model (Stable Diffusion; Rombach et al., 2022). DiffuseFace addresses key limitations of traditional databases by offering a diverse and rich set of stimuli. It comprises 1080 AI-generated headshots featuring young, middle-aged, and older women and men from 12 nationalities (e.g., Brazil, Nigeria, Finland, Vietnam), displaying 15 distinct emotional expressions (e.g., amusement, shame, contempt, fear). Building on prior research (Holland et al., 2019; Oosterhof & Todorov, 2008), we will collect online ratings of emotions and personality traits from a representative sample of Italian participants to test whether AI-generated stimuli possess characteristics comparable to those found in traditional face databases. Our work suggests the successful application of g-AI for generating realistic human facial portraits. By enabling the creation of diverse, high-quality stimuli at lower costs and with less effort, g-AI can significantly enhance psychological and neuroscientific research. Furthermore, the ease of sharing AI-generated stimuli among researchers has the potential to improve the robustness and generalizability of face research.

Session: Concepts & Emotions | Room V
Wednesday September 18, 2024
16:55 – 17:20

Increasing emotional distancing with prism glasses:
Dissociated gender and adaptation direction effects on alexithymia in healthy individuals?

Laura Culicetto, Selene Schintu, Chiara Lucifora, Massimo Mucciardi,
Alessandra Falzone, and Carmelo Mario Vicario
University of Messina; University of Trento; University of Bologna

Emotional processing is closely linked with spatial attention, which tends to prioritize emotional stimuli over neutral ones. The brain network responsible for directing spatial attention towards different sectors of the space also play a role in processing emotional stimuli. Recent evidence has identified a connection between the rightward shift in spatial attention, assessed through the line bisection task, and the challenges in comprehending one's own and others' emotional states—referred to as alexithymia. Based on this evidence, this study hypothesized that alexithymia, might be affected through prismatic adaptation (PA), a standard protocol to modulate visuospatial attention. A sample of 103 participants completed alexithymia questionnaires, Toronto Alexithymia Scale (TAS-20) and Perth Alexithymia Questionnaire (PAQ), in a counterbalanced order before and after a prismatic adaptation session (leftward, rightward, or neutral deviating prisms). Results showed that leftward PA significantly increased alexithymia scores in healthy individuals, with a selective effect in women compared to men. Our preliminary results suggest that the attentional shifts induced by leftward PA not only affect spatial tasks, but also emotional processing, particularly in how individuals perceive and interpret emotional proximity and distance. Consequently, alexithymia may be metaphorically likened to an impaired perception of emotional closeness and remoteness.

Session: Cooperation | Room VIII
Thursday September 19, 2024
10:00 – 10:25

Language-based game theory in the age of artificial intelligence

Veronica Pizziol, Valerio Capraro,
Roberto Di Paolo and Matja Perc
Università di Milano-Bicocca; University of Bologna

Understanding human behaviour in decision problems and strategic interactions has wide-ranging applications in economics, psychology, and artificial intelligence. Game theory offers a robust foundation for this understanding, based on the idea that individuals aim to maximize a utility function. However, the exact factors influencing strategy choices remain elusive. While traditional models try to explain human behaviour as a function of the outcomes of available actions, recent experimental research reveals that linguistic content significantly impacts decision-making, thus prompting a paradigm shift from outcome-based to language-based utility functions. This shift is more urgent than ever, given the advancement of generative AI, which has the potential to support humans in making critical decisions through language-based interactions. We propose sentiment analysis as a fundamental tool for this shift and take an initial step by analyzing 61 experimental instructions from the dictator game, an economic game capturing the balance between self-interest and the interest of others, which is at the core of many social interactions. Our meta-analysis shows that sentiment analysis can explain human behaviour beyond economic outcomes. We discuss future research directions. We hope this work sets the stage for a novel game theoretical approach that emphasizes the importance of language in human decisions.

Session: Cooperation | Room VIII
Thursday September 19, 2024
10:25 – 10:50

The effect of heterogeneous distributions of social norms on the spread of infectious diseases

Daniele Vilone, Eva Vriens and Giulia Andrighetto
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

The emergence due to the outbreak of the CoVid-19 disease, caused by the SARS-CoV-2 virus, suddenly erupted at the beginning of 2020 in China and has soon spread worldwide. This has caused an outstanding increase on research about the virus itself and, more in general, epidemics in many scientific fields. In this work we will focus on the dynamics of the epidemic spreading and how it can be affected by the dynamics of Social Norms. First of all, it is reasonable to expect that the level of compliance of Social Norms concerning health and hygiene by the members of a population may have a huge influence on the infection rate, and consequently on the dynamics and outcome of the pandemics. The level of compliance of Social Norms depends naturally on the behaviour of each individual, so that it can be represented by a distribution among the population. Up to now, studies about the influence of Social Norms complying on the spreading of the CoVid-19 have focused simply on the average value of Norms complying, finding a predictable result (the higher is the average level of complying, the better is the response of the population to the pandemics). On the other hand, many countries show similar average with different dynamics of the pandemics: it is then necessary to consider the higher moments of the distribution. In particular, in this work we focus on the standard deviation: fixed the main value, which distribution allows to addressing better the pandemics, a more heterogeneous one (i.e., with higher standard deviation), or a different one with smaller fluctuations with respect to the average? Here we present simulations based on a compartmental model of the pandemics which indicate that heterogeneous distributions allow a more efficient response to the spreading of the virus: this happens because, having fixed the average, a higher standard deviation implies more agents with higher level of norm compliance, who act as a more efficient barrier against the spreading of the pathogen. At the moment there are not available data yet to test rigorously this hypothesis, but similar studies about tax evasion show this same result. Therefore, we also present a theoretical study about time heterogeneity, that is, when norm compliance (therefore, the infection rates) changes in time. Finally, we propose suitable new studies and data collections to verify the robustness of our hypothesis.

Session: Cooperation | Room VIII
Thursday September 19, 2024
10:50 – 11:15

Tiny dictators:
Understanding altruism in young children

Marco Marini, Sebastiano Munini, Michela Carlino, and Fabio Paglieri
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

Previous studies suggest that young children demonstrate highly prosocial behavior, though it remains unclear whether this prosociality arises from intuitive or reflective decision-making processes. Additionally, it is well recognized that from early in ontogeny, a sense of ownership and values of equity influence altruistic sharing behaviors. This study aims to investigate these dynamics in preschool children using an adapted version of the dictator game paradigm. The study enrolled 124 children from six kindergartens, aged 4-5 years. Children participated in two experimental sessions approximately two weeks apart. The study employed a revised dictator game with three within-subjects conditions: Control (CC), Give, and Take. In the CC condition, children were instructed to distribute ten stickers between themselves and an absent peer. In the Give condition, children received ten stickers and had the option to donate any number to the peer. In the Take condition, all stickers were initially given to the peer, and children could choose to take any number for themselves. Each condition was conducted under two experimental settings in a between-subjects fashion: Free (no time constraint) and Time Pressure (decision within 10 seconds). Preliminary analysis of the control condition revealed that children understood equity, as evidenced by their sticker allocation behavior. Furthermore, children rated moral transgressions as highly unacceptable and deserving of severe punishment, demonstrating a clear understanding of moral norms related to fairness and property rights. In the experimental conditions, children shared more stickers in the Take condition than in the Give condition, indicating a high sensitivity to the initial allocation of resources. This finding suggests that children perceive taking from others as a more morally charged decision, prompting them to act more generously to mitigate any internal conflict or guilt associated with taking resources from peers. Moreover, time pressure increased the number of shared stickers regardless of the experimental condition. This suggests that when children are required to make quick decisions, they might rely more on intuitive prosocial tendencies rather than deliberate cost-benefit analyses. This consistent effect across conditions is significant because it demonstrates that the influence of time pressure on prosocial behavior is robust and independent of the initial allocation of resources. This supports previous evidence, which primarily focused on the Give condition, indicating that children's instinctive behavior under time pressure is more altruistic. Therefore, our study suggests that altruism in young children is an intuitive, non-deliberative response, rather than a behavior shaped by learned cultural norms. Furthermore, a significant interaction between condition and gender suggests that males were more influenced by ownership rules, showing higher variability in the number of stickers shared, whereas females were more consistent, adhering to equity norms. In summary, this study reveals that both time pressure and the status quo of resource allocation significantly influence sharing behavior in preschool children. The findings highlight the complexity of moral and prosocial development, with gender differences suggesting varying strategies for conflict resolution. These insights have important implications for educational practices aimed at fostering prosocial behavior and moral reasoning in early childhood.

Session: Language | Room IX
Thursday September 19, 2024
10:00 – 10:25

Do neural language models have narrative coherence?

Alessandro Acciai, Lucia Guerrisi and Rossella Suriano
University of Messina

This study investigates the cognitive linguistic abilities of Neural Language Models (NLMs) by testing their skills in generating and analyzing coherent autobiographical narratives.The surprising and peculiar linguistic ability of NLMs has spurred numerous studies, wherein evaluation methods typically utilized for humans are adapted from cognitive science and in this context, ”Machine Psychology” (Hagendorff, 2023), refers to the administration of psychological tests to NLMs. The scope of this research field is continuously expanding, with notable examples including NLMs performance on personality tests such as the Dark Triad and Big Five Inventory; judgment and decision-making tasks such as the Linda problem, Wason selection task, and Cab problem; attribution of cognitive biases, creative intelligence, inductive reasoning; the emergence of Theory of Mind through unexpected contents and transfer tasks. Engaging in this debate, this work proposes an in-depth examination of the narrative production capabilities of OpenAI models GPT-3.5 and GPT-4, subjecting them to a narrative linguistic production task that, in humans, implies extensive use of cognitive abilities beyond the mere construction of grammatically and syntactically correct sentences. The ability to appropriately narrate an event, project oneself in time and space through the story, grasping the subject and giving ”meaning” to the story, is an exquisitely human capability and it requires the use of a wide range of evolved cognitive functions. Therefore, we believe that narrative coherence is one of the most effective indicators in exploring specific linguistic and cognitive aspects in NLMs. We pursued this analysis through assigning text generation tasks to NLMs that require the narration of autobiographical experiences, simulating patients in psychotherapeutic sessions. Taking as reference the sample used for the standardization of the Narrative Coherence Coding Scheme (NaCCS) (Reese et al., 2011), a method commonly used in cognitive psychology for multidimensional analysis of narrative coherence, following an induction of age, mood, and gender, NLMs were tasked to generate autobiographical stories regarding “personally” significant recent events. Subsequently, after properly training the OpenAI models, the generated narratives were automatically analyzed by ad-hoc build model according to NaCCS. The results have proven particularly interesting demonstrating a high level of global coherence in NLMs similar to that of humans, and how this is primarily modulated by the variation of age and emotional aspects.

Session: Language | Room IX
Thursday September 19, 2024
10:25 – 10:50

How does sentence specificity shape uncertainty and curiosity in conversational dynamics?

