ISPSM 2025

ABSTRACT BOOK

INTERNATIONAL SOCIETY FOR THE PHILOSOPHY OF THE SCIENCES OF THE MIND

3RD ANNUAL WEB CONFERENCE

24-26 October 2025

ISPSM

www.ispsmind.com

INTERNATIONAL SOCIETY FOR THE PHILOSOPHY OF THE SCIENCES OF THE MIND

Funded in 2023 By Gualtiero Piccinini & Inês Hipólito
Current President: Fabrizio Calzavarini
Vice-president: Marco Facchin
Secretary: Liza Skidelsky
Council Members: Don Ross, Krushil Watene, Nir Fresco, Nicole Rust, Katsunori Miyahara, Inês Hipólito
ISPSM 2025 program chairs: Valeria Becattini, Liberty Severs
ISPSM 2025 program committee: Valeria Becattini, Lara Bräuchle, Jimena Clavel Vázquez, Niccolò Negro, Darius Parvizi-Wayne, Liberty Severs
Acknowledgements: The ISPSM sincerely thanks all the people that contributed to this conference by reviewing the submitted abstracts: Adrian Wieczorek, Adriana Alcaraz Sanchez, Alberto Barbieri, Alfredo Tomasetta, Ali Boyle, Andra Meneganzin, Andrea Blomkvist, Andrea Raimondi, Anil Patnaik, Anna Ichino, Antonella Tramacere, April Owens, Bartosz Radomski, Ben White, Benedetta Cogo, Carl Sachs, Chiara Caporuscio, Daniel Gregory, Daniel Munro, Danielle Williams, David Schoute, Davide Coraci, Dimitri Coelho Mollo, Edoardo Datteri, Elizabeth Waldberg, Ema Demsar, Eugenio Petrovich, Fabrizio Calzavarini, Farid Masrour, Federico Fantelli, Francesco Ellia, François Kammerer, Gabriele Ferretti, Giacomo Zanotti, Giulia Di Rienzo, Giulia Leonetti, Giulia Piredda, Gualtiero Piccinini, Guido Cassinadri, Hadeel Naeem, Hamza Naseer, Itamar Dreymann, Iwan Williams, Jabran Amanat-Lee, Jacopo Colelli, Janko Nešić, Jelle Bruineberg, Joel Krueger, Juan F. Álvarez, Keith Raymond Harris, Kimberly Van Orman, Lorena Sganzerla, Louis Loock, Lu Teng, Lucy Osler, Luis Favela, Mads Dengso, Marc Artiga, Marcin Miłkowski, Marco Degano, Marco Fasoli, Marco Viola, Mariela Destefano, Marta Caravà, Marta Pérez-Verdugo, Matej Kohar, Matteo Colombo, Matteo Grasso, Michelle Liu, Michiel Esseling, Miguel Segundo Ortin, Monika Dunin-Kozicka, Nara Figueiredo, Nathalia de Avila, Niccolò Nanni, Oryan Zacks, Paride Del Grosso, Pii Telakivi, Rafael Ambríz González, Raphaël Millière, Ritwik Agrawal, Sabrina Coninx, Sacha Behrend, Sanja Sreckovic, Shiho Sugiura, Silvia Larghi, Souichiro Honma, Tomy Ames, Uku Tooming, Urte Laukaityte, Valentina Martinis, Valentina Petrolini, Vicente Raja, Victor Kumar, Wanja Wiese and Wouter van Hooydonk.

Table Of Contents

Forewords

The ISPSM was created in 2023 by Gualtiero Piccinini (University of Missouri) and Inês Hipólito (Macquarie University), with the aim of creating the first truly global society devoted to the mind sciences and their philosophy.

Central to the society are the principles of inclusivity and diversity. The ISPSM is thus devoted to removing the geographical, social and economic barriers preventing knowledge exchange and participation in (philosophical) dialogues. This is why our annual meeting has been - and will always be - online: to eliminate registration fees and travel costs that impede a truly free knowledge exchange.

Introducing this 3rd annual meeting, we see a lot of new scholars who submitted their abstract for the conference. In fact, this year submissions have increased a lot, both for individual talks and for symposia. We are delighted to welcome you all. We take your presence to be a sign that the ISPSM is gaining momentum and notoriety in the philosophical community. But in this sea of new participants, we are also happy to see the well known names of those who participated in our yearly meeting from the first year. We are enthusiastic to welcome you back, and we hope you will notice that we really made use of your feedback from last year. In particular, we focused on making less numerous but denser and more thematically cohesive parallel sessions - which we hope they will make your experience in this yearly meeting even more enjoyable.

We mentioned that inclusivity and equality are central to the ISPSM. We have acted to make our commitment to these values more evident. The organizing committee is no longer a bunch of white males (finally!), and we are committed to expanding it, so as to have an organizing committee that's more representative of various geographical locations and their needs. We have also introduced a special keynote symposium to promote the work of what we feel are some under-appreciated junior scholars. Lastly, the fast turnaround of charges has been formalized, thereby allowing junior scholars to reach responsibility positions in the society.

There is, of course, still a great deal of work to do. For example, the conference is still centered on Europe - and so candidates from the East and the Americas are at times asked to give their presentations very early in the morning or late in the evening. The organizing committee still has some trouble recruiting BIPOC participants and keynotes. The ISPSM is still monolingual, thus excluding people that do not speak English. These are some of the issues that lie ahead - and that we are committed to address.

In the meanwhile, we wish you a pleasant conference.

Fabrizio Calzavarini & Marco Facchin


ISPSM

Keynote talks

Sabrina Coninx - Disruptive technology and Human Niche construction

Vrije Universiteit Amsterdam

In ecology, disruptive innovations refer to technologies that disturb existing market structures and alter competitive dynamics (Christensen & Bower, 1996). In the philosophy of technology, this concept has been expanded to include technologies that overturn entrenched norms and practices across social domains - economic, scientific, legal, ethical, or conceptual (Hopster, 2021). In situated cognition, technologies are also viewed as disruptive on a personal level, concerning cognitive, affective, epistemic, and interpersonal dimensions of individual lives (Fabry & Alfano, 2024; Krueger, 2024).
Unfortunately, there are crucial conceptual gaps in these still isolated fields of research, to be characterized along five guiding questions:

1. What is disrupted by 'disruptive' technology?
2. What does it mean for a technology to be 'disruptive'?
3. Why are we as humans particularly sensitive to technological disruptions?
4. Are technological disruptions inherently positive or negative?
5. How can we adapt to disruptive technologies, and to what extent might this exacerbate inequality?

To tackle these questions, we propose examining disruptive technologies through the lens of niche construction theory (Coninx, 2023; Fabry, 2021; Odling-Smee et al., 2003; Sterelny, 2018). This approach helps to provide a unifying framework of what makes technology socially and personally disruptive. In more detail, we are for the following claims:

1. Technological disruptions are specific forms of niche disruptions: dynamic processes that interrupt or overturn stable patterns in a social group's or individual's interaction with their environment.
2. The disruptiveness of technologies is best understood along multiple dimensions such as depth, range, pace, reversibility, and meaningfulness.
3. Human niches are inherently technological, making them especially sensitive to disruptions caused by technological change.
4. Technological disruptions create adaptive pressures at both individual and societal levels, with significant potential for cognitive and affective harm.
5. The extent of harm depends on the resources and abilities available to protect or reconstruct niches, often reflecting and reinforcing social inequalities.

Zoe Drayson - Mind Beyond mechanism: explanations in cognitive science

University of California, Davis

In the past twenty-five years, many philosophers of science have endorsed a mechanistic approach to explanation, on which more abstract (e.g. functional, mathematical, higher-level) explanations are proposed to be inferior to mechanistic explanations which focus on underlying concrete entities and their causal activities. Regardless of how well mechanistic explanation works in some areas of science, I think we should we wary of assuming that it is the main explanatory tool in cognitive science. In this paper I argue that neither computation nor representation, the two key concepts at the heart of cognitive science, play an explanatory role which can be understood wholly in terms of mechanisms.

Ying-Tung Lin - Self and other in observer memory and imagination

National Yang Ming Chiao Tung University (NYCU)

Episodic memory allows us to mentally travel back to the past, while imagination transports us to hypothetical, future, or fictional scenarios. These activities are often seen as solitary, and their social dimensions have received little attention. However, when these processes unfold from an observer perspective, they open up a space where self and other can intertwine. In this talk, I explore how observer-perspective memory and imagination are closely connected to social cognition. I begin by showing that observer perspectives dissociating the visual point of view from the represented body do not only enable us to view past or imagined events from an external vantage point (Nigro & Neisser, 1983; Lin, 2018; McCarroll, 2018). This shift also allows for various forms of self-experience: we can identify with the observer, the represented body, or both in an observer memory or imagination (Lin & Dranseika, 2021). I then argue that this flexibility has a social dimension. When taking an observer perspective, we can simulate how we might have seen ourselves or how others might have seen us. These forms of self-representation and their social dimensions reveal how memory, imagination, and social cognition may be interconnected. I propose that observer memory and imagination function as a kind of inner social laboratory, in which individuals adopt social viewpoints to better understand themselves and others. They not only transport us through time, but also into social scenes where we meet—and are met by—the selves we once were, the selves we might become, and the selves as seen through the eyes of others.

Michelle Liu - Mental Imagery and harmful language

Monash University

Research on pernicious language tends to focus on harmful beliefs and associations transmitted by such language. In this paper, I explore the idea that pernicious language often transmits harmful mental imagery. Empirical studies suggest that mental imagery is a pervasive feature of language processing. Furthermore, mental imagery prompted by language can influence our memories and judgements in an insidious way. Focusing on language containing misinformation about witnessed events, as well as generics and metaphors about social groups, this paper argues for the importance of mental imagery for theorising harmful language and suggests ways to combat the imagistic harm.

Fiona Macpherson - A Problem For Determining The Structural Features of Experience: A Pessimistic Meta-Induction

University of Glasgow

A dominant thought in consciousness studies is that we should investigate consciousness (either wholly, or at least in part) by studying the structural features of consciousness. Structural features of experience are necessary (or invariant) features of experience. Examples are found in the Kantian claim that all perceptual experiences must be experiences of space and time, and in the claim that experience of red are more similar to experiences of orange than they are to experiences of yellow. In contrast, I argue that we ought to be pessimistic about our ability to determine the structural features of experience. The argument takes the form of a pessimistic meta-induction: many claims as to what the structural features of experience are have turned out to be false. My diagnosis of why people have made claims about what the structure of experience is—claims that have turned out to be false—is that people are consistently fooled by the limitations of their own sensory imagination and they consistently underestimate the range of sensory experiences they—and other people—have not had. Indeed, so vast is the range of possible experience that, we should not be confident that we can ever determine the structural features of experience—if indeed there are any.

Giovanni Rolla - Will artificial systems ever be capable of human-like cognition?

University of Bahia

Will artificial systems ever be capable of cognition? To answer that question, I begin with the distinction between an artificial intelligence as a model for understanding cognition and as a potential display of cognition. While deep neural networks excel at tasks like language processing and pattern recognition, their operations fundamentally differ from human cognition. Unlike humans, who learn through embodied, intersubjective experiences tied to survival and adaptation, artificial intelligences rely on vast datasets and pre-programmed reward functions. Even if they can be useful models, it does not follow that they display cognition. But the possibility that they might eventually be genuinely cognitive is not yet excluded. I then turn to the enactivist and radically embodied framework to raise the embodiment challenge, according to which, due to the lack of biological embodiment and autonomy, artificial systems cannot replicate the self-sustaining, survival-driven processes essential to living systems. I conclude that creating a genuinely cognitive artificial intelligence would require achieving artificial life first, a feat that remains implausible.