Tommaso Lamarra, Caterina Villani, Claudia Mazzuca,
Anna M. Borghi and Marianna Bolognesi
University of Bologna; Sapienza University of Rome

Abstractness/concreteness of sentences affects the underlying dynamics of a conversation. Interlocutors might linguistically align their representations more easily with concrete concepts (e.g., spoon) compared to abstract ones (e.g., belief), the meaning of which needs to be negotiated with others (Borghi, 2022; Mazzuca & Santarelli, 2023) and consequently requires more social interaction (Borghi et al., 2018) than concrete concepts. In support of this, in a recent production study (Villani et al., 2022) where participants simulated being involved in conversations, abstract sentences elicited more expressions of uncertainty (e.g., “How is that?”) and requests for additional information (e.g., “Explain to me”) compared to concrete sentences. A further dimension that might affect communication efficacy is specificity, which refers to the degree of inclusiveness that a conceptual category affords. Specificity characterizes both concrete and abstract concepts alike (Bolognesi et al., 2020; Bolognesi & Caselli 2023): footwear is a concrete generic category that includes the more specific sandal, as much as religion is an abstract category that includes the more specific Buddhism. Generic categories are typically associated with more contexts than specific categories (Rambelli & Bolognesi, 2023) and refer to multiple instances. This wide extension of generic categories arguably leads to a decrease in the speaker’s certainty on word meaning (Borghi, Fini & Tummolini, 2021).Moreover, both generic categories and abstract concepts typically are low dimensional (Langland-Hassan et al., 2021; Borghi, 2022), referring to multiple elements that share less salient features. However, highly specific categories are typically less frequent and less familiar than generic ones (Bolognesi & Caselli, 2023), involving a possible lower confidence in the word’s meaning. Notably, abstractness/concreteness and specificity have a positive but mild/medium correlation ($r = 0.29$), anda recent study (Lamarra, Villani & Bolognesi., under review) reported that these two variables display a different effect over chronometric data across a lexical processing and a semantic decision task, suggesting that these two variables do not reflect the same aspect of referentiality. Currently, there is no empirical evidence of whether and how these dimensions interact and influence the alignment of representations in linguistic exchanges. This study is built upon a previous one (Mazzuca et al., accepted for COGSCI2024 conference) and addresses these questions with two preregistered experiments asking participants to judge the plausibility of conversations composed of sentences varying in abstractness and specificity, associated with different types of possible follow-up expressions or questions: uncertainty–certainty (Experiment 1); curiosity–end (Experiment 2). This latter is currently ongoing, but preliminary results of experiment 1 showed that uncertainty follow-up/abstract sentence pairings, both generic and specific, are considered more plausible than uncertainty follow-up/concrete sentence pairings. This suggests that not the extent of the referentiality but rather the easiness in recalling perceptual referents, provide support to the alignment of interlocutors’ representation. During linguistic exchanges, concrete sentences, both generic and specific, might be easier to represent in a situation than abstract concepts. This is consistent with the idea that with abstract concepts interlocutors need more “mutual monitoring” (Gandolfi, Pickering & Garrod, 2023).

Session: Language | Room IX
Thursday September 19, 2024
10:50 – 11:15

Lost in the labyrinths of stories:
The role of negation and contradiction in LLMs' understanding of narratives

Emanuele Bottazzi and Roberta Ferrario
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

Large Language Models (LLMs) exhibit significant linguistic capabilities; however, they struggle with managing logical inconsistencies and negations effectively. This shortfall leads to what some researchers term "strong hallucinations" — outputs that are logically incoherent (Asher and Bhar 2024). This issue is indicative of a profound limitation in LLMs: the failure to recognize incompatible situations, a critical feature in human language and reasoning (Berto 2015; Simonelli 2024). It is acknowledged that humans possess a superior ability to comprehend metaphors, even in novel instances, not previously encountered; LLMs, instead, face difficulties when addressing rare creative metaphors, due to their statistical design (Mao et al 2018; Li et al 2024). We argue that this limitation arises because human understanding of metaphors necessitates the ability to discern incompatibilities. In metaphors, certain aspects are selected for their suitability to the intended expression, while incompatible elements are omitted or ignored (Black 1955; Turbayne 1970). Claiming someone is "a lion" does not imply they are a lion in every respect, but rather in specific ways and not in others. Indeed, also humans sometimes struggle to understand metaphors. However, this poses a particular disadvantage for LLMs. Among humans, the interpretation of metaphors can often be clarified through dialogue: the recipient of a metaphor has the opportunity to engage with the metaphor's sender to verify understanding. The sender's confirmation or denial is crucial in determining whether the intended meaning was grasped. This dialogic element, only superficially present in interactions with an LLM, is precluded in these models due to their deficiencies in handling negation, rendering them unable to reliably disambiguate. This limitation extends to the realm of fiction, which fundamentally involves altering reality, presenting things as they are not. Fictional narratives often juxtapose elements one against the other. In contrast, LLMs do not withstand rigorous tests concerning their narrative capabilities, either in production (Chakrabarty et al 2024) or comprehension (Subbiah et al 2024). We believe that this is because complex stories, such as novels, are sorts of labyrinths of compatibility and incompatibility. This is further evidenced by LLMs' difficulties in orientation in described physical spaces, where they can generate invalid trajectories and end up trapped in loops (Momennejad et al 2023). Rather than to incompetence in planning tasks, we attribute these deficiencies to LLMs' inability to manage situations involving incompatibility and contradiction. Finally, we will conclude with an example proposed by Jane Rosenzweig (2024), in which ChatGPT-4 was asked: 'I need to get in touch with my sister’s only sister. Who is that?' and it responded, 'Your sister's only sister would be your sister herself [...]', without considering the possibility that a brother might say this referring to one of his two sisters. This highlights a further difficulty of LLMs, namely that of constructing new models of their reasoning based on their knowledge. This issue connects with those mentioned earlier, and with LLMs’ entrapment in the intricate pathways outlined by the subtle art of storytelling.

Session: Perception & Action | Room II
Thursday September 19, 2024
10:00 – 10:25

Action in multimodal object perception

Aleksandra Mroczko-Wasowicz and Spencer Ivy
University of Warsaw

In this paper, we address the following questions: Why is action so closely intertwined with perception; and, how does this relationship influence the structure of perceptual objects? We answer these questions by surveying a wide range of empirical evidence regarding the relationship between motor control and perceptual processing. We argue that there is a significant involvement of action in the creation and constitution of perceptual objects. This is because a basic function of perception is to enable appropriate movements with respect to environmental objects. By presenting evidence from a series of neuroscientific studies, we show how the brain's motor and perceptual systems are cooperative insofar as they aid in the processing of one another. Current philosophical theories of perceptual objects notably overlook the role of action and the motor system as constitutive elements for perceptual objects' structural composition (O'Callaghan, 2016; Green, 2019; Cohen, 2023). Explanations for this omission could take several forms. It might be thought that action and perception are of two different ontological and neurophysiological kinds. However, there is a substantial body of new data challenging the foregoing claim. Recent EEG and fMRI studies have shown strong correlations between the brain's motor system and perceptual processing where the neural components of action are levied to aid in perception (Binder et al., 2004; Zekveld et al., 2006; Wu et al., 2014; Michaelis et al., 2021; Schmalbrock & Frings, 2022). Nevertheless, it might be supposed that although action and perception go hand in hand, the former does not play a necessary role for the latter. To the contrary, we argue that action contributes significantly to the creation and structural composition of perceptual objects. To achieve this, we survey a series of neurophysiological and behavioral data regarding the impact that motor control has on perceptual processing. Empirical research on (1) multimodal view independent object representations, (2) action-influenced multisensory integration within peripersonal space, and (3) the event coding of multisensory stimuli with action, collectively suggest that the motor and perceptual systems are cooperatively intertwined. Consequently, we propose that the motor system often plays a constitutive role in the construction of perceptual objects. The conclusion to be drawn from the empirical findings can be summarized as follows: Motor action is often required to facilitate the integration of perceptual features into corresponding perceptual objects. Therefore, we conclude that perceptual objects may not only have an action-oriented etiology but also that – by virtue of this etiology – they may have an action-oriented structure and functional role. We suggest that this sort of structure for perceptual objects facilitates predictive perception and affords further action. Consequently, an accurate account of perceptual objects should include the fact that they can have these foregoing distinctly action-oriented aspects

Session: Perception & Action | Room II
Thursday September 19, 2024
10:25 – 10:50

Exploring inner speech influence on novel action acquisition and execution

Angelo Mattia Gervasi, Claudio Brozzoli and Anna Borghi
Sapienza University of Rome; INSERM Lyon Neuroscience Research Center

Inner speech (IS) is a subjective experience and a powerful cognitive tool used by many people which has aroused interest in the neuroscientific community. Many studies showed that it enhances, among others, attention, cognitive control, and memorization (Fernyhough & Borghi, 2023). However, its relation with actions and motor sequences has been only marginally investigated and, mostly, in sports contexts (Hatzigeorgiadis et al., 2011). For these reasons, we want to directly investigate the role of IS in action acquisition and, consequently, action execution in a controlled laboratory setting. Considering the studies on IS modulation of attentional and perceptual processes and the studies on the role of instructional self-talk in sport, we speculate that IS might drive attentional processes related to sensorimotor aspects as well, enhancing the ability to learn novel actions. Then, we hypothesize that interfering with IS impacts participants' ability to acquire and, consequently, reproduce novel actions. To test our hypothesis, according to the power analysis, we will recruit 104 participants. Participants will be divided into two groups and asked to sit at a table in front of a screen and to learn and reproduce two actions. Both actions comprise the same four motor sequences (manipulating a cylinder on the table) but differ in terms of order and side of execution. For both groups, the task consists of an encoding phase (actions observation and acquisition through videos on a screen), a practice phase (4 trials of procedure familiarization), and a test phase (50 trials of actions execution). During the encoding phase, participants of both groups will be asked to perform a concurrent task. In the experimental group (articulatory suppression group) participants will be asked to repeat the syllable "SA" with a frequency of two syllables per second while in the control group (motor suppression group) they will be asked to perform a tapping gesture (with the same frequency of the articulatory suppression) with their middle finger on a target area on the table. We will measure action performance taking into account action accuracy (whether participants correctly perform the actions or not), reaction times (time needed to recall the actions after the go signal) and action execution times (time needed to perform the entire actions). We expect worse action acquisition and performance results for participants belonging to the articulatory suppression group compared to the motor suppression group. In other words, we expect lower accuracy (less efficient action acquisition) and slower reaction times and action execution times for the articulatory suppression group compared to the motor suppression group. Besides that, we will investigate participants' mental strategies through the Internal Representation Questionnaire (IRQ, Roebuck & Lupyan, 2020) and we expect a higher impact of the articulatory suppression on the participants' actions acquisition ability for those showing higher scores in the internal verbalization subscale of the IRQ. Data collection is in progress and fifty out of one hundred and four participants have already been tested. Data will be analyzed soon, after preregistration submission.

Session: Perception & Action | Room II
Thursday September 19, 2024
10:50 – 11:15

The evolution of syntax:
Toward a minimal model of hierarchical cognition

Giulia Palazzolo
University of Warwick

According to the prevailing view in the scholarship, syntax is a uniquely human trait, the “Basic Property” that distinguishes human language from the communication systems of other animals (Berwick & Chomsky 2016; Chomsky et al. 2023). However, a growing body of empirical evidence suggests that animals can also combine signals into larger structures (see Palazzolo & Moore forthcoming for a review). Based on this evidence, some scholars have suggested that there are evolutionary precursors of human syntax in the communication systems of other animals (Townsend et al. 2018). In contrast, other scholars have rejected this hypothesis, arguing that there are important qualitative differences between animal combinations and human syntactic constructions (Bolhuis et al. 2018). In this paper, I will reconstruct and critically analyse the debate on the continuity between animal combinations and human syntax. I will identify two key questions in this debate: what I call the empirical question and the theoretical question. I will then propose a novel account for the study of the evolution of syntax, i.e. a model of “bounded hierarchy”. Finally, I will consider evidence for bounded hierarchy in animals.

Session: Trust | Room V
Thursday September 19, 2024
10:00 – 10:25

From expert testimony to lay belief: A Bayesian view

Pietro Avitabile and Gustavo Cevolani
IMT School for Advanced Studies Lucca

Modern societies are crucially dependent on the opinion of experts. The asymmetric relationship between experts and non-experts poses many socially impactful problems that have been subjected to the methodological lens of disciplines as diverse as philosophy of science, social epistemology, argumentation theory, and social psychology. We focus on one fundamental problem that lies at the intersection of these approaches: how lay reasoners should update their beliefs in some hypothesis or claim H given that some expert asserts that H. First, we propose to treat expert opinion as evidence from testimony, developing a broadly Bayesian model of both the experts' reliability and the evidentiary impact of their testimony (in terms of Bayesian confirmation measures). Second, we show how our model impacts the discussion concerning expert opinion both in social epistemology and in argumentation theory. As for the former, we show how the idea of "epistemic deference" amounts to ignoring the fallibility of real experts and fails as a general strategy of belief updating in the face of expert opinion. As for the latter, we elaborate on the recent discussion of "ab auctoritate" reasoning to clarify and evaluate this much-disputed argumentative strategy. By adopting a model of expert reliability based on Jeffrey conditionalization, we argue for a new characterization of ab auctoritate, more faithful to the complex epistemic interplay between laymen and experts.

Session: Trust | Room V
Thursday September 19, 2024
10:25 – 10:50

Evaluating trust dynamics with dependency networks

Alessandro Sapienza and Rino Falcone
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

The dynamics of social bonds are a widely explored subject within the realm of social sciences, encompassing both theoretical and empirical perspectives. This subject holds clear relevance within these fields. Understanding how these relationships develop, evolve, and influence human behavior is crucial for a deeper understanding of society. Furthermore, the dynamics of social bonds have a significant impact on individuals' and communities' decisions. People are often influenced by the opinions and actions of their social groups. Understanding how these dynamics influence behavior is essential for more accurate predictions and for the planning of social and economic interventions. The core of our investigation revolves around the intricate interplay between dependence and trust within a hybrid society, populated by human and artificial agents. Thus, building upon our theoretical framework, within this contribution, we introduce a simulation-based implementation of dependence networks in the context of block world to investigate their utilization and resultant effects. Specifically, our focus lies in conducting a comparative analysis between the concepts of dependence and trust, examining the roles they play in shaping interactions among agents. We are interested in examining how agents achieve better results the more they are capable of choosing their collaboration partners wisely. We investigate the agents' ability to accurately identify the dependencies that spontaneously arise and to use them profitably for their own goals. In our analysis, we refer to the concepts of agent and multi-agent systems, considering in particular the BDI—beliefs, desires, intentions—model of the rational agent. Our study offers valuable insights into the utilization of dependence networks and their impact on collaborative dynamics and resource management. Most notably, agents that leverage dependence, even at the cost of interacting with low-trustworthiness partners, achieve superior performance in resource-constrained environments. On the other hand, in contexts where the use of dependence is limited, the role of trust is emphasized. These findings underscore the significance of dependence networks and their practical implications in real-world contexts, offering useful practical implications in areas such as robotics, resource management, and collaboration among human and artificial agents.

Session: Trust | Room V
Thursday September 19, 2024
10:50 – 11:15

Artificial intelligence and institutional trust: Promise or peril?

Ginevra Prelle
University of Milan

The adoption of Artificial Intelligence (AI) has exploded. Companies in the private sector—ecommerce, real estate, and even human resources—have implemented AI. The public sector is almost certain to similarly adopt AI. Countries from across the world, including Italy, are planning to do so. Profiling systems using AI are applied in the US, Finland and Germany for purposes such as child welfare or fraud detection. Some social care agencies in the UK are developing models like Amazon's Alexa to complete tasks traditionally carried out by public employees (World Bank, 2023). As the EU White Paper on AI (2020) lays out, this is the future. What consequences will such a widespread adoption of AI in the public sector have on society? This study proposes to use a vignette experiment, combined with survey measures, to identify the consequences of AI adoption in the public sector on institutional trust. Specifically, this work will explore key mechanisms driving not only trust in algorithms but overall trust in a public welfare system implementing AI to profile and categorize citizens. While there are myriad plausible consequences of AI adoption, the focus will be on institutional trust. This is because institutional trust, the confidence in public institutions that allow to cooperate as a society (Putnam, 1995), has been linked to civic engagement, policy compliance, and political legitimacy and is argued to be fundamental for maintaining healthy democracies (Bornstein and Tomkins, 2015; Edlund 2006). Two primary theories, Procedural Justice Theory (Tyler 2006) and Automation Acceptance Model (Ghazizadeh, M., Lee, J.D., & Boyle, L.N. 2012), have led existing research to focus on three main drivers of trust: perceived fairness of decision-making procedures, perceived usefulness, and ease of use. Yet, existing studies have unveiled numerous mixed findings. Some found a tendency toward algorithmic appreciation, others found algorithmic aversion. These results necessitate further research on the mechanisms shaping trust in algorithms and in organizations adopting AI. This work will address three main gaps: the efficacy of transparency measures to enhance trust, the role of awareness about AI applications and pre-existing levels of institutional trust. Transparency will be conceived as the provision of simple information about purposes and benefits of AI along with access to explanations. Additionally, we will assess how awareness of AI adoption influence institutional trust. Finally, the role of pre-existing levels of trust as determinant to AI adoption in the public sector will be disclosed. A vignette experiment combined with survey measures will be employed. Using a 2x2x2 between subject design, participants will be exposed to scenarios inspired by an analysis of the Italian policy context. Data will be collected through questions pre- and post-scenarios specifically designed based on the algorithmic perception literature. Statistical analysis will be conducted to test the hypotheses. The expectation is to inform existing theories and to unveil new mechanisms. Providing further results will contribute to the investigation of AI adoption in the public sector as a game-changer for institutional trust. Given that recent years have seen declines in institutional trust across the board in the EU, understanding the effect of AI changes is an urgent matter with deep consequences.

Symposium: Recent work in the epistemology of imagery and imagination | Room VIII
Organizer: Alfredo Vernazzani

Thursday September 19, 2024

11:40 – 12:05
Aphantasia, unconscious imagery, and rationality
Joshua Myers
University of Barcelona

Aphantasia is a condition in which subjects report having no experience of mental imagery. Yet, extensive empirical results indicate that people with aphantasia show very few impairments on tasks that are typically taken to implicate mental imagery. This gives rise to two puzzles. First, the objective performance of aphantasics on imagery tasks suggests the presence of imagery. However, their subjective reports suggest the absence of imagery. How can we reconcile this tension between objective and subjective measures of mental imagery? This is the empirical puzzle. Second, aphantasics are as reliable at forming true beliefs on imagery tasks as non-aphantasics. However, while non-aphantasics can cite their experience of imagery as their reason for holding a belief, aphantasics cannot. There is a tension between objective and subjective factors that are epistemically relevant. What is the epistemic status of the beliefs formed by aphantasics? This is the epistemic puzzle. I will argue that aphantasia involves unconscious mental imagery. This view solves the empirical puzzle by holding that aphantasics use mental imagery to perform tasks but cannot report on this imagery because it is unconscious. This view solves the epistemic puzzle because unconscious mental imagery, I will argue, can justify belief. In developing this solution to the epistemic puzzle, it will emerge that aphantasia poses a challenge to a popular family of views that ground epistemic justification in phenomenal consciousness.

12:05 – 12:30
Maps of the imagination: a theory of artifact-based understanding
Alfredo Vernazzani
University of Witten-Herdecke; Ruhr-University of Bochum

The literature in philosophy of science often insists on the role of imagination in scientific understanding (Levy & Godfrey-Smith 2020). Scientific understanding is frequently bolstered by means of models, such as the Lotka-Volterra differential equations, or Schelling's model of racial segregation. Model-based understanding is an instance of artifact-based understanding, e.g. the use of some artifacts to foster understanding. Similarly, proponents of aesthetic cognitivism often emphasize the understanding-enhancing role of artworks (e.g. Elgin 2017). But how do artifacts enhance our scientific or aesthetic understanding?
I first clarify what is meant by understanding relying on a tripartite conception (Myers & Vernazzani ms; Vernazzani ms) as opposed to a monistic conception (Bengson 2018) or dual-conceptions (de Regt 2017; Elgin 2017). On this view, to understand an object (objectual understanding), broadly understood, involves the exercise of rational capacities in navigating an epistemic space driven by some epistemic concern. Next, I motivate the research question highlighting the lacunae of extant theories. I shall advance a new account of how artifacts enhance our understanding. My account is broadly inspired by the linguist Daniel Dor's theory of language as a social communication technology (2015). On this account, artifacts such as scientific models and artworks provide instructions for the imagination. A fruitful way of thinking about such artifacts is as providing maps for navigating an epistemic space thus enabling us to chart routes through the imagination, leading us to deepen our understanding of some object.

12:30 – 12:55
Imaginative justification and imagistic reasoning
Sofia Pedrini
Ruhr-Universität Bochum

Recently, there has been growing interest in the role of imagination in the justification of our contingent beliefs about the actual world (Myers 2021; Balcerak Jackson 2018; Dorsch 2016; Kind 2016, 2018): imagining, say, jumping a certain distance justifies your belief in being able to do so. At the heart of the debate on imaginative justification lies the question of the cognitive significance of imagination. In my talk, I investigate the notion of imaginative justification using Brewer's notion of imagistic reasoning (Brewer 1999, 2005). In accordance to the view that imagination has epistemic relevance in so far as we constrain it (Dorsch 2016; Kind 2016), I argue that in the cases of contingent beliefs about the actual world we learn something from our imagination (i.e., we are

justified in believing the particular contingent belief) thanks to three factors: the presence of relevant background knowledge, the presence of perceptual demonstratives which fix the reference and individuate the object of the imaginative act employed in the reasoning (I distinguish cases in which the perceptual demonstratives are still present during imagistic reasoning, cases in which we recall them from episodic memory, and hybrid cases) and the precision of the imaginative translation. This allows me to clarify the role that imagination plays for our reasoning.

12:55 – 13:20

The epistemic role of embodiment for imagination (and its lack in AI)

Zuzanna Rucińska
University of Antwerp

The epistemic relevance and usefulness of imagination is determined by whether we can learn anything from our imaginings. Since what we imagine is up to us, how can imagination provide us with new knowledge? This talk considers how strong embodiment sheds light on the epistemic role of imaginings (Rucińska & Gallagher 2021). I focus on two aspects of strongly embodied cognition - explicit motoric processes, and neuronal processes rooted in bodily and action processes - and describe how they can play distinctive roles in constraining imagining, complementing Kind's (2018) argument for the epistemic relevance of imagination and Balcerak Jackson's (2018) argument for justification by imagination. I will then discuss potential implications the strongly embodied perspective may have on our understanding of imagination and creativity of Artificial Intelligence. While some argue that even neural-network-based AI is imagining as it can synthesize new ideas (Buckner 2024), the way the synthesis occurs for AI systems (generating novel outputs based on combining patterns and information learned from vast datasets) will be different than the way in which humans synthesize information, following the strongly embodied account. The talk aims to open discussion on the types of imaginative and creative processes that may or may not be simulated in AI models.

Session: Modeling | Room IX
Thursday September 19, 2024
11:40 – 12:05

Ranking cognitive plausibility of computational models of analogical reasoning with the Minimal Cognitive Grid: Results and implications

Alessio Donvito and Antonio Lieto
University of Bari; University of Salerno

The creation, use and interpretation of analogies and metaphors expressed in natural language sentences represents a crucial abstractive capacity of human semantic competence and communication. In the context of computational cognitive modelling, many different systems and approaches have been proposed to endow, with the same ability, artificial systems. In this paper, we review the main, state-of-the-art, computational models of metaphors by using the epistemological lens of computational cognitive science: i.e. with the aim of analyzing their level of cognitive plausibility and - as such - the explanatory power of their produced output with respect to the existing theories in cognitive science aiming at explaining such a phenomenon. More in detail, in this paper we analyze and compare the following AI systems: 1) the Structure-Mapping Engine (SME), developed by Dedre Gentner and Kenneth Forbus, this is one of the most influential systems in the field of analogical reasoning whose underlying principles of identifying systematic correspondences between different domains has also been extended to metaphor understanding, (where metaphors are seen as a form of analogy) 2) the AnalogySpace. This system uses factor analysis to represent general common-sense knowledge in a mathematical space where both analogical reasoning and metaphorical interpretation can occur; 3) CogSketch: an AI system that provides a sketch understanding environment and cognitive modeling tool supporting both visual analogical reasoning and metaphor interpretation, enabling users to draw diagrams that the system interprets, supporting reasoning about both physical and abstract concepts 4) Large Langue Models (in particular GPT-3.5 and GPT-4) used for metaphors comprehension and generation. The methodological tool adopted for our analysis is the Minimal Cognitive Grid: a pragmatic framework proposed to rank the different degrees of structural accuracy of artificial systems in order project and predict their explanatory power (Lieto, 2021). The Minimal Cognitive Grid (MCG) considers three key dimensions that characterize the relationship between a model and its biological or cognitive target system:

- Functional/Structural Ratio: This dimension concerns the balance between functional and structural components in the model. It evaluates the extent to which the model relies on abstract functional descriptions versus detailed structural mechanisms. A lower ratio indicates a more mechanistic model, while a higher ratio suggests a more functional approach.
- Generality: This dimension assesses the breadth of phenomena that the model can represent. A highly general model can be applied to a wide range of cognitive functions or biological systems, while a narrow model is tailored to a specific task or domain.
- Performance Match: This dimension involves a direct comparison between the model's performance and that of the target system. It considers not only the overall accuracy of the model but also the similarity of its errors and execution times to those of the biological or cognitive system. A close performance match suggests a higher degree of psychological or biological plausibility.