Mpho Tshivhase - No Ghost in the Machine: Resisting Digital Appropriation of African Ancestors

University of Pretoria

African cosmologies understand ancestors to hold sacred place in the metaphysical real of being. Ancestors occupy an ontological realm that is significantly different from the physical and the digital realm. These spiritual beings are intricately interconnected with living beings (including the environment) and those who are yet to be born. The AI simulation of ancestors undermines this interconnectivity and the related rituals and customs that connect them with the living and those who are yet to be born. To think that the complex spiritual constitution and transcendent meaning of ancestors can be codified through data and algorithms is reductionist in ways that degrades the sacredness of ancestors from present spiritual beings to a form of digital performance. The impulse to digitize ancestors, reveals a troubling lack of epistemic humility––one that risks positioning those avatars as simulated beings with ancestral authority. Another consideration is the troubling spiritual commodification of lineage and memory which desecrates the spiritual domain as a realm that can be extracted for profit-making. Finally, I will illustrate that the avatarisation of ancestors misses the deeply relational framework of African ethics since the digitization isolates persons from the interactive communal customs and ritual that incorporate ancestors into the meaningful social fabric of ethics. My central argument is that attempts to digitize ancestors presents a category mistake that is grounded in the misapprehension of African cosmology and the accompanying ontological and ethical significance that ancestors hold for African lifeworlds.

Contributed talks

Adriana Alcaraz-Sanchez - Sucked in by the fantasy world: The case of Maladaptive Daydreaming

University of Edinburgh
adriana.alcaraz.sa@gmail.com

We spend roughly half of our waking hours engaged in thoughts unrelated to our immediate task (Killingsworth & Gilbert, 2010). Sometimes, we might become rapt to those thoughts, and lose ourselves in our imagined worlds, yet come back to our usual activities. But other times, we might get stuck in those worlds. This is the case in "maladaptive daydreaming" (MDD; Somer, 2002), a very compelling and extensive sort of fantasising. MDDers report engaging in highly realistic waking fantasies for hours at the expense of their daily responsibilities and interpersonal relationships. Such is the disruption of MDD that some researchers have advocated for its inclusion in the Diagnostic and Statistical Manual for Mental Disorders (DSM-V) considering it a form of "daydreaming disorder" (Somer et al., 2017) or "compulsive fantasising" (Bigelsen & Schupak, 2011). Despite those calls, MDD has not yet been acknowledged by the medical and psychological establishments, preventing MDDers from receiving adequate support and treatment (Bershtling & Somer, 2018). Similarly, other authors have expressed their scepticism about the alleged pathologisation of what is considered a normal mental activity and argue that it can be explained away by the presence of other underlying mental disorders. Yet, descriptions of MDD seem to point to a distinct phenomenon, one that is qualitatively different and not fully captured as an ordinary form of daydreaming. As such, we might wonder, what does the experience of MDD exactly involve? In this presentation, I put forward a descriptive proposal to guide future research on MDD by focusing on the lived experience of MDD. From the research available, I propose that MDD involves a significant shift of attention to the imagined world as well as a strong self-identification with the imagined ego, one that can be said to be reminiscent of dreaming as well as certain dissociative states. In MDD there is an inability to shift one's attention back to the factual. In a way, MDDers are stuck in their fantasies, due to an attentional process that is narrowly focused on their internal worlds, which in turn, prevents them from stepping aside their imaginative involvement.

References

Bershtling, O., & Somer, E. (2018). The Micro-Politics of a New Mental Condition: Legitimization in Maladaptive Daydreamers' Discourse The Micro-Politics of a New Mental Condition: Legitimization in Maladaptive. The Qualitative Report, 23(8), 1983–2002.

Bigelsen, J., & Schupak, C. (2011). Compulsive fantasy: Proposed evidence of an under-reported syndrome through a systematic study of 90 self-identified non-normative fantasizers. Consciousness and Cognition, 20(4), 1634–1648. https://doi.org/10.1016/j.concog.2011.08.013

Killingsworth, M. A., & Gilbert, D. T. (2010). A wandering mind is an unhappy mind. Science, 330(6006), 932. https://doi.org/10.1126/science.1192439

Somer, E. (2002). Maladaptive daydreaming: A qualitative inquiry. Journal of Contemporary Psychotherapy, 32(2), 197–212. https://doi.org/0022-0116/02/0600-0197/0

Somer, E., Soffer-Dudek, N., & Ross, C. A. (2017). The Comorbidity of Daydreaming Disorder (Maladaptive Daydreaming). Journal of Nervous and Mental Disease, 205(7), 525–530. https://doi.org/10.1097/NMD.0000000000000685

Jabran Amanat-Lee - The Little Voice in the Head: Does it Say Anything About Metacognition?

University of Nebraska-Lincoln
jamanat-lee2@huskers.unl.edu

From the first personal point of view, i.e., introspectively speaking, I am experiencing a little voice in my head right now as I am writing this sentence and silently reading it to ensure the absence of errors. There is a high probability that you are having a similar experience right now as you are reading this sentence, and you have an instant introspective awareness of it. Many report experiencing speaking and hearing to themselves silently in their own heads. This phenomenon is often labeled as Inner Speech. There seems to be a rising research interest in understanding the nature of the connection between Inner Speech and Metacognition. There are at least two prominent views on the relation between Inner Speech and Metacognition. A prominent view is that Inner Speech produces and sustains Metacognition—call this the Constitution View. Another prominent view is that Inner Speech facilitates Metacognition by bringing thoughts to consciousness. That is, Inner Speech is the vehicle through which we become more cognizant of our thoughts—call this the Catalyst View. However, there is a more general question of concern in the background: Is there a connection between Inner Speech and Metacognition? The general assumption in the literature seems to be that there is a connection between Inner Speech and Metacognition. This assumption seems to have gone largely unchallenged. This paper raises a number of challenges to this general assumption. Evidence that is often cited in support of the general assumption can be plausibly categorized as coming from the following two sources: Neural imaging data and introspective reports. The latter set of evidence is conceived to support the thought that we have introspective reasons to think that we are clearly aware of our thoughts as objects of our metacognitive reflection, and they occur as Inner Speech episodes— let us call this the Introspection Argument. The former set of evidence is regarded to support the thought that neuroimaging data shows that there is a relationship between Inner Speech and Metacognition—let us call this the Neural Correlates Argument. These two arguments appear to be the footing for the general assumption that there is a connection between Inner Speech and Metacognition. The evidence that is often cited in support of these two arguments will be the focus of evaluation. I argue that both of these arguments fail to substantiate the general assumption that there is a connection between Inner Speech and Metacognition. The claim is not that there is no connection; the primary claim is that the Neural Correlates Argument and the Introspection Argument fail to substantiate that there is a connection between Inner Speech and Metacognition.

Tomy Ames - Illusory Evidence, or How Perceptual Illusions Justify Beliefs

Washington University in St. Louis
tames@wustl.edu

I argue that perceptual illusions illustrate the nascent epistemological view that we can gain knowledge from non-knowledge. Given that perceptual illusions are non-veridical, this spells obvious epistemological problems, particularly for alethic virtue-reliablist accounts of knowledge. Some kinds of perceptual illusions such as peripheral drift, simultaneous contrast, Ponzo, and Ebbinghaus illusions, however, can be paradigmatic examples of non-factive knowledge under Nolfi's action- oriented virtue-theoretic account. Importantly, this account maintains that epistemic success is determined by the degree to which a doxastic state can be used as a tool by the agent to effect some action. Within this framework, some perceptual illusions, despite being non-factive, provide evidence and reasons that subserve beliefs which are used as predictive tools that allow the subject to navigate and interact with normal environments and circumstances. Such perceptual illusions are thus not errors, but are a normative function of perceptual systems that are well-suited for generating knowledge toward situationally appropriate agential action. I argue that perceptual illusions, within the action-oriented virtue-theoretic account, provide an extraordinary empirical example of one way that what we perceive can be, by design, non-factive and yet still lead to knowledge that undergirds our actions.

Laida Arbizu Aguirre - The Mind(s) We Deny: Epistemic Denial and the Marginalization of Cognitive Diversity

University of the Basque Country
laida.arbizu@ehu.eus

Who defines what constitutes a mind? Dominant paradigms across cognitive science, AI research, and neuroscience recognize certain cognitive forms while systematically excluding others. This exclusion reflects not merely empirical oversight but a deeper epistemic denialism: the refusal to acknowledge cognitive diversity. By privileging rationalist, computational, and disembodied models, mainstream frameworks enforce exclusionary epistemologies that equate intelligence with historically androcentric and Eurocentric ideals of reason. This denialism operates through both overt rejection and epistemic injustice (Fricker, 2007), rendering non-human, collective, and marginalized human intelligences illegible. Drawing on feminist critiques (Haslanger, 2012; Scheman, 2011), this work challenges essentialist models that prioritize formal logic over embodied, affective, and relational ways of knowing. Testimonial injustice—where cognitive systems deviating from rationalist norms are discredited—silences marginalized groups and non-human agents. Such injustices extend beyond marginalization to actively deny the existence of alternative cognitive forms, reinforcing a narrow ontology of intelligence. For instance, Haraway's (1988) critique of the "god trick" reveals how scientific objectivity masks exclusionary practices, while posthumanists like Hayles (1999) and Braidotti (2013) advocate for ecological frameworks that recognize distributed cognition across human, non-human, and artificial agents. Current epistemic architectures not only exclude but shape the tools used to study minds. To counter this, the work proposes a pluralist epistemology embracing distributed (Hutchins, 1995), affective (Colombetti, 2014), and ecological (Clark, 1997) intelligences. This shift requires abandoning rule-based models in favor of frameworks that value relationality and embodiment. For example, collective intelligences—from swarm behavior to AI-human collaborations—challenge individualist notions of cognition, while affective reasoning disrupts hierarchies that privilege abstraction over emotion. By interrogating exclusionary practices, this analysis contributes to philosophy of mind, cognitive science, and AI ethics. It asks: What minds remain unseen due to epistemic denialism? How might recognizing diverse intelligences transform research and technology design? The work argues that overcoming denialism demands more than inclusivity—it necessitates reimagining cognition itself. This entails dismantling the illusion of neutrality in scientific methods and centering marginalized perspectives in epistemic practices. A radical epistemic pluralism contests not just the definitions of intelligence but also the approaches employed to investigate it. By combining feminist, posthumanist, and ecological viewpoints, this framework enhances our comprehension of cognition, providing tools to acknowledge the minds that are presently overlooked by prevailing paradigms. This change holds the potential to democratize the creation of knowledge and encourage technologies that embody the entire range of cognitive diversity.

References

Braidotti, R. (2013). The Posthuman. Polity Press.

Clark, A. (1997). Being There: Putting Brain, Body, and World Together Again. MIT Press.

Colombetti, G. (2014). The Feeling Body: Affective Science Meets the Enactive Mind. MIT Press.

Fricker, M. (2007). Epistemic Injustice: Power and the Ethics of Knowing. Oxford University Press.

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. Feminist Studies, 14(3), 575–599.

Haslanger, S. (2012). Resisting Reality: Social Construction and Social Critique. Oxford University Press.

Hayles, N. K. (1999). How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics. University of Chicago Press.

Hutchins, E. (1995). Cognition in the Wild. MIT Press.

Scheman, N. (2011). Shifting Ground: Knowledge and Reality, Transgression and Trustworthiness. Oxford University Press.

Ritesh Bansal - Teasing Out the Nature of the Design and Application of GPS from Nudging in the Context of Ethics of Nudging

Indian Institute of Technology Jammu
ritesh.bansal@iitjammu.ac.