We report the obtained results, discuss the epistemological implications of such analysis, and suggest how it can inform the design of the next generation of artificial systems aiming at tackling such cognitive ability.

Session: Modeling | Room IX
Thursday September 19, 2024
12:05 – 12:30

The Minimal Cognitive Grid+, universal cognition and perceptual performance

Selmer Bringsjord, Paul Bello and James T Oswald
Science Rensselaer Polytechnic Institute (RPI); Naval Research Laboratory (USA)

Lieto's Minimal Cognitive Grid (MCG) for assessing artificial agents, augmented as the method MCG+, has two implications: (1) MCG+ can advance the mathematical science of universal intelligence/cognition. (2) (a) pre-Lieto, this science lacks of coverage of perception; (b) heralded artificial agents of today are devoid of human-level perceptual intelligence. In Cognitive Design for Artificial Minds, Lieto (2021) introduces the Minimal Cognitive Grid (MCG); applied to artificial agents1 produced by either computational cognitive science (CCS) or AI it returns verdicts regarding the intelligence and explanatory power of these agents. Lieto e.g. applies MCG to the Watson system and AlphaGo; in both, MCG reveals acute deficiencies.2 MCG applies three sub-metrics to a given artificial agent a: (i) ratio of functional to structural components; (ii) level of generality (higher the more cognitive faculties meaningfully present in the agent); and (iii) how well the agent performs, as determined by tests in line with Psychometric AI (Bringsjord & Schimanski 2003).3 We have augmented and formalized MCG, to produce a method for determining, precisely, what can be viewed as the overall cognitive power of an artificial agent. The method is MCG+; its application to the agents Lieto has analyzed with MCG yields formal outcomes concordant with his.

The first implication is that while hitherto no scholars to our knowledge have noticed that Lieto's framework relates directly to the mathematical science of universal intelligence and cognition, it does, in a substantive, consequential way. E.g., the formal theory of universal artificial intelligence given by Hutter (2005) and elaborated in (Legg & Hutter 2007), identifies the intelligence of an artificial agent a with the level of reward maximization achievable by that agent across environments, which completely ignores the rich, nuanced, and illuminating information returned by the application of MCG+ to some agent a. Put starkly and simply, it's entirely possible for the cognitive power of some agent by MCG+, Πa, to be vanishingly small, while Hutter's framework Y declares the agent maximally intelligent. This profound divergence surely must be investigated. The second implication is that if MCG or MCG+ adopted, one sees that AI agents of today receiving much attention display a serious deficiency: they are devoid of one of the chief cognitive faculties that make human agents cognitively powerful: the ability to, in an environment, attend to and perceive objects therein ways that enable and inform other cognitive faculties (such as reasoning and decision-making). he scene. For instance, without looking back at it: Were there less than seven birds? Was there a zebra right of a bird? Was there an artifact commonly used to enhance human vision below at least two objects? There are also questions used on the test from which we draw. In the larger paper summarized here, the perception lacuna revealed by Lieto's work is addressed by turning to the ARCADIA cognitive architecture (Lovett, Bridewell & Bello 2019), which places perception at the heart of the cognitive faculties.

Session: Modeling | Room IX
Thursday September 19, 2024
12:30 – 12:55

Deductive flexibility in humans and beyond:
Testing the tool with synthetic datasets

Mariusz Urbanski, Paweł Łupkowski,
Tomáš Ondráček and Ganna Stoyatska
Adam Mickiewicz University, Masaryk University, Oles Honchar Dnipro National University

Our aim in this research is to compare the results of studies involving the Deductive Flexibility test (DFT) as carried out on human subjects vs synthetic datasets created with Large Linguistic Models (LLMs) in order to study the usefulness of the latter as a reliable means to test a psychometric tool and validate already gathered results.
The name of the construct of deductive flexibility was coined by Urbański and Żyluk [2] in analogy to cognitive flexibility - an ability to switch between thinking about different concepts and thinking about multiple concepts simultaneously. Deductive flexibility can manifest in determining premises that imply a certain conclusion: this is the idea underlying the construction of the Deductive Flexibility Test (DFT). The phrase “can manifest” is used here because, although deductive flexibility could easily be characterised in logical terms (referring to the relation of logical entailment), its psychological operationalisation - in terms of a more expanded list of manifestations – still requires further analysis. DFT exhibits good psychometric properties. For example, in our two previous studies, we found its results to be normally distributed and the reliability of the tool varying between .72 (Cronbach’s alpha) to .82 (Guttman’s lambda2), on par with Raven’s Advanced Progressive Matrices (APM), used in parallel. Preliminary results of running DFT on LLMs (ChatGPT v. 3.5, 4, and 4o) suggest that, in general, they perform well in solving these types of tasks, achieving results ranging from 70% to 80% of correct solutions.
We shall compare the results of the previous studies involving DFT with the ones involving synthetic datasets [3] designed to match the sociodemographic properties of the already existing sets of human subjects. Synthetic datasets, in our case created using ChatGPT 4o, promise to achieve more representative and diverse groups of participants than those recruited using traditional methods. These make them an interesting option for the aforementioned testing of psychometric tools and validating already gathered results.
We shall create synthetic datasets to match two very different groups of human subjects. The first one consisted of 47 Polish students of different curricula at the Adam Mickiewicz University in Poznań, Poland (the study was conducted in Polish), aged 20 to 25. The second one consisted of 102 people, representatives of the educational sphere (students and teachers) of the Dnipro region in Ukraine, aged 21 to 72 (this study was conducted in Ukrainian). We shall carry out this study employing three different language versions of DFT: Polish, Ukrainian, and English.

Session: Modeling | Room IX
Thursday September 19, 2024
12:55 – 13:20

Evaluating Dream Semantics to discover patterns in personality traits and creative abilities

Aldo Gangemi, Chiara Lucifora and Claudia Scorolli
University of Bologna; Institute of Cognitive Sciences and Technologies, National Research Council (CNR-ISTC), Italy

Dreams are universal experience (van Wyk et al., 2019) involving thoughts, images and emotions (Zadra & Stickgold, 2022) capable of influencing the lives of subjects in terms of mood, problem solving and creative abilities (Pagel & Vann, 1992; Schredl, 2000; Schredl & Erlacher, 2007). Based on the correlation between creativity and personality traits of subjects (Clark & DeYoung, 2014; Lucifora et al., submitted), in this study our aim is to explore potential correlations between personality traits, creative thinking and dreams. Unlike previous studies (Klepel et al., 2019; Price, 2023), which focus either on the frequency of dreams and the ability to recall them, or on the frequency or statistical association of the words from dream reporting (Fogli et al., 2020; Elce et al., 2021), our study proposes a deep semantic analysis of dreams recorded in a dream diary. Our semantic analysis is performed automatically with both artificial intelligence tools and human validation. We consider five types of dreams (fantasy dream, observed dream, dream experience, memory, mixed dreams), nine narrative features (familiar elements, fiction elements, inconsistent narrative, suffered/observed/performed violence, metadreams, inability to do something, feelings), and two textual features (level of details and recall confidence). Additionally, we recorded the creative abilities of the subjects using the K-DOCS (Kaufman, 2012) test, and their personality traits and ability to identify emotions, using the Big-Five Inventory (John & Srivastava, 1999) and TAS20 (Taylor et al., 1992) test. So far we collected a total of 126 dreams from a sample of 21 subjects with a mean age of 23,33. A preliminary analysis on 33 dreams by 11 subjects shows good consistency among our categories: fantasy dreams positively correlate with fiction elements (r 0.902 p 0.032).

Session: Decision Making | Room II
Thursday September 19, 2024
11:40 – 12:05

Truth approximation, calibration and bias in human judgment

Davide Coraci and Gustavo Cevolani
IMT School for Advanced Studies Lucca

Human reasoning and decision-making under uncertainty is well-known to deviate from normative standards of rationality. Over the past decades, cognitive scientists have widely investigated a number of biases in human reasoning, as well as reasoning heuristics departing from theoretical prescriptions (Tversky and Kahneman 1983; Gigerenzer 2015). At the same time, philosophers have investigated the different "cognitive" or "epistemic" utilities—such as probability, accuracy, confirmation, explanatory power, truthlikeness, etc.—governing the reasoning of both scientists and laymen in different contexts (Sprenger and Hartmann 2019; Oddie and Cevolani 2022; Pettigrew 2016). The psychological and the philosophical approaches, however, have run mainly in parallel, without significant overlap despite some relevant exceptions (e.g., Crupi et al. 2008). In this talk, we put together empirical and philosophical work, with a focus on the (mis)calibration of human judgment and estimation (Yaniv and Foster 1997; Moore and Healy 2008; Fellner and Krügel 2012). As we argue, philosophical models of rational inquiry may shed light on empirical results, allowing for a better appraisal of their significance for the study of human reasoning and rationality. We proceed as follow. First, we review some results concerning overconfidence and miscalibration in judgmental estimation and forecasting, and their implications for reasoning, memory recall, and testimony in legal cases (Moore, Carter, et al. 2015; Koriat et al. 2000; Weber and Brewer 2008; Mazzoni 1996; Liberman and Tversky 1993). We also discuss some interesting theoretical models proposed by psychologists to account for such phenomena, some of which pointing to a trade-off between informative content and accuracy in human judgment (e.g., Yaniv and Foster 1995; Liberman and Tversky 1993). Second, we assess those results through the lenses of truthlikeness theory as developed in the philosophy of science (Oddie and Cevolani 2022; Cevolani and Festa 2021). A judgment or belief is truthlike (or verisimilar) when it is "close to the truth" in the sense of conveying much true information about the relevant domain. As we show, theories of truthlikeness precisely formalize the combination of accuracy and informativeness which is relevant in the empirical analysis of judgment and forecasting. To further strengthen this link, we also discuss some original (still unpublished) results concerning people's estimates of truthlikeness, and compare them to the results in the psychological literature. Finally, we discuss how truthlikeness theory can make sense of empirical results concerning the rationality of human estimation and judgement. In particular, we suggest that overconfidence and miscalibration can often be seen, in a more positive light, as rational strategies to approximate the truth under uncertainty. Some implications of our analysis for other phenomena, like the conjunction fallacy in probabilistic legal reasoning (Cevolani and Crupi 2022), are also discussed.

Session: Decision Making | Room II
Thursday September 19, 2024
12:05 – 12:30

Unmasking stress:
Gender differences in decision making under mild hypoxia

Stefania Pighin, Alessandro Fornasiero, Marco Testoni, Barbara Pellegrini,
Federico Schena, Nicolao Bonini and Lucia Savadori
University of Trento

In our daily lives, decisions often need to be made without complete certainty about their outcomes. Decision-making under uncertainty is a complex cognitive process that governs behavior, involving a blend of cognitive and emotional elements. It becomes particularly relevant in unpredictable and threatening situations, where it operates alongside stress responses aimed at monitoring and regulating internal states and external behaviors. Research has consistently shown that acute stress can profoundly influence this cognitive process; however, the effects vary between males and females. Under various stressful conditions, males often exhibit a tendency toward greater risk-taking behavior, while females tend to become more risk-averse. However, most prior investigations have primarily focused on stressors that participants are consciously aware of, such as time constraints or social pressures. The aim of this study was to investigate whether gender differences in decision-making under stress are triggered by stress awareness. To this end, the effect of a mild oxygen deprivation, a systemic stressor, on decision-making under uncertainty was explored. Indeed, one intriguing aspect of mild hypoxia is its effect on the body without conscious recognition, as individuals under mild hypoxia may be unaware of the stress they are experiencing. From an experimental standpoint, this allows for a manipulation capable of triggering a physiological stress response without conscious awareness. The present study involved 64 participants (53% female, 47% male; Mage=22.9; SD=3.4), who took part in three research sessions which were separated by a 7-day interval: a familiarization session, a control session in normoxic environment (oxygen concentration of 20.9%, simulating an altitude of 0 m above sea level), and an experimental session in mild hypoxic environment (oxygen concentration of 14.1%, simulating an altitude of 3,000 m above sea level). The sequence of conditions was counterbalanced across participants. Moreover, awareness was manipulated between participants, such that half of the participants were made aware of the oxygen levels, while the others were not informed of the manipulation (due to ethical considerations, they were informed that they could be in a mildly hypoxic environment during none, one, two, or all of the sessions). In each session, participants were asked to complete a repeated-measure version of the Iowa Gambling Task (Xiao, Wood, Denburg, Moreno, Hernandez, & Bechara, 2013). Coherently with the hypothesis that gender differences are triggered by stress awareness, the results showed that male participants made aware of the oxygen manipulation increased the number of disadvantageous, risky choices, whereas females decreased them. Both male and female participants, however, made more disadvantageous, risky choices under mild hypoxia compared to normoxic conditions when they were unaware of the oxygen manipulation. The implications of these findings for the debate on the complexity of human decision-making processes under stress will be discussed.

Session: Decision Making | Room II
Thursday September 19, 2024
12:30 – 12:55

Moral and social nudges for promoting cooperation in wicked social dilemmas: a theoretical and experimental investigation on waste sorting behavior

Sebastiano Munini, Marco Marini and Fabio Paglieri
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

Social dilemmas are described as “wicked” when they involve additional obstacles to cooperation, such as unclear individual impact, doubts about the link between cooperation and outcomes, and difficulties tracking contributions. Addressing wicked social dilemmas is not only challenging but also urgent: some of the most pressing societal problems fit well this description. Among these, are environmental problems such as waste sorting, which is the main focus of this paper. Waste sorting can be conceptualized as a cooperative behaviour in which the individual cooperates with others (e.g., neighbours) to maximize collective interest and achieve a common goal, by enabling recycling and thus reducing climate change. Despite the numerous positive effects of recycling on the environment, people are often reluctant to sort their waste, for a variety of reasons. One is the fact that human cooperation is conditional, therefore, if people experience a lack of cooperation from others, they may refrain from recycling themselves. Nonetheless, it remains imperative to incentivize and encourage participation in waste sorting, especially in environments where cooperation levels are low. Previous research has demonstrated this can be achieved by addressing economic, social, and moral factors. Given reliance solely on economic incentives poses challenges due to potential funding limitations, exploring the efficacy of cost-effective strategies such as nudges, and non-monetary drivers such as social and moral factors becomes crucial. Social and moral information is often conveyed through and embedded within social norms. These regulate behavior within a group by outlining the preferred and appropriate conduct among its members. Not only do norms limit egoism and self-interest, but also promote prosociality and cooperation. Given the relevance of social influence in waste sorting, in the present article, we study the efficacy of injunctive and descriptive norm-nudges, a sub-category of nudges that operate by utilizing social norms to influence behavior, in promoting waste sorting behavior. While moral nudges focus on injunctive norms and appeal to doing the right thing encouraging individuals to prioritize moral values, social nudges refer to descriptive norms and motivate individuals to act as the virtuous majority of people. Using a 3x2 between-factor design participants receive either a moral, social, or no nudge, and experience high or low levels of cooperation from the surrounding environment. The task used to measure waste-sorting behaviour consists of an 8-round computerized waste-sorting game in which, in each round, participants are allocated 8 kilograms of garbage, and must choose how and if to allocate their garbage in separate bins. Choosing to sort is costly in terms of time and requires waiting 10 seconds per recycled kilogram, while non-differentiating is immediate and effortless. The number of kilograms sorted is the main dependent variable. We aim to verify the efficacy of each nudge strategy to promote recycling behaviour, and to understand if strategy efficacy is moderated by the experienced level of cooperation. The outcomes of this study will help develop interventions aimed at fostering sustainable practices, and pro-environmental behaviours.

Session: Decision Making | Room II
Thursday September 19, 2024
12:55 – 13:20

Integrating VR and neuropsychometrics:
Predicting consumer preferences via submental muscle activity

Francesca Ferraioli, Carmelo Mario Vicario, Chiara Lucifora, Viviana Betti, Matteo Marucci
University of Messina; University of Bologna; IRCCS Santa Lucia

The rapid advancements in artificial intelligence (AI) have increased the need of the integration of experimental research with innovative technologies. Our ongoing research leverages virtual reality (VR) and neuropsychometrics to analyze consumer decision-making processes, utilizing electrophysiological measures such as electrodermal activity (EDA) and facial electromyography (EMG). Works form our team highlight the involvement of submental EMG in purchase decision (Ferraioli and Vicario, in Press) and more in general his linkage with reward circuitry (Vicario et al., 2014; Vicario et al., 2017; Vicario et a., 2020; Vicario et a., 2022a; Vicario et a., 2022b). Preliminary result from our study demonstrated a positive correlation between submental muscle activity and declared preferences, suggesting its potential as a novel biomarker in marketing studies. Moreover, we use virtual environment as VR ensures greater ecological validity to the experimental settings (Daher et al., 2021; Lucifora et al., 2022; Vicario et al., 2023; Waterlander et al., 2011; Melendrez-Ruiz et al., 2022; for review Wang et al., 2021). Further, recent neuromarketing studies have examined the influence of fragrances on purchasing behavior (Doucé & Adams, 2020; Mancini et al., 2021; Morrin & Tepper, 2021; Morrison et al., 2011). Taking this evidence together, in our ongoing study we aim to replicate result of our pilot study on a larger sample and adding odor influences on studied indices. The ongoing study will involve 120 healthy participants from the University of Messina, in a mixed experimental design where all participants perform the shopping task in no-odor condition (as in the pilot study), after that they will be randomly assigned to one of the two experimental groups: pleasant odor and unpleasant odor. Before the main experiment, a pre-test will be conducted to determine the most effective pleasant and unpleasant fragrances. Thirty participants will rate 20 different fragrances (10 pleasant, 10 unpleasant) using a scale from 1 to 10 for pleasantness. The fragrances with the highest and lowest pleasantness ratings will be selected for the main experiment. In the main experiment, participants will interact with a virtual supermarket using Oculus Quest 2 VR headsets and controllers. Participants will complete three types of shopping tasks designed to elicit different motivational contexts: daily shopping, where they choose items, they typically buy for daily meals; hedonic shopping, where they select items, they particularly enjoy; and dislike shopping, where they choose items they do not like. These tasks are balanced across participants to ensure consistency. Throughout the immersive VR sessions, electrophysiological measures will be recorded through BIOPAC MP36 system will be used for the simultaneous acquisition and real-time visualization of these signals, with a trigger box from Brain Trends marking events of interest. Additionally, subjective measures will be collected, participants will ask to rate for each selected product pleasantness, healthiness, desire to eat, and purchase likelihood on a 10-point scale five seconds after interacting with the product. This comprehensive methodology aims to provide robust, ecologically valid insights into consumer behavior, leveraging advancements in VR and neuropsychometric measures.

Symposium: Multimodal Integration between Perception and Action:
Cognitive, Neural, and Computational Mechanisms | Room V
Organizer: Luca Tummolini

Thursday September 19, 2024

11:40 – 12:05

Decoding haptic information and motor preparation in the early visual cortex

Simona Monaco, Luca Turella, Doug Crawford, Samantha Sartin
University of Trento

In this talk, I will present a series of research projects that fill a niche in action and perception by investigating their relationship with other forms of cognition, such as motor imagery, and by putting emphasis on the top-down aspects of neural processing. Specifically, I will review fMRI data from three experiments that span three conceptual themes of my ongoing research interests. First, I will present evidence that haptic exploration of unseen stimulus size can be decoded from the activity patterns within the primary visual cortex and as expected, in the primary somatosensory cortex. Second, I will show that action intention can be decoded as early as in the primary visual cortex even before participants start to move, and that motor preparation differentially modulates the activity pattern in early visual and somatosensory-motor areas. With the third project, I will explain how the neural representations for planning vs. imagining hand movements rely on overlapping but distinct neural substrates in the primary visual cortex. In sum, I aim to show that action is not only a product of the motor system, but rather the unitary output generated by a cascade of neural mechanisms that encompass the perceptual, motor, and cognitive domains.

12:05 – 12:30

Peripersonal space:
A multisensory interface for the interaction between
the body and the surrounding objects

Claudio Brozzoli
INSERM & Université de Lyon

Despite the fact that the space around us appears continuous, the brain distinguishes between the space near the body, also known as peripersonal space (PPS), and the space far from us. This differentiation occurs thanks to the mechanism implemented in a specific network of sensorimotor neurons located in parietal, premotor, and subcortical areas, initially identified in monkeys and later in humans. PPS can be viewed as a buffer zone between the self and the world, better preparing us for defensive purposes and for guiding voluntary actions and navigation. In this talk I will present and discuss findings focusing on the contribution of PPS to the guidance of voluntary manual actions in humans. By measuring the interaction between visual stimuli on a target object and touches on the acting hand during planning and execution of actions, we revealed a rapid remapping of peripersonal space triggered by actions. The concurrent kinematic recordings allowed us to link the multisensory modulation to the motor behavior. In conclusion, the multisensory mechanism for the peripersonal space representation, not only contributes to defensive behavior but also to appetitive actions.

12:30 – 12:55

Goal formation in multimodal space:
A topological alignment approach

Francesco Mannella, Julian Zubek and Luca Tummolini
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

The acquisition of action control rests on the ability to form goals. Goals are multimodal representations of future effects that can be used to select an action among its alternatives, guide the action toward an outcome, and regulate its unfolding until a new goal is selected. Despite their centrality in cognition, how goals develop to play such a trio of roles is unclear. In this contribution, we propose a computationally specified process model to demonstrate how multimodal goal representations can be acquired in interaction with the environment. In this process model, multimodal sensory patterns are mapped in the same low-dimensional representation space. The motor repertoire is also represented in the same space via a topological mapping. We discuss how the alignment of motor topology with sensory ones amount to a measure of agent's competence in achieving an effect that can be used to drive learning. We show in simulation that the computational mechanism of topological alignment eventually results in a multimodal system that can form and select its own goals. We conclude by discussing the biological plausibility of topological alignment as a neural mechanism.

12:55 – 13:20

From motor representations to language and back

Gabriele Ferretti and Silvano Zipoli Caiani
University of Bergamo; University of Florence

What mental states are required for an agent to know-how to perform an action? Answering this question requires establishing the nature of the mental representations involved in practical knowledge. It is commonly assumed that practical knowledge has two distinctive components: one prescriptive component, in a conceptual format, concerning what action has to be performed, and one practical component, in a motor format, concerning how to execute that action. If so, we must explain how these two components can interact. Here, we offer a unified account capable of explaining how the conceptual structures related to action can be linked to motor processing, reviewing behavioral and neurological evidence on the functioning of the motor system. This will allow us to show the relation between motor representations, language and skills, in a way that is coherent with an interdisciplinary framework covering neuroscience and philosophy of action.

Symposium: On the attribution of cognitive and emotional states
to autonomous and intelligent systems | Room VIII
Organizers: Silvia Larghi, Marco Facchin and Giacomo Zanotti
Thursday September 19, 2024

16:20 – 16:45

How people understand robots' mind:
folk-psychology vs. folk-cognitivism

Silvia Larghi and Edoardo Datteri
Università di Milano-Bicocca

We distinguish between two styles people may adopt to model the functioning of robots (and other sorts of artificial intelligent agents) in their interaction with these systems. One modeling style is based on the attribution of rationality and propositional attitudes to the system. The other, called folk-cognitivist, is more akin to the cognitivist account of the human mind, based on the functional decomposition of the system in cognitive modules processing representations. In this contribution we shed light on the characteristics of folk-cognitivism outlining and analyzing in depth some of its sub-categories, and explore the positioning of this modelling strategy with respect to Dennett's intentional and design stances (Dennett 1971, 1987). These claims will be supported with reference to the preliminary results of experimental studies of people's explanation of robot behavior.