In the literature of nudging, Thaler and Sunstein, major advocates of nudging, state that GPS is a prime example of nudging (Sunstein, 2014a; 2014b; 2017; 2019) (Sunstein, Reisch, & Kaiser, 2019) (Thaler, 2018). Sunstein (2015b) claims that GPS is a prime nudge because GPS identifies the best route for people and steers them in the best direction. People have the liberty to drive on the route suggested by GPS or they can choose their own route (Sunstein, 2020a). In this way,GPS preserves freedom of choice and makes people better off as judged by themselves (Sunstein, 2018). The concept "better off" is typically interpreted as welfare within the framework of nudging. Sunstein often claims that nudges like default rules and disclosures operate like a GPS device (Sunstein, 2020a). Consequently, by referring GPS as a prime example of nudge, Sunstein has also made attempts to counterargue against several ethical concerns raised against nudging in general. This paper challenges the view that GPS qualifies as a nudge. There are significant differences between nudges and GPS in terms of their designs and the nature of the benefits they provide. In the context of navigation, the primary purpose of GPS is to provide users with efficient and accurate route guidance and enhance user experience based on explicit user input. GPS doesn't intend to influence broader behavior and decision making of the user. In contrast, in the context of nudging, choice architects deliberately design nudges to influence agents' behavior toward specific choices as judged by themselves. Choice architects often rely on insights from behavioral economics to address cognitive biases and improve agent's decision-making. Unlike navigation by GPS, which focuses on usability, nudges aim at the welfare of the stakeholders through subtle behavioral influence, which may introduce certain ethical considerations related to autonomy, transparency, manipulation, and paternalism. In the literature, GPS is often cited to justify the ethical concerns of nudging interventions. The arguments suggest that since GPS technology is widely accepted as ethically uncontroversial, similar acceptance should be extended to other forms of nudging as well. Sunstein states; "As the GPS example suggests, many nudges have the goal of increasing navigability ... Insofar as the goal is to promote navigability, the ethical objections are greatly weakened and might well dissipate" (Sunstein, 2015a, p. 426); "A GPS insults no one's dignity". (Sunstein, 2015a, p. 441); A GPS does not undermine human agency;" (Sunstein, 2015b, p. 512). However, this comparison overlooks critical distinctions in the objectives and outcomes of nudging and GPS. By distinguishing the design and purposes behind the design of GPS and nudging, this paper highlights the need for a productive and critical discourse on the ethical challenges associated with nudges.

References

Sunstein, C. R. (2014a). Choosing Not To Choose. Duke Law Journal, 1-52.

Sunstein, C. R. (2014b). Nudging: A Very Short Guide. Journal of Consumer Policy, 583-588. doi:https://doi.org/10.1007/s10603-014-9273-1

Sunstein, C. R. (2015a). The ethics of nudging. Yale Journal on Regulation, 413-450.

Sunstein, C. R. (2015b). Nudges, Agency, and Abstraction: A Reply to Critics. Review of Philosophy and Psychology, 511-529. doi:https://doi.org/10.1007/s13164-015-0266-z

Sunstein, C. R. (2017). Nudges that fail. Behavioral Public Policy, 4-25. doi:https://doi.org/10.1017/bpp.2016.3

Sunstein, C. R. (2018). "Better off, as judged by themselves": a comment on evaluating nudges. International Review of Economics, 1-8. doi:https://doi.org/10.1007/s12232-017-0280-9

Sunstein, C. R. (2019). Which nudges do people like? A national survey. In Handbook of Behavioural change and public policy (pp. 285-303). Edward Elgar Publishing. doi:https://doi.org/10.4337/9781785367854.00028

Sunstein, C. R. (2020a). Behavioral science and public policy. Cambridge University Press. doi:https://doi.org/10.1017/9781108973144

Sunstein, C. R., Reisch, L. A., & Kaiser, M. (2019). Trusting nudges? Lessons from an international Survey. Journal of European Public Policy, 1417-1443. doi:https://doi.org/10.1080/13501763.2018.1531912

Thaler, R. H. (2018). Nudge, not sludge. Science, 361(6401), 431. doi:https://doi.org/10.1126/science.aau9241

David Barack - Neuron doctrines

Lingnan University
dbarack@gmail.com

The neuron doctrine has long guided neuroscientific research into the brain's functioning, both cognitive and otherwise. However, the neuron doctrine suggests that the only cognitively relevant findings in the brain are from single neurons. This is far from accurate: there are decades of findings about the brain basis of cognition that do not involve neurons. Neural phenomena associated with cognition span a wide range of spatiotemporal scales and types of entities. The challenge is to formulate a thesis that can describe the brain basis of thought while accommodating the wide range of cognitively relevant findings. This talk has three aims. First, I aim to state the neuron doctrine and what would count as evidence against it. Stating the neuron doctrine in a fair and comprehensive fashion turns out to be challenging because it is a generic, like 'tigers have stripes' or 'birds fly'. I will provide a fair and comprehensive statement of the neuron doctrine as a generic. Second, given the range of neural activity intuitively speaking against it, I want to explain the surprising resilience of the neuron doctrine. Finally third, I wish to motivate an alternative to the neuron doctrine by presenting a range of cases that imply that the relative functional contribution of neurons to cognition is less than expected once understood as a generic. I define relative functional contribution and relate it to the neuron doctrine as a generic. I first distinguish three different neural doctrines, anatomical, sensorimotor, and cognitive. I state the cognitive neuron doctrine, which consists of two theses, one regarding the unit in the brain that functionally contributes to cognition and the other the signal, briefly discuss how generic theses should be understood, and relate my statement to past statements of the neuron doctrine. I next discuss what it would take for some findings to count as evidence against the neuron doctrine. Stating principles sufficient for countervailing evidence turns out to be no easy task and goes part of the way toward explaining the resilience of the neuron doctrine. I also situate the cognitive neuron doctrine in the context of the success of the anatomical and sensorimotor doctrines. Framed in terms of establishing a Bayesian prior on the neuron as the relevant unit and the spike the relevant signal, the focus on neurons is no longer surprising. I then canvas a wide range of evidence drawn from across the neurosciences regarding the brain basis of cognition. For each case study, I evaluate whether it counts for or against the cognitive neuron doctrine and I consider what revisions, if any, are required to accommodate the evidence. I end by describing findings that I take to challenge the cognitive neuron doctrine and recommend a search for an empirically better approach to understanding the mind.

Carlos Barth - Relevance sensitivity as a cognitive gadget

Faculdade Jesuíta de Filosofia e Teologia (FAJE)
carlos@cbarth.me

Sensitivity to what is circumstantially relevant across an open-ended set of con- texts — a capacity sometimes termed "relevance realization" (Vervaeke, Lilli- crap, and Richards 2012) — is central to human cognition. Even picking a beverage at a restaurant involves sensitivity to indefinitely many elements that are neither about restaurants nor beverages (are my colleagues around?). This capacity challenges the cognitive sciences to provide a naturalistic and non-circular account of how human cognition is able to navigate the seeming intractability posed by it. Its history goes back the frame problem of AI. At first, it was presented as the difficulty to model complex dynamic domains (McCarthy and Hayes 1969; Janlert 1987), but it was quickly reinterpreted as an issue regarding relevance (Dennett 1987). Even though it was originally identified within the classical cognitivist paradigm (Fodor 1987) and largely discussed using a representational and computational vocabulary, it is neither representational, nor computational in nature. The issue can be formulated, for instance, in terms of selecting the relevant set of affordances to attend to at any given context (Bruineberg and Rietveld 2014). Likewise, one can give it a predictive-processing (Clark 2016) gloss by formulating it as the need to select the contextually adequate set of priors. Thus, no contemporary framework is free from it. Theories associated with em- bodied, embedded, enacted, extended, and affective cognition (4EA cognition) are all equally subject to the same challenge. Although bodily and environmental scaffolds may shape and constrain cognitive processes, they do not inherently indicate what is circumstantially relevant. Similarly, emotions and affective states fall short of independently resolving the fundamental issue (Barth 2024). The issue has a remarkable resilience. Attempts to simplify or decompose relevance sensitivity into smaller, more manageable problems have not succeeded (Rietveld 2012). Moreover, solution attempts appealed to mechanisms or ele-ments that presupposed this capacity rather than explaining it. A good exam- ple is the attempt to explain it through learning. Acquiring relevance sensitivity seems to require either 1) an all-encompassing and domain-independent learning strategy, which is demonstrably implausible (Wolpert and Macready 1997; Sterkenburg and Grünwald 2021), or 2) indefinitely many permutations of domain-specific or model-specific learning biases, which begs the question of how to find the articulation that is appropriate to the domain and context in question. This work draws on the cognitive gadgets framework proposed by Heyes (2018), and suggests that we should conceptualize relevance sensitivity as a culturally evolved cognitive tool. In this picture, culturally transmitted developmental tweaks and biases actively shape cognitive processes, aligning them with cultur- ally established understandings of contextual relevance. Such cognitive gadgets do not merely exploit cultural information but are partially constituted by culturally transmitted learning and developmental biases and processes. I'll argue that this perspective affords a plausible and non-circular explanation of how relevance sensitivity is acquired and exercised. Moreover, the hypothesis is reasonably

framework-neutral, and it is potentially useful for both cognitivists and ecological psychologists or enactivists.

References

Barth, Carlos. 2024. "Is There a Place for Emotions in Solutions to the Frame Problem?" Síntese: Revista de Filosofia 51 (161): 527. https://doi.org/10 .20911/21769389v51n161p527/2024.

Bruineberg, Jelle, and Erik Rietveld. 2014. "Self-Organization, Free Energy Minimization, and Optimal Grip on a Field of Affordances." Frontiers in Human Neuroscience 8 (August). https://doi.org/10.3389/fnhum.2014.005 99.

Clark, Andy. 2016. Surfing Uncertainty. Oxford: Oxford University Press.

Dennett, Daniel. 1987. "Cognitive Wheels: The Frame Problem of AI." In The Robot's Dillemma: The Frame Problem in Artificial Intelligence, edited by Zenon W. Pylyshyn, 41–64. Ablex.

Fodor, Jerry A. 1987. "Modules, Frames, Fridgeons, Sleeping Dogs and the Music of the Spheres." In The Robot's Dillemma: The Frame Problem in Artificial Intelligence, edited by Zenon W. Pylyshyn, 139–49. Ablex.

Heyes, Cecilia. 2018. Cognitive Gadgets. Harvard University Press.

Janlert, Lars-Erik. 1987. "Modeling Change - the Frame Problem." In The Robot's Dillemma: The Frame Problem in Artificial Intelligence, edited by Zenon W. Pylyshyn, 1–40. Ablex.

McCarthy, John, and Patrick J. Hayes. 1969. "Some Philosophical Problems from the Standpoint of Artificial Intelligence." Machine Intelligence 4: 463– 502. https://doi.org/10.1016/b978-0-934613-03-3.50033-7.

Rietveld, Erik. 2012. "Context-Switching and Responsiveness to Real Rele- vance." In Heidegger and Cognitive Science, edited by Julian Kiverstein and Michael Wheeler, 105–35. Hampshire: Palgrave Macmillan. https: //doi.org/10.1007/978-1-137-00610-3/_3.

Sterkenburg, Tom F., and Peter D. Grünwald. 2021. "The No-Free-Lunch Theorems of Supervised Learning." Synthese 199 (3-4): 9979–10015. https://doi.org/10.1007/s11229-021-03233-1.