16:45 – 17:10

Enactive Intentionality in HRI:
From Attribution to Detection

Martina Bacaro
University of Bologna

This contribution seeks to elucidate the advantages of adopting an enactive framework of intentionality, particularly in the context of HRI research. Departing from conventional views of intentionality, which center on internal mental processes, the enactive approach underscores the active engagement of agents with their environment (Hutto 2012; Di Paolo 2015). This reevaluation holds profound implications for understanding human-robot interactions and for emphasizing the pragmatic nature of intentionality. It advocates for a transition from "intentionality attribution" to "intentionality detection", thereby prompting a comprehensive reassessment of the cognitive science approach within the domain of HRI.

17:10 – 17:35

Making emotional transparency transparent

Giacomo Zanotti and Marco Facchin
Politecnico di Milano; Universiteit Antwerpen

Certain autonomous intelligent systems mimic human emotional expressions, thereby interacting in emotionally salient ways with their users. To this end, their emotionless nature must "fade in the background": whilst users may be reflectively aware of it, they pre-reflectivey interact with these systems as if they genuinely have emotional states. Facchin & Zanotti (2024) called this feature emotional transparency, but left it unanalyzed. We will fill-in this lacuna providing a rigorous definition of emotional transparency, and showing that it may be a direct, hardly avoidable, and normatively problematic consequence of widely adopted design principles in AI and robotics.

17:35 – 18:00

Substituting/complementing humans:
A cognitive and affective analysis

Guido Cassinadri
Sant'Anna School of Advanced Studies Pisa

Given that AI systems such as LLMs may induce cognitive diminishment due to excessive cognitive offloading, some suggest to use these tools in a complementary way, rather than merely in a substitutive one (Cassinadri 2024). We argue that the same should apply for affective artifacts such as social chatbots, which may cause over-attachment, potentially substituting human relationships, as well as emotional dependence (Kretzschmar et al., 2019; Vaidyam et al., 2019). On the basis of the right to mental integrity, we argue that social chatbots should be used in a complementary way, preventing the diminishment or lack of develop

Session: Perception | Room IX
Thursday September 19, 2024
16:20 – 16:45

Block on non-conceptual color perception

Ivan Cotumaccio
Université Paris 1 Panthéon-Sorbonne

Block (2023) argued that perception is constitutively non-conceptual. One of Block's central arguments for that claim seeks to demonstrate that 6 to 11 months-old infants' color perception is non-conceptual. My aim in this talk is to show that Block's argument falls short of establishing its conclusion. The argument a positive component — that 6-11 month-olds perceive colors — and a negative one — that 6-11 month-olds do not deploy color concepts. I am prepared to concede the positive component for the sake of the argument. Block's case for infants' failure to deploy color concepts relies on the results of Wilcox (1999)'s study. These results indicate that below 11 months of age infants are not able to draw on color information when reasoning about object identity during occlusion events. Block argues that infants fail to draw on color information because they fail to notice color change, where 'noticing' is taken to be a cognitive process (Block 2023: 280). According to him, the fact that infants do not engage in the cognitive process of noticing color change indicates that color concepts are not activated in color perception, and therefore that their color perception is non-conceptual. I argue that Block's interpretation of Wilcox's study becomes untenable in light of other empirical results. My argument is the following. If Block were right that color concepts are not activated in 6-11 month-olds' color perception, then infants should not deploy color concepts even in less cognitively demanding tasks than Wilcox (1999)'s. However, this is not the case. In order to show that in less demanding task than Wilcox (1999)'s infants deploy color concept when perceiving colors, I draw from the results of a study conducted by Bremner et al. (2013). This study investigated whether shape and color change affects 4 month-olds' perception of object identity by investigating whether it affects their perception of the trajectory of moving objects, as follows. Infants were habituated to an object moving behind an occluder and changing shape, or color, or both, when reappearing at the other side of the occluder. In the test phase they were shown the same type of event to which they had been habituated, but without the occluder. This time the object was moving either discontinuously (i.e. disappearing and reappearing) or continuously. Importantly, while in Wilcox's study the object went out of sight for a distance of 15.5 cm and for a duration of at least 1 second, in Bremner et al. (2013) the object was totally out of sight behind the occluder for 67 ms. The results indicate that color change affected infants' perception of motion continuity, and therefore of object identity (Bremner et al. 2013: 3). This means that infants must have noticed color change, and therefore deployed color concepts. It follows that starting from 4 months of age infants are able to deploy color concepts, and therefore that Block's argument fails to establish that infants' color perception is non-conceptual.

Session: Perception | Room IX
Thursday September 19, 2024
16:45 – 17:10

Perceiving emotions:
A multimodal approach

Niccolò Nanni
University of Lugano

In recent years, the debate on the admissible contents of perception has taken an empirical turn. Rather than relying on the armchair methodology associated with the phenomenal contrast strategy (Siegel, 2005), philosophers have been trying to determine which properties are perceived by looking at a variety of empirical phenomena studied by the sciences of the mind. The general idea behind this methodology can be summarized as follows. We can use evidence from the sciences of the mind to identify the distinctive features that characterize the processing of uncontroversially perceptual properties. Once those features have been identified, we can look at whether they are also exhibited by the processing associated with any controversial property. If they are, we can infer, by analogy, that the controversial property in question is also perceptual. Aspects of the processing of perceptual properties that philosophers have used to this end include its speed (Fish, 2013), (Smortchkova, 2017), its cognitive impenetrability (Toribio, 2018), or the involvement of the perceptual properties in adaptational effects (Block 2014), (Varga 2018), among others. Despite the popularity of this methodology, most of its applications have suffered from an important limitation: they have been set within an exclusively unimodal framework, that treats the workings of different sense modalities as, for the most part, isolated from one another. However, there is abundant evidence that challenges this unimodal picture of perception. In fact, the most accurate way to look at perception seems to be as a profoundly multimodal phenomenon, that involves a constant interaction between different sense modalities (O'Callaghan 2012, 2016), (Stokes 2014). Recently, it has been argued that ignoring the multimodal nature of perception in the debate on perceptual content is problematic, insofar as it leads to ignoring the possibility of genuinely multimodal content, over and above that associated with individual modalities (Cavedon-Taylor 2020). In my presentation, I will contend that it is problematic for an additional reason. If we want to use the processing of paradigmatically perceptual properties as a starting point in an argument for the perception of controversial properties, our picture of such processing has to be as true as possible to how it actually works. And if perceptual processing is pervasively multimodal, this will be taken into account when building our argument. In the last part of the presentation, I will show how this multimodal approach can be implemented to build an argument for the direct perception of emotion properties.

Session: Perception | Room IX
Thursday September 19, 2024
17:10 – 17:35

Amodal completion as a means to perceptual beliefs

Hamza Naseer
Università della Svizzera Italiana

A study by Cooke et al. (2015) studies the formation of memories within the visual cortex. The study was conducted on mice, and the main aim was to ascertain whether there is modification in the V1 area suggestive of visual recognition memory. The study judged whether certain cortical areas contribute to learning and memory by locally manipulating portions of the cortex involved in electrophysiological modifications that consequently "prevent or reverse memory demonstrated behaviourally." (Cooke et al., 2015, pg. 2) and found that, "experience-dependent plasticity in primary visual cortex is a substrate for visual recognition memory, manifest behaviourally as long-term habituation to familiar stimuli." (ibid) Seeing that the visual cortex does have the capacity to store memory which would help in amodal completion by allowing background beliefs to influence the filling in of partially occluded objects, we're now left with some empirical gap that needs to be filled in about familiarizing oneself with an object to amodally complete it. Hazenberg et al. (2014) showed that knowledge does have an impact on our amodal completion, "almost as early as 150 ms after prime onset." (Hazenberg et al., 2014, pg. 28). This suggests that background knowledge can play a role in our act of amodal completion quickly enough to keep up with the act of amodal completion, which, "in the early cortices happens within 100–200 milliseconds of retinal stimulation." (Nanay, 2023, pg. 76) I have made the claim that amodal completion requires familiarity as an essential component. From a purely philosophical standpoint, it seems difficult to envision that anyone would be able to perform the act of amodal completion if they were placed in a situation where the objects around them didn't adhere to spatio-temporal laws that we're accustomed to. For example, if Mr. H in this world, with its 3-D structure and physical laws, woke up one day and found himself in a world where such laws did not hold, it'd be impossible for him to comprehend the objects in such a world, let alone amodally complete them. Hence, an ability to familiarize oneself with the objects around oneself ALONG with a role played by background beliefs is necessary for amodal completion to occur. The question that arises is: when we perceive something via amodal completion, does it share that content with what is seen? Note: there is no "seeing" in amodal completion. Can one share content with what they have not seen? The radical answer is yes. After all, IF there is a mind that has arrived at perceptual belief X, and X happens to coincide with the external object X, then it does not matter what the origins of perceptual belief X are - there can be sharing with the object without direct interaction with the object. However, I will go on to argue more in favor of the "shared object approach" (Helton and Nanay, 2023, pg. 95) and highlight that the indirect relationship between external object X and perceptual belief X is suggestive of their shared content.

Session: Perception | Room IX
Thursday September 19, 2024
17:35 – 18:00

Time experiences for survival

Antonella Tramacere
University of Roma Tre

Research has not yet reached a consensus on the evolutionary function of consciousness, where consciousness here refers to the subjective aspect of an experience, to what it is like to have an experience, or phenomenal consciousness. An evolutionary investigation of consciousness is important to determine which species possess phenomenal consciousness, helping to link changes in its manifestation to ecological and neurophysiological variables. I will contribute to the investigation of the evolutionary function of consciousness by focusing on the subjective experience of the time of events. Investigation of experience of event timing is not common in the philosophy of mind; in fact, research on phenomenal consciousness has typically focused on vision and pain. I will show, however, that by focusing on organisms' perception of events duration, it is easier to demonstrate how subjective experience may have evolved as an adaptive response of organisms facing threats in the environment. Human beings perceive events as shorter or longer than they are, depending on their perceptual, cognitive, physiological and emotional state (Droit-Volet & Gil, 2009). The bodily and environmental conditions that can distort the perception of the time of events are manifold and can produce both a compression and a dilation of time. Although there are several cases of temporal distortion, a few examples may help to grasp a general dynamic.
Consider temporal binding (Hoerl, et al. 2020), which refers to the phenomenon whereby individuals perceive events occurring close together in time as causally related or as part of the same phenomenon, or conversely whereby individuals perceive the timing of events considered causally related as compressed. Although the subjective experience (of the timing of events) appears prima facie to be an epiphenomenon of the ability to group objects, there are reasons to hypothesize that this experience must have been under the effect of natural selection and thus played a role in the survival of organisms throughout their evolutionary history.
No animal could survive with time distortions that are too wide, because this would lead individuals to misjudge situations in which it is important to act quickly. On the other hand, perceiving certain events as occurring over a shorter time span could be useful for anticipating behavioural reactions in a range of potentially dangerous situations. These observations allow us to make some predictions. Firstly, for temporal distortions of events to be adaptive, they cannot involve too long-time intervals, as this could lead individuals to be de-synchronised in a potentially deleterious way by the dynamics of the social and non-social world. Secondly, in situations where reacting faster has a high priority, individuals should perceive events as more compressed in time than they actually are. I will conclude by showing that these predictions are supported by evidence and offer concrete proposals for testing the phenomenon of temporal binding in non-human animals.

Symposium: (Allegedly) AI-generated media: how do they make us feel? | Room II
Organizers: Dominique Makowski & Marco Viola
Thursday September 19, 2024

16:20 – 16:45

Are androgynous faces uncanny?