Vervaeke, J., T. P. Lillicrap, and B. A. Richards. 2012. "Relevance Realization and the Emerging Framework in Cognitive Science." Journal of Logic and Computation 22 (1): 79–99. https://doi.org/10.1093/logcom/exp067.

Wolpert, D. H., and W. G. Macready. 1997. "No Free Lunch Theorems for Optimization." IEEE Transactions on Evolutionary Computation 1 (1): 67–82. https://doi.org/10.1109/4235.585893.

Sacha Behrend - The three faces of mental imagery: Towards a New Definition of Mental Imagery

Université Paris 1, Panthéon-Sorbonne
sachabehrend1991@gmail.com

This paper addresses the surprisingly difficult task of defining mental imagery. It argues against recent influential definitions, those by Bence Nanay (2023) and Margherita Arcangeli (2020), and proposes an alternative definition that avoids their limitations. Bence Nanay (2023) defines mental imagery as perceptual processing that is not triggered directly by sensory input. However, this definition faces several challenges. First, it is either too permissive or too restrictive (Green & Lande, 2024). Second, it fails to delineate a coherent natural kind, as the notion of indirect triggering applies to a heterogeneous set of phenomena that lack a unifying explanatory property (Green & Lande, 2024). Finally, it leads to the counterintuitive classification of amodal completion as a form of mental imagery. On the other hand, Margherita Arcangeli (2019) distinguishes two senses of mental imagery: as a psychological attitude and as a type of mental content. In the first sense, mental imagery refers to a re-creation or simulation of perception, producing mental experiences that are phenomenologically and/or functionally similar to perceptual ones. In its second sense, mental imagery refers to a specific type of content rather than a mode of mental processing. She proposes to restrict mental imagery to this second sense. Despite these insights, Arcangeli's account has several shortcomings. First, I argue that she misconstrues certain subpersonal processes as personal-level phenomena. In particular, the first sense of mental imagery is better understood as a subpersonal process rather than a conscious psychological attitude. Additionally, Arcangeli's account does not provide a definition that would allow psychologists and neuroscientists to apply her framework in empirical research. To address these limitations, I propose a new definitional approach to mental imagery. Building on Arcangeli's approach, I first examine the varied uses of "mental imagery" in both philosophical and scientific literature. These diverse uses reveal conceptual ambiguity, requiring a clearer theoretical framework. I argue that they can be reduced to three core referents necessary for crafting a precise definition:
(a) Phenomenal experience,
(b) A representational format, and
(c) A mental/neural process.
Additionally, I challenge the assumption that phenomenal experience is either necessary or sufficient for defining mental imagery, although it often accompanies it. Based on this analysis, I propose the following definition: Mental imagery is (i) a top-down, quasi-perceptual process (ii) producing analog/depictive mental representations. This definition captures both the processual and representational aspects of mental imagery, ensuring it remains distinct from perception. By refining the concept in this way, this paper contributes to greater clarity in the theoretical and empirical literature on mental imagery, imagination and perception.

Luuk Brouns - Folk Psychology in Cross-Cultural Context: A Non-Universalist Account Based on Interactive-Mindshaping Practices

Radboud University
L.brouns@hotmail.com

This paper argues that folk psychology is not a universal occurrence but a culture-specific adaptation of WEIRD societies (societies categorised as Western, Educated, Industrialised, Rich, and Democratic). This opposes the standard mindreading approach. The convention holds that (a) folk psychology results from a universal cognitive capacity called "mindreading," and (b) the resulting core folk psychology is not affected by cultural variation (Apperly & Butterβill, 2009; Carruthers, 1996; Fodor, 1998; Gopnik & Wellman, 1998; Nichols & Stich, 2003; Scholl & Leslie, 1999). The non-universalist approach I defend in this paper is based on the integration of mindshaping (McGeer, 2007, 2021; Zawidzki, 2013) and interactionism (Gallagher, 2004; Gallagher & Hutto, 2008). The integrated approach holds that (i) social practices such as (over)imitation, norm enforcement/following, pedagogy and narrative practices, shape the minds of individuals such that they can regulate their behaviour in accordance with a variety of culture-specific norms, and (ii) these cultural practices vary so significantly that analysing them in terms of belief and desires is unwarranted and insufβicient. Some cultures, like WEIRD societies, have developed narrative practices that include folk psychological narratives, involving conceptual tools such as beliefs and desires (Hutto, 2008). These are useful within their cultural context. However, the actual analysis of a culture that differs significantly from WEIRD cultures, that of Samoa, shows that not all societies require folk psychology. On my account, Samoa exemplifies how a well-functioning society can thrive without engaging in folk psychological practices. I thus conclude that, when actual social practices are considered, it turns out that folk psychology is not a universal occurrence.

References

Apperly, I. A., & Butterβill, S. A. (2009). Do Humans have two Systems to Track Beliefs and Belief-like States? Psychological Review, 116(4), 953–970. https://doi.org/10.1037/a0016923

Carruthers, P. (1996). Simulation and Self-knowledge: A Defense of Theory Theory. In Theories of Theories of Mind (pp. 22–38). Cambridge University Press.

Fodor, J. A. (1998). Psychosemantics: The Problem of Meaning in the Philosophy of Mind (4th ed.). MIT Press.

Gallagher, S. (2004). Understanding Interpersonal Problems in Autism: Interaction Theory as An Alternative to Theory of Mind. Philosophy, Psychiatry, & Psychology, 11(3), 199–217. https://doi.org/10.1353/ppp.2004.0063

Gallagher, S., & Hutto, D. D. (2008). Understanding Others Through Primary Interaction and Narrative Practice. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), Converging Evidence in Language and Communication Research (Vol. 12, pp. 17–38). John Benjamins Publishing Company. https://doi.org/10.1075/celcr.12.04gal

Gopnik, A., & Wellman, H. M. (1998). The Theory Theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), Mapping the Mind: Domain SpeciŸicity in Cognition and Culture (Reprint, pp. 257–293). Cambridge University Press.

Hutto, D. D. (2008). Folk psychological narratives: The sociocultural basis of understanding reasons. MIT Press.

McGeer, V. (2007). The Regulative Dimension of Folk Psychology. In D. D. Hutto & M. Ratcliffe (Eds.), Folk Psychology Re-Assessed (pp. 137–156). Springer Netherlands. https://doi.org/10.1007/978-1- 4020-5558-4_8

McGeer, V. (2021). Enculturating Folk Psychologists. Synthese, 199(1–2), 1039–1063. https://doi.org/10.1007/s11229-020-02760-7

Nichols, S., & Stich, S. P. (2003). Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds (1st ed.). Oxford University Press. https://doi.org/10.1093/0198236107.001.0001

Scholl, B. J., & Leslie, A. M. (1999). Modularity, Development and 'Theory of Mind.' Mind & Language, 14(1), 131–153. https://doi.org/10.1111/1468-0017.00106

Zawidzki, T. W. (2013). Mindshaping: A new Framework for Understanding Human Social Cognition. MIT Press.

José Carlos Camillo - An explanatory model for observer memories

Université de Genève
josecarloscamillo@gmail.com

There are several models that aim to explain observer memories—memories in which a person recalls a past event from a third-person, "visual" perspective. Most of these models focus on the role of emotional and self-related content (emotional-self models). However, alternative models explain observer memories based on the type of information available or requested during retrieval. These are known as the informational-availability model and the informational-request model, respectively. This talk will integrate these latter models with current research on the neural representations involved in encoding, consolidation, and retrieval, offering a more comprehensive explanation of observer memories. It will begin by outlining how to assess the explanatory power and completeness of models, followed by a review of the main models discussed in the scientific literature. I will argue that empirical evidence supports both the informational-availability and informational-request models. Although often presented as competing views, I will propose that they are complementary and together form an informational-request-and-availability model. To complete this integrated model, I will examine the role of neural representations in memory processes—an aspect largely overlooked in studies of observer memories. Neuroimaging and optogenetic research show that during perceptual experiences, multiple types of representations are formed, ranging from abstract to concrete/perceptual. These include egocentric and allocentric spatial representations, as well as perspective-laden and perspective-neutral object representations, each encoded in distinct brain regions. Although these representations undergo some transformation during consolidation, their core content and type typically remain stable. At retrieval, abstract and concrete/perceptual representations interact to meet the demands of the retrieval request. Abstract representations are more closely linked to allocentric spatial representations and perspective-neutral object representations, while concrete/perceptual representations are more connected to egocentric spatial representations and perspective-laden object representations. Thus, when retrieval calls for abstract information, memories tend to adopt an observer perspective. Conversely, when concrete/perceptual information is needed, memories are more likely to be recalled from a first-person, own-eyes perspective. Importantly, when concrete/perceptual representations are unavailable—such as in remote memories, in older adults' memories, or in early stages of Alzheimer's Disease—an observer perspective often emerges. Notably, these groups show reduced access to concrete/perceptual representations even in non-mnemonic tasks. This integrated model has important implications for debates on the accuracy of remembering. Scholars such as Georgia Nigro, Ulric Neisser, and Christopher Jude McCarroll argue that observer memories can be accurate because they include impressions formed by thoughts, imaginings, and emotions, thus reflecting what was consciously experienced. By contrast, the neuro-representational model I propose suggests that memory reconstructs information originally processed during perception—whether consciously experienced or not. This view maintains that observer

memories are not necessarily distortions but reconstructions based on real, albeit sometimes unconscious, representations formed during perceptual states.

Berke Can - External cognitive tool use: Modelling uniquely human cognitive development

University of Warwick
berke.can@warwick.ac.uk

Adult humans can perform cognitive feats like counting or complex toolmaking that no other species can. How are such capacities acquired? While it is common to approach this question by asking whether a given capacity is innate or learned, we can gain better insights by focusing on the structure of the developmental process involved. I will argue that complex capacities often only performed by humans develop through a process that I will call external cognitive tool use. This is a developmental process where the acquisition of a capacity requires the acquisition of at least one prerequisite capacity that belongs to a distinct context from the one that it will be applied to, this distinct capacity serving as the external tool. This is similar to the biological process of endosymbiosis, where two cells with distinct evolutionary histories merge to form a more complexly organized cell (Mast et al., 2014). Something similar can happen in cognitive development, for instance in counting or complex toolmaking. Learning to count involves combining two capacities with distinct developmental histories: going through the counting procedure (i.e., saying "one", "two", "three" while pointing) and determining the quantity of a given set (Carey, 2009). This combination yields a complex capacity where the counting procedure serves as an external tool, enabling the exact determination of large quantities which wasn't possible before. The same developmental structure can be observed in complex toolmaking, like stone hand axes made by earlier hominins (Birch, 2021). The tool needs to be manufactured in one context, where materials are collected and shaped, and applied to another, like butchery. The manufacture of a stone tool serves as an external cognitive tool in butchery, resulting in a new and complexly structured capacity that couldn't have been acquired within the constraints of a single context. I argue that such developmental structures are possible not because our ontogeny has complex starting points nor because we experienced a sudden jump in our cognitive capacities in our evolutionary history, but rather because of our sensitivity to social contexts. We are uniquely sensitive to feedback from others, as subtle as gaze directions or facial expressions. These sensitivities create contexts where external tools can be acquired in ontogeny and develop through cultural evolution, such as counting systems or toolmaking techniques (Everett, 2017; Henrich, 2016). For instance, an agent who may not understand that toolmaking will lead to greater butchery capacities may still acquire it because making the tool elicits rewarding social feedback from peers. This developmentalist perspective alleviates the need to posit complex starting points in ontogeny or mysterious cognitive jumps in phylogeny, enabling a gradualist account of the emergence of complex cognitive capacities in both time scales. These capacities are made possible by the structure of the developmental process underlying them, which is in turn made possible by uniquely human forms of sociality.