Antonio Olivera-La Rosa
Universidad Católica Luis Amigó

Recent research suggests that the uncanny valley hypothesis may constitute an insightful framework for explaining negative social inferences from faces. One potential explanation is that stimuli that are difficult to categorize may evoke a negative emotional response. Do androgynous faces follow a similar pattern? Some studies showed that the challenge of categorizing androgynous faces into binary sex categories serves as a metacognitive factor that can contribute to negative social judgments. We conducted cross-cultural research to explore how categorical uncertainty influences social judgments of androgynous faces. Method: Categorical uncertainty was measured using reaction times in the Face Evaluation Task. Trustworthiness, creepiness, and perceived shared moral values were measured using Likert scales. In Study1 (N=76), androgynous faces were rated as more trustworthy, less creepy, and more morally similar compared to sex-typical faces. Although androgynous faces were more difficult to classify into a binary sex category (female vs. male) than typical-sex faces, this cognitive difficulty did not affect the social judgments of the faces. Similar results were found in Study2 (N=45). Study 3 (N=85) revealed an overall positive bias towards androgynous faces, particularly compared to male targets, even after accounting for morphing procedures in stimuli selection. This research indicates that under certain circumstances, a positive social bias towards androgynous faces can exist independently of categorical uncertainty.

16:45 – 17:10

Emotional response toward fiction and the underlying cognitive mechanisms

Marco Sperduti
Université Paris Cité

In philosophy there has been a long-lasting debate on the nature of emotion toward fictional characters and events. Neuroscientific studies have underlined that processing of fictional entities reduces the recruitment of cortical regions involved in self-referential processing, and boost activity in fronto-parietal regions involved in executive processes. In a series of studies we tested the hypotheses that appraising a stimulus as fictional would downregulate emotional reaction, and that this effect would be modulated by selfreferential processes and interindividual variability in executive functions. In all the studies we presented emotional laden material – videos (Study1, N=29) or pictures (Study2, N=37; Study3, N=33) – as either fictional or real. We collected subjective, physiological and neuronal data characterizing participants' emotional reaction. In all studies, we reported that negative stimuli presented as fictional were subjectively appraised as less intense and less negative. Moreover, we reported that they elicited lower physiological (skin conductance, heart-rate deceleration), and neural response (late positive potential). We also showed that self-referential processes complexly modulated the effect of fictional appraisal. Finally, we reported that the amount of down-regulation in the fictional condition was predicted by interindividual variability in updating performances. We will discuss the relevance of our findings for the current debate on the negative bias toward AI-generated material.

17:10 – 17:35

Real is the new sexy

Alessandro Demichelis
IMT School for Advanced Studies Lucca

Extant studies suggest that believing that faces are or can be fake decrease their trustworthiness, whereas allegedly fake non-sexual videoclips are less arousing than real ones. Thus, to test whether the "purportedly unreal = less arousing" effect also applies to sexual contents, we performed 2 pre-registered online studies. In Study1, participants (N=57) saw 60 images of male or female models in underwear. They were told that pictures could be either genuine photos or AI-generated (unbeknownst to them, all stimuli were real). For each stimulus, they were asked both whether the stimulus was real and how much they were sexually aroused by that image. In study 2, participants (N=108) were shown the same images, but this time they were presented with 30 real and 30 allegedly artificially generated pictures in two different blocks. They had to rate their sexual arousal. We found that realness, be it self-rated (Study1) or experimentally manipulated (Study2), was a significant positive predictor of higher

arousal. These findings have significant implications for understanding the impact of deepfakes and suggest a robust correlation between the assessment of authenticity and the potential for experiencing arousal.

17:35 – 18:00

MusicAI bias: listeners like music less when they think it was performed by an AI

Alessandro Ansani
University of Jyväskylä

The notion that contextual cues influence aesthetic judgments is known both in visual art and music. Contextual cues influencing how we judge music involve knowing the composer's identity and personality. Nowadays, in the AI era, it has been shown that people like music compositions less when they think (or are told) that they are composed by AI. Musical composition might be conceptualized as a somewhat algorithmic activity, where some degree of schematicity must be maintained beyond mere creativity, possibly with no emotion involved. Indeed, algorithms have been created which compose pieces à la Bach, hardly distinguishable from Bach's actual pieces. On the contrary, the performing act seems to be a solely human endeavour, harder to be imitated credibly by non-human entities. In the current cross-over design experiment, the performative act is analyzed. The participants (N=50) rated three videos of classical musical performances in two different versions: in one, a professional pianist sat on a piano, pretending to play; in the other, the same (reproducing) piano played the piece automatically, allegedly thanks to an AI. Actually, the audio was identical in both versions. Irrespective of musical training and attitudes toward AI, the participants rated the music as more likeable, engaging, higher in emotional valence, and of higher musical quality when the pieces were "performed" by the pianist. Interestingly, when asked what differences they noticed between the two renditions, participants confabulated about dissonances, tempo variations, differences in rhythm and dynamics.

Symposium: Music perception and cognition:
crossmodal, cross-cultural, and cross-species approaches | Room V
Organizer: Nicola Di Stefano
Thursday September 19, 2024

16:20 – 16:45
Crossmodal associations involving musical stimuli
Cross-cultural evidence
Nicola Di Stefano
Institute of Cognitive Sciences and Technologies
National Research Council of Italy

16:45 – 17:10
Music perception and action:
Embodiment, dyadic dance, and interpersonal synchronization
Giacomo Novembre
Neuroscience of Perception and Action Laboratory (NPA Lab)
Istituto Italiano di Tecnologia

17:10 – 17:35
Rhythm and sound production across species
Andrea Ravignani
Department of Human Neurosciences
Sapienza Università di Roma

Music perception and cognition encompass the intricate processes through which humans perceive, interpret, and respond to music. At its core, music perception involves the processing of the fundamental dimensions of auditory stimuli, including pitch, rhythm, melody, harmony, and timbre. These elements are integrated by the auditory system to create a coherent musical experience.

Cognition refers to the processes that lead to understanding, interpreting, and making sense of music. This includes higher-level cognitive functions such as memory, attention, expectation, emotion, and decision-making. For example, listeners use memory to recognize familiar melodies, attention to focus on specific musical elements, expectation to anticipate upcoming musical events, emotion to interpret and respond to the affective content of music, and decision-making to evaluate musical preferences or make judgments about musical structure.

Music perception and cognition are influenced by several factors, such as individual differences in auditory processing abilities, cultural background, musical training, and personal preferences. To account for all these factors, research draws from disciplines such as psychology, neuroscience, cognitive science, musicology, and ethnomusicology to investigate the underlying mechanisms and processes involved in music perception and cognition.

Valuable insights into human music perception and cognition arise from comparing listeners with diverse cultural backgrounds and languages, aiming to reveal music universals across cultures. For similar purposes, researchers have investigated rhythmic and melodic abilities in non-human species. Additionally, crossmodal approaches seek to uncover domain-general processing mechanisms that hold across the senses.

During this symposium, speakers will present examples illustrating how crossmodal, cross-cultural, and cross-species approaches contribute to our understanding of music perception and cognition. **Nicola Di Stefano** (CNR) will present cross-cultural studies examining crossmodal associations with auditory stimuli, highlighting the role of shared emotional meanings in mediating these associations. **Giacomo Novembre** (IIT) will offer an overview of his recent research on music perception and action, focusing especially on interpersonal synchronization in dyadic dance. **Andrea Ravignani** (Sapienza) will explore rhythm and sound production across species, including non-human primates and seals, from an evolutionary perspective.

Symposium: Ethical and cognitive perspectives on socio-technical hybrid societies | Room IX
Organizer: Ludovica Marinucci
Friday September 20, 2024

10:20 – 10:45

Decision-making and self-control with AI in the loop

Vieri Giuliano Santucci
Institute of Cognitive Sciences and Technologies
National Research Council (CNR-ISTC), Italy

As argued by the anthropologist Leroi-Gourhan, technologies can be interpreted as an externalization of human skills to tools, in order to efficientize processes and free up time and cognitive resources. The most recent advances in Artificial Intelligence fit into this framework, but they exasperate its effects and, above all, its potential consequences o human cognitive capacities such as decision-making, and more generally on the very concept of autonomy, which is being questioned and redefined precisely with AI. Adopting a conception of autonomy as self-control derived from Dennett's theory, this contribution will examine (as an example of how AI is affecting human autonomy) the ambivalent effects that recommender systems and generative AI can exert on decision-making and creative dimensions. The former can in fact play a dual role with respect to decision-making autonomy: by filtering information, they can both increase self-control in decision-making and act as mechanisms of distraction, attention control and exploitation, thus blocking degrees of freedom to exert a kind of remote control over the human user. Regarding generative AI, it can be seen both as a powerful selection and suggestion system - similar to standard recommendation algorithms - and as an information production tool, thus opening up new perspectives in terms of creativity and autonomy.

10:45 – 11:10

The ethics of using large language models to predict patients' preferences: a proposal

Marco Annoni
Centro Interdipartimentale per l'Etica e l'Integrità nella Ricerca, CNR

The contemporary model of clinical decision-making is largely based on informed consent and a procedural understanding of patient autonomy. However, patients may sometimes be incapacitated, thus precipitating the question of how their treatment preferences ought to be reconstructed in such cases. Advance directives (ADs) and surrogate decision-making (SDM) provide common means of determining patient preferences in these situations, but they are both problematic. Only a tiny percentage of patients compile an AD, while SDM often relies on uninformed and precarious guesses by surrogate decision-makers. Moreover, in any case, treatment decisions are often difficult due to cognitive biases, poor clinical knowledge, cognitive limitations, and lack of ethical preparedness. To solve these issues, it has been proposed to develop "Personalized Patient Preference Predictors" (or P4s) using large language models and other generative models of artificial intelligence. Combining different training methods, these tools have been hailed as capable of predicting patients' preferences in a way that matches or surpasses in accuracy surrogate decision-makers – and perhaps even patients themselves. In this talk, I explore the ethics of P4s, charting the different practical, ethical, and legal challenges that the development and implementation of such tools may have in healthcare. As I will conclude, while P4s to enhance standard advance directives are rather uncontroversial, the development of P4s to support substitute judgments requires a more prudent approach, as it may infringe substantively on the respect of patient autonomy.

11:10 – 11:35

Can we get rid of empathy in AI-driven healthcare?

Elisabetta Sirgiovanni
Sapienza University of Rome

Common ethical concerns surrounding clinical AI include that AI systems may eradicate human empathy, thereby dehumanizing the doctor-patient relationship. It is believed that this could compromise the trust between doctors and patients, ultimately hindering positive therapeutic results. This talk challenges the idea that we should incorporate empathy in artificial devices used in healthcare. It will be argued that natural empathy, while often touted as essential, carries its own set of negative aspects such as bias, favoritism, and paternalism (i.e. interfering with the patient's individual choices in a kin-like manner), which could limit patients' autonomy and equality and their right to receive adequate treatment and care. These inherent flaws in human empathy, rooted in neuro and psychological processes, are likely to be mirrored in AI devices, raising doubts about the wisdom of fostering

empathic relationships between AI systems and patients. Highlighting the propensity for biases to manifest throughout the AI information-processing stages, the paper underscores the challenge of identifying and addressing such biases due to the inherent opacity of AI systems. Additionally, human-machine interactions often perpetuate biases rather than mitigating them, further complicating the issue. Despite critiquing empathy's efficacy as a moral driver, the paper does not advocate for the complete dismissal of affective AI programs in healthcare. Instead, it suggests exploring alternative moral emotions such as sympathy, which involves a more detached commitment to others' welfare, potentially mitigating paternalistic tendencies even in AI applications. Moreover, the paper emphasizes the necessity of establishing some form of direct normativity for clinical AI systems.