References

Birch, J. (2021). The skillful origins of human normative cognition. Analyse & Kritik, 43(1), 191–202. https://doi.org/10.1515/auk-2021-0010

Carey, S. (2009). The Origin of Concepts. Oxford University Press.

Everett, C. (2017). Numbers and the making of us: Counting and the course of human cultures. Harvard University Press. Henrich, J. P. (2016). The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter. Princeton University Press.

Mast, F. D., Barlow, L. D., Rachubinski, R. A., & Dacks, J. B. (2014). Evolutionary mechanisms for establishing eukaryotic cellular complexity. Trends in Cell Biology, 24(7), 435–442. https://doi.org/10.1016/j.tcb.2014.02.003

Marta Caravà - On the rationality of holding onto one's 'past' emotions

Purdue University
mcarav@purdue.edu; marta.carava@gmail.com

When we remember past experiences, we often feel that we are re-living them, right here, right now (Tulving, 1983). The emotions we feel in this process play a great role in building up the sense of re- reliving the past that characterizes episodic remembering (Perrin, Michaelian, & Sant'Anna, 2020). Based on these considerations, one might be tempted to claim that the emotions we feel in remembering are themselves memories of past emotions. Like others have done before (Debus, 2007), I will argue that this claim is mistaken. While we may feel that we are remembering past emotions, the affective states that accompany episodic remembering are, in fact, present emotions. This does not mean, however, that 'holding onto these emotions' is an irrational thing to do. Quite the contrary. These emotions are, indeed, epistemically significant: they are central for gaining new knowledge in many ways. Or so I argue. My core claim is that the emotions we feel in episodic remembering often trigger inferential processes of memory search that are beneficial for retroactive learning. They allow us to look back at our personal past from a different perspective, making particular details of past experienced events salient and available for further inspection from a privileged epistemic point of view - from the point of view of our present, more experienced, self. This emotion-based process of inferential memory search has benefits for our knowledge of the world: it allows us to learn something new about past events and facts (Boyle, 2019). Its epistemic power does not stop here though. Indeed, this process also contributes to increasing our self-knowledge: it allows us to get a better sense of who we were in the past and of who we are right now. Thus, in some sense, it consolidates, deepens, and complexifies our personal identity (Schechtman, 2024). These improved self-knowledge and increased awareness of our personal identity have benefits for other, non-strictly, cognitive domains. For example, they allow us to engage in self-directed moral processes such as self-forgiveness (Caravà & McCarroll, 2025), they afford adaptive forms of emotion regulation (Nørby, 2018), and they make us better at social interaction, thus enhancing the quality of our social relationships (Tenney et al., 2013). Considering all these benefits, it should be clear why, even if the emotions we feel in episodic remembering are not themselves memories of past emotions, 'holding onto such emotions' and using them as tools for different types of learning and self-discovery is a rational thing to do. At least if we understand 'rationality' in a prudential sense.

References

Boyle, A. (2019). Learning from the past: Epistemic generativity and the function of episodic memory. Journal of Consciousness Studies, 26 (5-6): 242-251.

Caravà, M. & McCarroll, C. J. (2025). Forgiving unbound: Emotion, memory, and materiality in extended moral processes. Synthese, 205 (4):1-26.

Debus, D. (2007). Being emotional about the past: On the nature and role of past-directed emotions. Noûs, 41(4):758-779.

Nørby, S. (2018). Mnemonic emotion regulation: A three-process model. Cognition and Emotion, 33(5):959-975.

Perrin, D., Michaelian, K., & Sant'Anna, A. (2020). The phenomenology of remembering is an epistemic feeling. Frontiers in psychology, 11, 1531. https://doi.org/10.3389/fpsyg.2020.01531 https://doi.org/10.1371/journal.pone.0069605

Schechtman, M. (2024). Personal identity and mental time travel. Philosophy and the Mind Sciences, 5. https://doi.org/10.33735/phimisci.2024.10639

Tenney, E. R., Vazire, S., & Mehl, M. R. (2013). This examined life: The upside of self-knowledge for interpersonal relationships. PloS One, 8(7), e69605.

Tulving, E . (1983). Elements of episodic memory. Oxford, UK: Clarendon Press.

Mark-Oliver Casper - Modeling Cognition's Complexity: A Dual Approach to the Motley Crew Argument

University of Kassel & University Göttingen
moc@uni-kassel.de

A basic enactivist claim states that a circuitry of action, environmental structures, and perception is constitutive for cognitive phenomena. However, by maintaining that ever more variables need to be considered in cognition research, enactivism is vulnerable to criticisms that are methodological in nature. Critics fairly maintain that it is crucial to rely on methods properly suited to capture the wild patchwork of things and processes enactivists aim to focus on. This critique is known as "the motley crew argument". Other scientific disciplines, such as behavioral biology, investigate the intricate network of variables underlying behavior. Behavioral biology in particular developed sophisticated strategies for handling the inherent complexity of behavioral profiles. This talk introduces the key conceptual commonalities between both research accounts, proposing to integrate enactivism with the empirical rigor of behavioral biology. The commonalities are illustrated through compelling ethological case studies, e.g., the change of risk evaluation in gregarious birds due to swarm size and lived habitat. Such case studies show how relevant variables for cognitive phenomena are identified and related. The development of such networks of variables is an important step to mathematically analyze the strength of their connection and the dynamics between them. In a further step, this talk introduces structural equation modeling (SEM) as a potent analytical tool for modeling the multifaceted relationships between biological, behavioral, and environmental factors. SEM is a statistical technique that allows researchers to examine complex relationships among observed and unobserved (latent) variables. SEM enables the simultaneous estimation of multiple interrelated dependencies, making it ideal for evaluating causal hypotheses and theoretical constructs, while being an enriching alternative to differential equations used in the context of dynamical system theory. SEM can therefore bridge the gap between enactivist theory and empirical data, providing a data-driven method for testing enactivist-behavioral hypotheses. The dual connection, enactivism with behavioral biology and enactivism with SEM, is a dual approach to confront the motley crew argument and to advance not just the enactivist but the 4E study of cognition.

Cong Chen - Trivialisms about explanatory gap

Zhejiang University
congchen_phi@zju.edu.cn

A spectre, a spectre of anti-physicalism is haunting the realm of philosophy of mind. We seem to be forever unable to explain how phenomenal experiences (such as the vanilla scent of ice cream or the pain of a breakup in an intimate relationship) suddenly emerge from the physical, functional, or even so-called protophenomenal and whatnot. No matter what advanced theoretical tools we use to explain it, there is always a residual mystery—an abyss staring back at us, reminding us of an explanatory gap. In this paper, I do not want to repeat clichés about how profound the gap is, but rather attempt to defend a position I call Trivialism, or more accurately, a set of various Trivialisms, which suggests that perhaps the explanatory gap for the phenomenal is not as crucial as we thought. Please note that the various forms of Trivialism I discuss do not all claim that the explanatory gap for the phenomenal is meaningless; rather, most of them suggest that in our attempts to explore the nature of reality, we encounter various explanatory gaps across different fields, and the explanatory gap for the phenomenal is at least no more special than other explanatory gaps. In section 1, I will survey Schaffer's original version of what I also take to be the most modest form of Trivialism, which I characterize as consisting of anti-exceptionalism and the thesis of non-specialness. The former claims that explanatory gaps are everywhere—this is a descriptive thesis that attempts to defend the idea that explanatory gaps arise whenever there is a connection between the more fundamental and the less fundamental. The latter is a normative thesis, arguing that there is nothing uniquely special about the explanatory gap for the phenomenal, or at least that it is no more special than other explanatory gaps. I will then briefly outline the implications of Schaffer's version of Trivialism (which apply across all versions of Trivialism). Further, I will present a strong objection— perhaps the best currently available in the literature—Aleksiev's critique of the thesis of non- specialness, and attempt to respond to it in defense of Schaffer's position. In section 2, I introduce and defend what I labels Radical Trivialism, suggesting that while explanatory gaps are everywhere, there are fewer epistemic reasons (that is, epistemic insensitivity) to posit an explanatory gap specifically concerning the phenomenal (a position I call exceptionalism), and that the so-called explanatory gap for the phenomenal is not as special as other kinds of explanatory gaps (a more radical version of the thesis of non-specialness). I will also argue that this version can, in some respects, preempt certain objections that have been raised in the literature; thus, even though it appears more radical than Schaffer's version, it may actually be easier to defend. In section 3, I will respond to several potential objections, including objection concerning the very concept of epistemic sensitivity, challenge from metaontological deflationism, and objection that challenging epistemic insensitivity can be considered a genuine defeater by independent third factors, among others.

References

Aleksiev, Damian (2022). Lightweight and Heavyweight Anti-physicalism. Synthese 200 (112):1-23.

Bennett, Karen (2021). Why I am not a dualist. Oxford Studies in Philosophy of Mind 1:208-231.

Schaffer, Jonathan (2017). The Ground Between the Gaps. Philosophers' Imprint 17.

Alice Andrea Chinaia - Are There Norms of Rationality for Perception? Evaluating Experience-forming Processes under Bayesian Norms

IMT School for Advanced Studies Lucca
alice.chinaia@imtlucca.it

According to Susanna Siegel's Rationality of Perception, perceptual experiences and their underlying processes can be rational or irrational depending on whether they "give prior outlooks too much weight and fail to give proper weight to perceptual inputs" (Siegel 2017, 5). Arguments for this thesis are consistent with the existence of norms of rational perception–that is, norms governing the perceptual states and processes of rational subjects of experience, providing us with standards for the appraisal of particular cases. But these norms, if they exist, have yet to be identified. Here, we consider standard Bayesian norms of rational credence and explore whether they can apply to perception. Our main claim is that experience-forming processes can rationally be appraised based on a norm of probabilism, which is a coherence norm saying that one's perceptual states and processes ought to probabilistically fit together. Violations of probabilism in perception, we suggest, would characteristically be accompanied by a phenomenology comprising disunified conscious experiences and by epistemic and practical costs. The argument in support of our claim develops in two steps. We start by examining whether and how standard Bayesian norms of rational credence can ground the general types of irrational "perceptual hijacking" Siegel (2017) distinguishes: inherited inappropriateness, jumping to conclusions, and circularity. After concluding that none of these types of perceptual hijacking involves an apparent violation of a Bayesian norm, we then turn our attention to two specific Bayesian models used in psychiatry to account for perceptual anomalies in schizophrenia, namely: a circular inference model (Jardri & Denéve, 2013a,b) and a predictive coding model (Adams et al., 2013). We show that the circular inference model, but not the predictive coding model, can give us a principled reason to appraise an experience-forming process as irrational, based on the norm of probabilism. This is because the circular inference model can allow for "double counting" of the prior information or evidence processed by the perceptual system in generating experiences; and double counting is a plausible case of perceptual irrationality grounded in the violation of probabilism. Such a violation constitutes a case of probabilistic incoherence and can result in accurate, but irrational, perceptual experiences characteristic of certain psychiatric maladies. We conclude by exploring the phenomenological character of violations of probabilism, which comprise the disunified conscious experiences typically found in patients with schizophrenia. Our discussion considers both the cases for and against the rationality of perception thesis. If Bayesian norms cannot be used to determine the (ir)rationality of hijacked perception, and if Bayesian norms are constitutive of what it is to be Bayesian, then the human perceptual system is not Bayesian. If they can, then the best source of evidence to identify which norms apply to perception is in

computational modelling in psychiatry. Still, computational modelling in psychiatry indicates that the rationality of perception is pragmatically encroached, as it does not depend only on epistemic, truth-related factors like reliability or evidence, but also on pragmatic factors like what is practically at stake or valuable for the subject.