11:35 – 12:00

Ethical framework for deception in human-robot interactions

Ludovica Marinucci

Centro Interdipartimentale per l'Etica e l'Integrità nella Ricerca, CNR

Telling lies, especially white lies, is common in human interactions. This emotionally deceptive communication serves important pro-social role functions that are also becoming relevant in human-robot interactions. A notable example of “self-deception” was people’s reaction to Weizenbaum’s Eliza chatbot: even people who knew quite well that it was just a keyword-based rule program became very fond of it. The Eliza effect has been exploited not only in newer and powerful chatbots like ChatGPT, but also in social robots designed specifically so that humans project emotional states onto the robot. Therefore, despite their effectiveness in monitoring health and providing companionship, ethical concerns have been raised about such technology, including deception and infantilization. The talk will focus on examples of ethically ambiguous situations in order to illustrate how different types of deceptive behaviors of artificial agents (e.g., tactful deception, nudging, self-deception, etc.) can be acceptable and even desirable by users. The aim is to challenge the idea that the morally upright approach of non-deception is not always useful in some situations. Therefore, we should understand if and how to incorporate emotionally deceptive behaviors in artificial agents, taking into account that different cultures have very different values and attitudes when it comes to dealing with ethically ambiguous situations.

Session: (Anti)Representationalism | Room II
Friday September 20, 2024
10:20 – 10:45

Representational realism is not a tenet of cognitive science

Claudio Fabbroni
Humboldt Universität zu Berlin

Mental representations have a prominent role in cognitive science. A widely accepted idea is that they do so because the standard position among theorists is robust realism, according to which representations would be understood as real entities that bear cognitive content due to their physical properties. This presentation challenges such a view and has a twofold focus: firstly, I argue that robust realist positions seem to have unavoidable shortcomings and cannot satisfactorily explain the multiplicity of empirical data; secondly, I claim that the realist stance is not necessary nor a central working assumption in cognitive science, because cognitive scientists do not normally have a position so ontologically committed. Rather, I argue that a pragmatic approach, which understands representations not as real objects but just as useful heuristic concepts, is closer to both the practice and data of neuroscience. The talk is divided in three brief sections. In the first one, I recapitulate the realist's claims on the neural realization of mental representations. Then, through some case studies, such as place cells, which realists like Nicholas Shea hold to be a paradigmatic example of a realized representational relation between a neural subsystem and its target, I argue that realist positions cannot expound the diversity of empirical data even in such a "paradigmatic" example. Moreover, I demonstrate, through some recent surveys on the use of the concept of "neural representation" among hundreds of psychologists, cognitive scientists and neuroscientists, that realism does not seem necessarily implied in cognitive science research. In the third part, I suggest an alternative approach: the pragmatic (or deflationary) one, claiming that it aligns more with neuroscientific practice and data. It is thus concluded that if there is a "standard scientific position" on mental representations, this appears to be more pragmatist than robust realist.

Session: (Anti)Representationalism | Room II
Friday September 20, 2024
10:45 – 11:10

The propositionalist view on emotion and its relevance for emotional attributions to robots in HRI

Ivan Zanzarella
University of Bari

Emotions are complex phenomena, involving both mental and physical components (see e.g. [Scarantino and de Sousa, 2021]). As mental states, they are generally classified within the range of intentionality: Emotions seem in fact to have that "aboutness" which characterizes all intentional mental states (see e.g. [Brentano, 1874; Goldie, 2002; Crane et al., 2009]). If A fears b, for instance, A's fear is directed at, is about, b, which amounts to the object of A's emotion. Most emotional intentionalists, however, deny that propositions can be the object of emotions – emotions, instead, are even evoked as an evidence for the admissibility of non-propositional attitudes (see e.g. [Grzankowski, 2015; Grzankowski and Montague, 2018]). For example, take again A's fear of b: "Fearing" can be hardly conceived as a propositional attitude (similarly, say, to "believing"), nor can be the object b of fear entirely captured by a proposition. Yet, beyond all the phenomenological, physiological and expressive-behavioral elements which constitute an emotion, also evaluative and aspectual representations of intentional objects are involved in emotional states (see e.g. [Green, 1992, 61-76]). In the case of A fearing b, for instance, A not only experiences unpleasant (subjective) feelings towards b, has increased heart rate and flees in presence of b, but also appraises b as displeasing, horrible, frightful, etc. Now, if so, emotions could be treated, at least in part, as propositional mental states –which would indeed falsify "hard" non-propositionalist accounts of emotions (I do not namely claim that emotions can be entirely reduced to propositions). A's aspectual assessment of the emotional object b, in fact, can assume the form of a proposition in the sense of A believes that <b is unpleasant>, whereby <p> is a proposition. Being in an emotional state for A, in other words, also amounts to having beliefs about b, whereby beliefs are of course attitudes towards propositions. Having such beliefs about b is also what establishes the conditions of appropriateness of A's emotional state towards b. Beliefs, in fact, can be true or false of b, and if A believes that <¬ p>, this means that A's emotion towards b is inappropriate. Furthermore, it can be argued that evaluative aspects of emotions also involve implicit reasoning based on propositionally structured background knowledge: For B(x): x is a bear, F(x): x is frightful, and b ∈ B, for example, it holds that

1. $\forall x(B(x) \rightarrow F(x))$
2. $B(b)$
3. $\therefore F(b)$

whereby A fearing b (also) consists in A's believing (3) relying on a more general background knowledge (1) of propositional form. It seems in fact difficult to think that –at least for some emotional states– A can have emotions towards b without having previous structured knowledge about b (whereby knowledge can be but propositional).
The aim of this proposal is to investigate the relation between emotions and propositionality. By defending a judgementalist account of emotions, I will claim that emotions can be treated –at least partly– as propositional mental states because of evaluative components and background knowledge. Furthermore, I will examine the implications of the propositionalist view on emotions on the research about the attribution of emotional states to robots in the field of HumanRobot-Interaction (HRI). Since robots have entered human social context, several scientists within HRI have been trying to understand how the attribution of mental states to robots works (see e.g. [Thellman et al., 2022]). In this respect, (minor) efforts have been made also for understanding how humans attribute specifically emotions to robots. In most cases, however, the (empirical) research in the field failed to recognize the complexity of emotional phenomena as displayed by philosophers and cognitive scientists. Emotions have been often indistinctly subsumed and investigated under the general category of "mental states" together with beliefs, knowledge, desires, intentions, etc. (see e.g. [Brüne et al., 2007]). Moreover –from within an implicit and uncritical acceptance of a reductivist (and maybe non-intentionalist) theory of emotion– the emergence and the attribution of emotional states have been mostly correlated only with physical and bodily factors, without sufficient attention to the other phenomenological, evaluative, expressive-behavioral elements however constitutive of those particular states [Hortensius et al., 2018]. Here, my aim is to contribute to establishing a more precise epistemology for HRI research about human attribution of emotions to robots. In particular, my investigation will be directed towards the relation between emotional and propositional attributions to robots: For A : Robot, does attributing to A fear for b somehow relates to attributing to A the belief that <b is agreeable>? Is the attribution of belief logically implied or presupposed by the attribution of emotion? Does emotional attributions to robots necessarily involve the attribution of other (propositional) mental states to them at all?

Session: (Anti)Representationalism | Room II
Friday September 20, 2024
11:10 – 11:35

Grounding values in which environment?

Francesco Abbate
Sapienza University of Rome

I will focus on the key role played by normative capacities within a perspective based on grounded cognition and radical enactivism. I will argue that, accepting this theoretical framework, a homeostatic approach coupled with sensorimotor capacities is unavoidable and should be pursued in robotics. In fact, it is a matter of developing intrinsic goals and capacities starting from sensorimotor capacities, e.g. object avoidance, but also of attributing non-arbitrary values to the external objects, particularly to salient objects (Craig, 2009). I will argue for the importance and need for self-attunement (Gibson, 1966) with the environment in order to ground one's values in a non-arbitrary way. I will define this self-attunement as operationally conditioned semantic or homeostatic sense-making (Di Paolo, 2005). For example, the fact of understanding sucrose as a nutrient is not an immediate data for bacteria but something that must be deciphered by them thanks to their metabolic capacities. Sucrose has significance as food only in connection with such metabolic capacities, thus in the milieu that the organism brings into existence (Thompson, 2007). Two fundamental consequences follow from this. The first refers to the fact that both cognitive capacities and the agent's milieu emerge thanks to value attributions which are always particular, i.e. they refer to specific requirements of that agent and of a certain milieu. The second refers to the fact that there are also universal values which are not negotiable, i.e. are not subject to value attribution but constitute the condition of grounded value attribution. In order to define a rigorous criterion for identifying these universal values, I will draw on the materialist tradition derived from Spinoza (1992), according to which all things are conative and responsive, possessing the inherent drive to persist in their being and to be affected and to affect in turn. From this distinction I will draw on a recent proposal by Blok (2024), who suggests the existence of things that partly escape this materialist definition, that is, that are conative but not responsive in the sense that they affect without being affected in turn. He calls it the domain of the elementary Earth in which he includes the geosphere of tectonic plates and oceans. I propose to add to the latter, gravity and the second law of thermodynamics, understood not as universal laws but as processes inherent in all things. I will define the ability to tune in to this non-negotiable nor manipulable domain as the prerequisite for being able to subsequently properly ground normative capabilities, defining what could be called as Value Grounding Problem.

Symposium: The social media debate:
Do social media really represent a threat to our society? | Room V
Organizer: Alberto Acerbi
Friday September 20, 2024

10:20 – 10:45

The causal impact of Instagram usage on psychological well-being

Valerio Capraro
Università di Milano-Bicocca

Numerous studies have identified a negative correlation between social media usage and psychological well-being. However, the causal relationship between these factors remains a subject of ongoing and intense debate; it is indeed possible that people with lower psychological well-being might be more inclined to use social media. This study presents the findings from an experimental design in which participants were divided into two groups: one group was asked to reduce their Instagram usage for four weeks, while the control group maintained their usual social media habits. Post-intervention assessments revealed that participants who decreased their Instagram usage reported significantly higher levels of happiness and a stronger inclination towards finding deeper meaning in life, alongside lower levels of materialistic attitudes. These findings suggest a causal link between reduced Instagram usage and improvements in several aspects of psychological well-being.

10:45 – 11:10

Does the problematic use of social media constitute a pathological condition? Possible underlying psychobiological mechanisms

Tania Moretta
Università di Padova

The extent to which problematic use of social media (PUSM) constitutes a psychopathological condition remains under discussion, with some researchers proposing that social media may work as a vehicle for expressing an individual's addictive focus on specific behaviors, which represent the “true” problematic focus. The question is whether people who may be addicted to online behaviors are those who may also be addicted to the same offline behaviors. In this view, it would be important to characterize how online behaviors differ from their offline counterpart. One of the key features of PUSM, which is often not shared with any other offline behaviors, is the number of available visual, auditory, and tactile cues to which many users are frequently exposed. This feature may help answer why some people compulsively search/”surf” online. It can be hypothesized that conditioned environmental cues significantly influence online behavior by promoting early attentional bias to social media rewards and enhancing conditioning to such rewards. The incentive properties of social media-related cues leading to urges to use social media and future directions for the clinical characterization of the PUSM will be highlighted.

11:10 – 11:35

The skepticism puzzle: A critical examination of disinformation intervention effects

Folco Panizza
IMT School for Advanced Studies Lucca

In this commentary we discuss one of the unintended consequences resulting from interventions aimed at contrasting disinformation, namely increased skepticism towards reliable information. While the goal of an intervention should be to increase true beliefs and reduce false beliefs, there is evidence that some interventions are successful at reducing false beliefs because they reduce beliefs overall. We review evidence from the literature finding this heightened skepticism, and distill a set of recommendations for addressing this phenomenon: targeting fake news super-spreaders, promoting the recognition of valuable and trustworthy news content, and providing the correct base rate of misinformation prevalence. Finally, we argue that tackling misinformation requires a nuanced approach that goes beyond fast and frugal solutions. Instead, we suggest prioritizing strategies that strike a balance between reducing susceptibility to misinformation and preserving individuals' ability to evaluate truthful news critically. We advocate for strategies that promote a more discerning and informed public in the evolving landscape of online information.

11:35 – 12:00

The social media debate: A roundtable

Moderator: Alberto Acerbi