References

Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D., & Friston, K. J. (2013). The computational anatomy of psychosis. Frontiers in psychiatry, 4, 47. doi: 10.3389/fpsyt.2013.0004

Jardri, R., & Deneve, S. (2013a). Circular inferences in schizophrenia. Brain, 136(11), 3227-3241.

Jardri, R., & Denève, S. (2013b). Computational models of hallucinations. The neuroscience of hallucinations, 289-313.

Siegel, S. (2017). The rationality of perception. Oxford University Press.

Nicola Chinchella - Enacting Recovery: Virtual Reality, Active Inference, and Cognitive Behavioural Therapy

University Of Bologna
nicola.chinchella2@unibo.it

Attempts to successfully utilise virtual reality (VR) technology in the treatment of depression have thus far yielded disappointing results. Applications span from encouraging physical activity and simulating therapeutic interactions to more imaginative approaches like engaging with one's child self or exploring rich, soothing environments. Overall, though, even the most effective applications yield short-lived benefits, focus on eliciting positive affect, and generally fail to get to the fundamental dynamics that underpin depression as a chronic condition. In this paper, we argue that these disappointing results stem from a lack of consistent theoretical principles hindering the effective utilisation of VR's therapeutic potential. Here, we propose the active inference framework as a foundation. Active inference is a neuroscientific theory grounded in probabilistic belief updating and action, which furnishes an understanding of action and learning as uncertainty minimisation. Our overarching hypothesis is that VR can be used to present manageable uncertainty so that patients can continuously perform "just better than expected", thus resulting in a gradual update of their beliefs. Building on the potential of actions to bring about change and the malleability of VR, we propose a set of core principles for the design and use of VR in a therapeutic setting to reap its full potential. To produce core principles, we proceed as follows: we first review neural findings on depression, which highlight a difficulty in learning and updating one's beliefs. We side this with a phenomenological first-person account of depression, where a shrinkage of action possibilities and difficulties in future planning result in a reduced field of affordances and a narrowed lived space. We then read this evidence through active inference as stuck priors, warped learning rates and precision mechanisms, and reduction in policy number and length, leading to the failure of confidently reducing uncertainty in diverse contexts. We then propose how VR can be designed to tackle these challenges. Firstly, the VR environment can be leveraged to provide affordances that induce a consistent and calibrated learning rate, always allowing a "better than expected" performance. This practically translates into environments that yield manageable uncertainty, neither too easy nor too difficult to resolve, enabling a gradual update of the rigid sub-personal beliefs characteristic of depression. Crucially, this process must respect existing expectations about uncertainty minimisation, which cannot be challenged head-on but must be gently eroded over time. Secondly, to ensure therapeutic efficacy beyond VR, the virtual environments must be designed for maximal generalisability, allowing action-outcome mappings established in VR to 'seep' into the patient's offline world, i.e., avoiding fantastical scenarios. Lastly, we stress the critical role of supporting the extended temporal horizons of effective engagement with affordances, where VR treatment is uniquely placed to support learning over extended timescales (e.g., long sequences of actions), thereby maintaining and repairing the "temporal thickness" of an agent's learning. In conclusion, we hope that practitioners can pick up our

suggestion to design VR environments that marry well with the patients' need for an embodied and care-oriented type of treatment.

Benedetta Cogo - Getting Personal in the Philosophy of Psychiatry. Broadening Our Minds … and More

University of Wollongong
cogo.benedetta@gmail.com

Philosophical explorations about psychiatry target different topics—among them, issues of classification and characterisations of the various mental disorders, explanations of the different processes intertwined in mental conditions, ethical questions correlated to these accounts, and other aspects that constitute the reality of mental disorder. In this paper I will show that a proper characterisation of persons is needed in philosophy of psychiatry, if theorising about psychiatry ultimately aims at furnishing insights into the factual, lived reality of psychiatric conditions. I will justify and develop my argument in three steps. First, I will motivate the need for pursuing an appropriate philosophy of persons in psychiatry by analysing how three prominent accounts in philosophy of psychiatry fail, in different ways, to reflect the way the person is understood in the psychiatric practice – in their concrete, real-life situations (Gerrans 2014-2024, Fuchs 2021, Gallagher 2024). Second, I will show one viable direction for developing a philosophy of persons that does not engage with unnecessary and unhelpful metaphysical issues, and that can be applicable to the clinical practice. I will do so by relying on George Graham's (2010) theory of mental disorders as conditions that exhibit a "truncated rationality" (Graham 2010, p. 126). By stressing the relevance of rationality as one important factor in understanding psychiatric conditions, I will show that the three analysed exemplary accounts fail, in different ways, to give congruous consideration to rationality in mental disorders. Third, I will advance some correctives and clarifications that, starting from Graham's conception of mental disorders, are applicable to a philosophy of psychiatry that aims at being valuable for the clinical practice. I will focus on two main issues identified in Graham's account: the role and extension of the notion of rationality for the understanding of persons, and the appropriateness of the personal/subpersonal distinction in explaining various processes pertaining to the person. Ultimately, the criterion of success for my argument – and for a coherent characterisation of person in philosophy of psychiatry – is the applicability of this framework to the practice in a way that is of genuine value for the understanding and treatment of psychiatric conditions.

References

Fuchs, T. (2021). In Defense Of The Human Being: Foundational Questions Of An Embodied Anthropology. Oxford: Oxford University Press.

Gallagher, S. (2024). The Self and Its Disorders. Oxford University Press.

Gerrans, P. (2014). The Measure Of Madness: Philosophy Of Mind, Cognitive Neuroscience, And Delusional Thought. Cambridge, Massachusetts: Mit Press.

Gerrans, P. (2024). Anatomy of an Avatar. The virtual self in pathologies of self awareness. Oxford University Press.

Graham, G. (2010). The Disordered Mind. An Introduction to Philosophy of Mind and Mental Illness. Routledge.

Jacopo Colelli - Challenging Enactivist Anti-Representationalism: From Cognitive Mediation to Phenomenological Constitution

Università degli studi Roma 3
jcolelli@uniroma3.it

Enactivist approaches challenge the mediational epistemology of cognitive science, particularly by rejecting the central role of internal representations and symbolic content in guiding cognition. Enactivism instead emphasizes embodied interaction, sensorimotor coupling, and practical intentionality to "deny that cognition requires presence of a physical structure (a vehicle) carrying or encoding a content and which, in virtue of this content, stands for (or stands in for) events and properties outside of the agent" (Steiner, 2023, p. 6). Within this context, different varieties of enactivism reinterpret intentionalism and the role of content in cognition. Autopoietic enactivism (AE) links biology and phenomenology by proposing that content can be understood as a normative constraint imposed on the object as it is aspectually intended, guiding interaction without assuming intracranial representation. Radical enactivism (RE), by contrast, rejects content entirely, naturalizing intentionality through teleosemiotics—as informational sensitivity to natural signs, without mediation by contentful states. These differing correlations between intentionality and content reflect divergent critiques of the dominance of representation in cognitive science. Setting aside the ontological dimension, I focus on the epistemological question of whether abandoning—or refining—the traditional notion of content as a descriptive tool, and replacing it with intentionality as object-directedness without representation, can advance the explanatory aims of cognitive science. AE maintains that the aspectuality of experience cannot be reduced, as RE attempts to do, to behavioral dispositions or naturalistic mechanisms. Instead, it must be understood through phenomenological analysis of how objects are given in different normative modes. Yet both frameworks face limitations. AE must show whether—and how—its notion of phenomenological content can retain epistemological relevance within cognitive science if not linked to material vehicles. RE must clarify whether its account of object-directedness genuinely captures intentionality or merely describes behavioral tendencies, without addressing the experiential structure of directedness. In light of these challenges, I propose an alternative path grounded in phenomenology, yet oriented toward integration with cognitive science. Rather than treating phenomenology as a source of modes of presentation—expected to be mapped onto material or neural vehicles to ensure causal relevance—I suggest focusing on modes of constitution as a sui generis form of functional analysis. That is, we should examine how sequences of experiential processes generate and structure the aspects of the experienced object. This object, not the natural object as such, is constituted as the kind of object it is through the processes that produce it. On this view, content is not propositional or representational, but the experienced object itself, defined by its dependency on constitutive processes. The epistemological benefit of this approach lies in constructing phenomenological architectures—structured descriptions of the processes responsible for producing types and

classes of experienced objects. Content plays a central explanatory role not because it refers to a representational entity, but because it lets us identify fundamental differences between experiential types. These distinctions then guide the analysis of the underlying processes. Such architectures, obtained through functional decomposition, need not be mapped onto mechanisms. What matters is that cognitive models capture the correlations between constitutive processes and the aspects they give rise to.

References

Dengsø, M. J. (2025). Cognitive science and Cartesianism: Representational pull and methodological individualism. Journal of Theoretical and Philosophical Psychology.

Hutto, D. D., & Myin, E. (2013). Radicalizing enactivism: Basic minds without content. MIT press.

Hutto, D. D., & Myin, E. (2017). Evolving enactivism: Basic minds meet content. MIT press.

Hutto, D. D., & Myin, E. (2021). Re-affirming experience, presence, and the world: Setting the RECord straight in reply to Noë. Phenomenology and the Cognitive Sciences, 20, 971-989.

Kee, H. (2021). Phenomenology and naturalism in autopoietic and radical enactivism: exploring sense-making and continuity from the top down. Synthese, 198(Suppl 9), 2323-2343.

Noë, A. (2021). The enactive approach: a briefer statement, with some remarks on "radical enactivism". Phenomenology and the cognitive sciences, 20(5), 957-970.

Pokropski, M. (2020). Phenomenology and functional analysis. A functionalist reading of Husserlian phenomenology. Phenomenology and the Cognitive Sciences, 19(5), 869-889.

Steiner, P. (2023). The structure of intentionality: Insights and challenges for enactivism. Inquiry, https://doi.org/10.1080/0020174X.2023.2216753

Steiner, P. (2025). Not thinking about the same thing. Enactivism, pragmatism and intentionality. Phenomenology and the Cognitive Sciences, 24(1), 9-32.

Thompson, E. (2007). Mind in life: Biology, phenomenology, and the sciences of mind. Harvard University Press.

Thompson, E., (2018); Review of evolving enactivism: basic minds meet content

Varela, F. J., Thompson, E., & Rosch, E. (2017). The embodied mind, revised edition: Cognitive science and human experience. MIT press.

Mads Dengsø - The Mind and the Territory

University of Wollongong
madsdengsoe@gmail.com

Does embodied cognition break with representationalism and the mediational picture of the mind? Despite the insistence of various accounts from 4E approaches in cognitive science, the representational picture of cognition remains dominant. In response to this, researchers from 4E approaches have sought to come to grips with what has been termed as 'representational pull'. I will argue that 4E cognitive science faces underappreciated challenges in overcoming the representational standard in cognitive science. I will argue that this challenge, rather than a misapplication of cognitive science, stems from the methodological setup of cognitive science. In particular, I will argue that representationalism and mediational picture may be seen as a result of a methodological commitment to investigating minds and cognition at the level of individual organisms and their interactions. I will argue that this setup can be dated all the way back to the Cartesian foundations of early modern philosophy of mind. The overly intellectualized Cartesian view of the mind is best understood as a result of the constraints of methodological individualism. Contrary to the assumptions of much of 4E cognition, embodied and anti-intellectualist approaches based on individual organisms and their interactions have been around since the time of Descartes. Moreover, then, as today, embodied and anti-intellectualist approaches based around individual organisms and their interactions remain ineffectual at combating representationalism – insofar as the representational standard stems from the methodological commitment to individualism about minds. In the second half of the talk, I will argue that the methodological individualism, which continues to define cognitive science, is unsupported by a biological and biogenic approach to minds and cognition. Rather than a product of individual organisms, minds and cognition are emergent products of assemblages of multiple organisms. Cognitive systems may thus be distinguished from non-cognitive biological systems by their sympoietic (rather than autopoietic) organization. From this I will conclude that the methodological commitment of cognitive science to individualism about minds and cognition is not an apt fit for its explanandum. I will thus conclude by arguing that 4E cognition is right to resist representationalism, but that its current means are insufficient. However, the root cause of the error is not in a misapplication of cognitive science, but in its foundational methodological presuppositions. It is the methodological setup of cognitive which needs to be addressed if its manifestation in variously overly intellectualized, mediational, and representationalist guises are to be countered. Insofar as 4E accounts seek to break with mediational epistemology, they will have to confront the active role of the methodological contingencies of cognitive science, including its focus on individuals.

Mariela Destéfano - Doing Without Public Meanings

Consejo Nacional de Investigaciones Científicas y Técnicas
mariela.destefano@gmail.com

Inner speech can be defined as the subjective experience of language in the absence of overt and audible articulation (Fernyhough 2008). Currently, both empirical and theoretical approaches to inner speech have emphasized its cognitive function. Under this perspective, inner speech supports domain-general functions such as working memory, reasoning, behavioral and cognitive control, language processing, and problem-solving (Spelke, 2003; Baddeley, 1986; Law et al., 2013; Gilhooly, 2005; Hardy, 2006). For instance, inner speech is essential for maintaining and manipulating information in our minds (Baddeley, 1986), which is a core component of working memory. Additionally, it aids in the logical processing of information and forming coherent thoughts, thus facilitating reasoning (Spelke, 2003). Furthermore, inner speech helps regulate our actions and thoughts, contributing to self-control and executive function, which are critical aspects of behavioral and cognitive control (Law et al., 2013). In the realm of language processing, inner speech plays a significant role by facilitating the understanding and production of language (Gilhooly, 2005). Moreover, it allows for the conceptualization and solving of problems through internal dialogue, thereby enhancing problem-solving abilities (Hardy, 2006).

However, inner speech is not merely a cognitive capacity required for navigating and reflecting on the world. As Vygotsky (1934) already anticipated, there is a subjective component that underscores the personal and private character of inner speech. In this respect, Vygotsky defends the predominance of "senses" or personal meanings over public meanings in our internal speech, through which much of our affective and situational experience is filtered. This linguistic subjectivity in part informs our inner speech, producing an intimate construction space that goes beyond the boundaries of public semantics. In this talk, I will defend the thesis that inner speech includes what I call "idiosyncratic contents". That is, it is composed of linguistic representations, heavily tied to concrete personal experiences, and, therefore, less likely to be easily shared or understood through public language alone.

I will provide a detailed description of what I consider the two main conceptions of inner speech: the cognitive conception, which highlights its useful functions, and (ii) the expressive conception, which underscores it relationship with egocentric experience and subjective expression (Fossa 2022, 2017) Then, I will address the notion of "idiosyncrasy", drawing on perspectives from philosophy of language (Gendlin 1997), developmental psychology (Vygotsky 1934, Piaget 1972), and illustrate with examples how inner speech might involve these private contents not found in public language. After that, I will delve into the strongest objections found in the literature, examining empirical (Hulburt & Heavey, 2018) and theoretical (Bermúdez, 2018) challenges to idiosyncrasy. I will provide responses to these criticisms concluding that a perspective acknowledging the idiosyncratic nature of inner speech can be reconciled with, and enrich, an overall view of its cognitive functions. Moreover, I will discuss how to integrate this with the idea that inner speech also serves essential cognitive functions. Thus, the idiosyncratic element should not be seen merely as a surplus or an anomaly, but as a fundamental constituent of that inner voice shaping our mental life.

References

Alderson-Day, B. & Fernyhough, C. (2015). Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology. Psychological Bulletin, 141(5), 931-965.

Baddeley, A. D. (1986). Working Memory. Oxford: Oxford University Press.

Bemúdez, J. L. (2018). Thinking about Propositional Attitudes, or: Sharing Thoughts about Thoughts. In P. Langland-Hassan & A. Vicente (Eds.), Inner Speech. New Voices (pp. 60-79). Oxford: Oxford University Press.

Fernyhough, C. (2008). Getting Vygotskian about Theory of Mind: Mediation, Dialogue, and the Development of Social Understanding. Developmental Review, 28(2), 225-262.

Fossa, P. (2017). The Expressive Dimension of Inner Speech. Psicologia USP, 28(2), 318-326.

Fossa, P. (2022). Inner Speech: An Emancipatory Perspective. Manuscrito inédito, Universidad de Chile.

Gendlin, E. (1997). Experiencing and the Creation of Meaning: A Philosophical and Psychological Approach to the Subjective. Evanston, IL: Northwestern University Press.

Gilhooly, K. J. (2005). Working Memory and Strategies in Reasoning. In V. A. Green & B. Van Laar (Eds.), Logic and Decision Making (pp. 41-66). Mahwah, NJ: Erlbaum.

Hardy, L. (2006). Speaking Clearly: A Critical Review of the Self-Talk Literature. Psychology of Sport and Exercise, 7(1), 81-97.

Law, J., Rush, R., Schoon, I. & Parsons, S. (2013). Modeling Developmental Language Difficulties from School Entry into Adulthood: Literacy, Mental Health, and Employment Outcomes. Journal of Speech, Language, and Hearing Research, 56(4), 1409-1422.

Martínez-Manrique, F. & Vicente, A. (2015). The Activity and Format Approaches to Inner Speech. Frontiers in Psychology, 6, 1-8.

Piaget, J. (1972) Psicología de la inteligencia, Editorial Psique, Buenos Aires.

Spelke, E. (2003). What Makes Us Smart? Core Knowledge and Natural Language. Language in Mind: Advances in the Study of Language and Thought, 8(1), 277-311.

Vygotsky, L. S. (1934). Pensamiento y lenguaje. Madrid: Paidós

Dan Durso - LLMs, Plant Minds, and Why Multiple Realization is Not Enough

University of Illinois at Urbana-Champaign
ddurso3@illinois.edu

Discourse on whether Large Language Models (LLMs) like ChatGPT and OpenAI have minds, in some sense, is receiving significant attention in both the public sphere and academic circles. Whereas the public sphere is engaged in broader conversations that engage with the possibility of LLMs becoming sentient, researchers on the other hand, are taking a more specific approach and are attempting to uncover whether LLMs have certain mental capacities. These capacities range from whether LLMs can form propositional attitudes like beliefs and desires (Hase et al., 2021; Herrman & Levinstein, 2025; Chalmers, 2025), have a theory of mind (Holterman & van Deemter, 2023; Zhu et al., 2024; Jung et al., 2024) or a genuine folk psychology including the production of internal representations and dispositions to act (Goldstein & Levinstein, 2025). Skeptical challenges have already emerged challenging the potential mental capacities of LLMs, including claims that they are just "stochastic parrots" or that sensory grounding is a necessary yet missing component for mind realization. With this paper, I wish to add to these skeptical challenges by pointing out that many attributions of LLMs having minds are built on multiple realization arguments. I will argue, following Bechtel and Mundale (1999), that multiple realization arguments, at the cost of showing the sameness of mental states, often employ a mismatched broad- grained criterion. I will add that this kind of multiple realization argument usually leads to trivialization since it leaves open the possibility that many things can qualify for having mental states. I will highlight this point by demonstrating that many claims for LLMs having minds (e.g. propositional attitudes and theory of mind) permit most plants as qualifying for minds as well since a broad-grained criterion is assumed and this trivializes claims for LLMs having minds. I introduce research on plant biology (Seguno-Ortin & Calvo, 2021; Lee, 2023; Hansen, 2024) which reveals 2 that plants can be viewed as having propositional attitudes, theory of mind, and folk psychology when a broad-grained criterion is applied that mirrors the same criterion for LLM minds. The overall goal of this paper is not to defend the possibility that plants have minds, this depends on one's criterion, but rather to use the possibility of plant minds as an argument that highlights the weakness of multiple realization arguments that employ a broad-grained criterion. Additionally, I will argue that we must employ a fine-grained criterion when determining whether LLMs and other AIs can achieve mental capacities that meet the high standards of human minds.

References

Bechtel, W., & Mundale, J. (1999). Multiple realizability revisited: Linking cognitive and neural states. Philosophy of science, 66(2), 175-207. https://doi.org/10.1086/392683

Chalmers, D. J. (2025). Propositional interpretability in artificial intelligence. arXiv preprint arXiv:2501.15740. https://doi.org/10.48550/arXiv.2501.15740

Goldstein, S., & Levinstein, B. A. (2024). Does ChatGPT Have a Mind?. arXiv preprint arXiv:2407.11015. https://doi.org/10.48550/arXiv.2407.11015

Hansen, M. J. (2024). A critical review of plant sentience: moving beyond traditional approaches. Biology & Philosophy, 39(4), 13. https://doi.org/10.1007/s10539-024-09953-1

Hase, P., Diab, M., Celikyilmaz, A., Li, X., Kozareva, Z., Stoyanov, V., ... & Iyer, S. (2021). Do language models have beliefs? methods for detecting, updating, and visualizing model beliefs. arXiv preprint arXiv:2111.13654. https://doi.org/10.48550/arXiv.2111.13654

Herrmann, D. A., & Levinstein, B. A. (2025). Standards for belief representations in LLMs. Minds and Machines, 35(1), 1-25. https://doi.org/10.1007/s11023-024-09709-6

Holterman, B., & van Deemter, K. (2023). Does chatgpt have theory of mind?. arXiv preprint arXiv:2305.14020. https://doi.org/10.48550/arXiv.2305.14020

Jung, C., Kim, D., Jin, J., Kim, J., Seonwoo, Y., Choi, Y., ... & Kim, H. (2024). Perceptions to beliefs: Exploring precursory inferences for theory of mind in large language models. arXiv preprint arXiv:2407.06004. https://doi.org/10.48550/arXiv.2407.06004

Lee, J. (2023). What is cognitive about 'plant cognition'?. Biology & Philosophy, 38(3), 18. https://doi.org/10.1007/s10539-023-09907-z 4

Segundo-Ortin, M., & Calvo, P. (2022). Consciousness and cognition in plants. Wiley Interdisciplinary Reviews: Cognitive Science, 13(2), e1578. https://doi.org/10.1002/wcs.1578

Zhu, W., Zhang, Z., & Wang, Y. (2024). Language models represent beliefs of self and others. arXiv preprint arXiv:2402.18496. https://doi.org/10.48550/arXiv.2402.18496

Lotem Elber-Dorozko & Devin Y. Gouvêa - ‘Neuronal representation’ is not a defective concept: Ambiguity as a sign of science in progress

The Hebrew University of Jerusalem
lotem.elber@mail.huji.ac.il

neuroscience and concluded that its ambiguity and imprecision constitute serious, perhaps even fatal, flaws (Baker, Lansdell, and Kording 2022; Favela and Machery 2023; Pohl et al. 2024; Vilarroya 2017). Inspired by the literature on polysemous and imprecise concepts in philosophy of science, we defend a very different conclusion: the ambiguity of 'representation' is not a sign that the concept itself is flawed. Instead, both the concept’s ambiguity and its accompanying discontent share a common cause—the uncertainty currently inherent in the field. Philosophers of science commonly defend conceptual ambiguity by attributing to it benefits such as facilitating interdisciplinary research. Such defenses often appeal to the epistemic goals served by these concepts (Brigandt 2010; Haueis 2024; Neto 2020). We in turn suggest that 'representation' serves two central epistemic goals of neuroscience—to explain mental phenomena by appeal to neural phenomena, and to explain cognitive capacities. To understand why scientists rely on such an ambiguous concept in pursuit of these goals, however, we think it is not enough to identify any benefits it might currently offer (see also Novick 2023). Instead, we make the much stronger claim that the ambiguity of ‘representation’ is necessary given the epistemic goals it serves and the current state of its empirical evidence. Focusing on the epistemic goal of explaining cognitive capacities, we note that it is still unclear what form(s) successful explanations will take and how to distinguish explanatory from non- explanatory scientific findings. This problem does not go away even if one fully buys into some well-developed philosophical framework of scientific explanation, such as functional (Cummins 1975) or mechanistic (Craver 2007; Piccinini 2021) analysis. For too little is known today to say which among the possible correlational and causal relations that scientists identify is relevant to the explanation of cognitive capacities. The result is that scientists use the term 'representation' to indicate that they seek relations that are explanatorily relevant, without knowing how to determine which those are. Thus the concept ‘representation’ is ambiguous and imprecise, attributed to different phenomena without a clear way of distinguishing which is and which isn’t a ‘representation’. Our analysis of ‘representation’ suggests that its ambiguity cannot be resolved through philosophical analysis or substitution with other concepts. Because the ambiguity in the concept reflects uncertainly in the scientific field, the way to resolve this ambiguity is by mitigating the field’s uncertainty. This can be done through empirical work. We suggest that the standards for neuroscientific explanation, and thus for 'representation' itself, will become clearer through the accumulation of generally accepted examples. Moreover, the ambiguity in ‘representation’ is not expected to hinder the pursuit of such examples, which relies on scientific intuition about explanation. We support this claim through an example of a successful neuroscientific project, which yielded a generally accepted explanation despite its reference to ‘representations’, namely sound localization in the owl (Carr and Konishi 1990).

References

Baker, Ben, Benjamin Lansdell, and Konrad P Kording. 2022. "Three Aspects of Representation in Neuroscience." Trends in cognitive sciences 26(11): 942–58.

Brigandt, Ingo. 2010. "The Epistemic Goal of a Concept: Accounting for the Rationality of Semantic Change and Variation." Synthese 177(1): 19–40. https://doi.org/10.1007/s11229-009-9623-8.

Carr, C E, and M Konishi. 1990. "A Circuit for Detection of Interaural Time Differences in the Brain Stem of the Barn Owl." The Journal of neuroscience : the official journal of the Society for Neuroscience 10(10): 3227–46.

Craver, Carl F. 2007. Explaining the Brain. Oxford University Press. Cummins, Robert. 1975. "Functional Analysis." The Journal of Philosophy 72: 741–65.

Favela, Luis, and Edouard Machery. 2023. "Investigating the Concept of Representation in the Neural and Psychological Sciences." Front. Psychol.

Haueis, Philipp. 2024. "A Generalized Patchwork Approach to Scientific Concepts." The British Journal for the Philosophy of Science 75(3): 741–68. https://www.journals.uchicago.edu/doi/10.1086/716179.

Neto, Celso. 2020. "When Imprecision Is a Good Thing, or How Imprecise Concepts Facilitate Integration in Biology." Biology & Philosophy 35(6): 58. https://doi.org/10.1007/s10539-020-09774-y.

Novick, Rose. 2023. "The Neutral Theory of Conceptual Complexity." Philosophy of Science 90(5): 1021–30. https://www.cambridge.org/core/journals/philosophy-of-science/article/neutral-theory-of-conceptual-complexity/3A616D9B882A550750F78247F2C501BD.

Piccinini, Gualtiero. 2021. Neurocognitive Mechanisms. Oxford University Press.

Pohl, Stephan et al. 2024. Desiderata of Evidence for Representation in Neuroscience. https://arxiv.org/abs/2403.14046.

Vilarroya, Oscar. 2017. "Neural Representation. A Survey-Based Analysis of the Notion ." Frontiers in Psychology 8. https://www.frontiersin.org/article/10.3389/fpsyg.2017.01458.

Francesco Ellia - Intrinsic Universal Structures and Extrinsic Local Functions

ATR - Computational Neuroscience Laboratories & Monash University
francescoellia@gmail.com

Recent advances in consciousness science have exposed a critical impasse: competing theoretical frameworks—structuralist vs. functionalist (Fleming & Shea 2024, 2024b; Song 2024; Ellia & Tsuchiya 2024; Kleiner 2024), universal vs. local (Albantakis et al. 2023; Kanai & Fujisawa 2024; Fleming 2024), and intrinsic vs. extrinsic (Doerig et al. 2019; Tsuchiya et al. 2020; Negro 2020; Ellia et al. 2021)—are inducing philosophical deadlocks and conceptual standstills that hinder further empirical testing. Although these debates have generated valuable insights, they have largely proceeded in parallel without a systematic framework for understanding their relationships and implications. In this work, we argue that these parallel disputes reflect deeper tensions in conceptualizing consciousness and can only be resolved by recognizing three fundamental dimensions defining all theories: (1) intrinsic (first-person) vs. extrinsic (third-person) perspectives, (2) universal (substrate-independent) vs. local (human-specific) scope, and (3) structure (formal organization) vs. function (cognitive roles) as explanatory priorities. Our three-dimensional framework offers a meta-theoretical approach that resolves theoretical gridlock and bridges disciplines by reframing debates as differences in dimensional priorities. This approach clarifies longstanding disagreements and identifies testable predictions while connecting foundational debates to translational applications. In doing so, it addresses underlying assumptions, epistemic trade-offs, and potential overlaps among different positions. The intrinsic/extrinsic dimension distinguishes between theories that prioritize a system's internal perspective and those that rely on external measurements. Extrinsic approaches align with traditional scientific methods—measuring neural correlates and behavioral markers—whereas intrinsic approaches argue that consciousness must be understood from within the system itself. This distinction is critical for investigating consciousness in novel systems where conventional behavioral correlates may be unclear or misleading. The universal/local dimension reflects a theory's intended scope. Local theories focus on specific manifestations of consciousness, typically human, and are grounded in well-understood neural and cognitive mechanisms. Universal theories, by contrast, seek principles applicable to any conscious system regardless of its physical substrate or evolutionary history. This distinction is especially relevant for exploring consciousness in artificial systems or non-human entities. The structure/function dimension captures whether theories emphasize organizational patterns or functional roles. Structure-focused theories highlight the formal properties that give rise to consciousness, often employing mathematical formalisms to describe isomorphisms between these properties and the target system. Function-focused theories, on the other hand, center on cognitive capabilities—such as attention, report, or metacognition—that may indicate or correlate with consciousness. These dimensions appear to interact and form clusters of two camps, such as

intrinsic-universal-structural theories as opposed to extrinsic-local-functional ones. For instance, intrinsic approaches often lean toward universal scope, defining consciousness through properties that could exist across various systems (Kanai & Fujisawa, 2024; Fleming, 2024). While this perspective can expose potential blind spots in current frameworks, it also points to interesting theoretical implications of exceptions. For example, mathematical approaches of category theoretical approach may unify extrinsic-intrinsic characterization of qualia through the Yoneda lemma (Tsuchiya & Saigo 2021). Recognizing these interactions clarifies why certain approaches cluster together and exposes potential blind spots in current frameworks.

Doudja Boumaza - Multiple functions hypothesis and episodic memory

Centre for philosophy of memory (Grenoble, France)
Keio University (Tokyo, Japan)
boumazadoudja@gmail.com

The literature on episodic memory has expanded significantly in the past decade. Neuroimaging studies showing that the same cognitive systems are activated when one remembers and when one imagines (Addis & Schacter 2007) have led to the response that philosophical literature has started to explore the nature of episodic memory. Thus, philosophers have attempted to provide criteria to distinguish episodic memory from other mental states, such as imagination, in light of new empirical data. Moreover, the function of episodic memory has emerged as a topic of growing interest in recent analytic literature. The primary focus has been on the biological function of memory, with philosophers debating whether this function is present in humans and, more recently, in animals. Today, three primary hypotheses regarding the evolutionary function of episodic memory are defended in the philosophical literature. The first suggests that episodic memory serves to retain past personal information (Boyle, 2019; Boyle, 2022). The second, aligning with neuroimaging findings, posits that memory evolved to enable future planning, decision-making, and counterfactual thinking (Addis & Schacter 2007; Michaelian 2011; Michaelian 2016; Robins & Schulz 2023). The third hypothesis argues that episodic memory provides epistemic authority for the rememberer in social communication, i.e., the justified ability to base beliefs on personal past experiences (Mahr & Csibra 2018). These three views represent different functions of episodic memory, oriented toward different scales of action, such as learning, anticipating, or justifying. The first, the mnemonic function, suggests that episodic memory specifically enables an orientation toward the past, while the simulative function motivates an orientation toward the future. The third, the communicative function, suggests that episodic memory is a tool for enabling social bonds by providing individuals with specific attitudes. My claim in this paper is that episodic memory is a multifaceted system serving multiple, sometimes independent, but mainly coextensive biological functions. The co-occurrence of these functions, I defend, underscores the flexibility and plasticity of episodic memory, a feature central to its evolutionary success. For instance, I submit, an important function of episodic memory is overlooked in the philosophical literature, despite being well-defined in the psychological literature. This is the sensorimotor function of episodic memory, which posits that episodic memory underlies perception and enables action in the present by reusing preserved sensory inputs (Bergson, 1896). My argument in defense of this thesis is run as follow: from a Darwinian framework I propose that episodic memory has been shaped by a variety of evolutionary pressures, encompassing not only biological needs but also social and cultural factors. I will run two arguments in support of my multiple function hypothesis. First is the necessity to consider cultural evolution as a cause for the emergence of a new function. To defend a functional approach to episodic memory, one should first consider the selected effect theory. But in this

perspective, the function is the result of a long process of selection and some actual functions are too recent to be due to biological evolution (Tomasello, 2001). My argument is that cultural evolution theory allows us to ascribe recent functions to an item developed in and by a cultural context without disproving existing functions. A second argument is the brain recycle hypothesis (Dehaene & al, 2010) defending an adaptation of the brain to cultural changes by creating new cognitive skills built on existing ones. I suggest that if the areas of the brain activated for visual recognition can be recycled for other functions than their evolutionary one(s), it could theoretically be possible that other brain areas such as the hippocampus could evolve and episodic memory could therefore also have new functions culturally adapted, such as a sensorimotor function.

References

Addis, D. R., A. T. Wong, and D. L. Schacter. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. Neuropsychologia 45 (7): 1363- 1377.

Boyle, A. (2019). The impure phenomenology of episodic memory. Mind and Language 35 (5):641-660. DOI : 10.1111/mila.12261

Boyle, A. (2022). The mnemonic functions of episodic memory. Philosophical Psychology, 35

Darwin, C. (1859). On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: John Murray.

Dehaene & al (2010) How learning to read changes the cortical networks for vision and language. In Sciences, vol.330.

Mahr, J. and Csibra, G. (2018). Why do we remember? The communicative function of episodic Memory. Behavioral and Brain Sciences, 41, 1-63.

Michaelian, K. (2016). Mental Time Travel. Episodic Memory and Our Knowledge of the Personal Past. MIT Press.

Millikan, R. (1987) In defense of Proper Function. Philosophy of Science, Vol. 56, pp. 288- 302.

Schulz, W, A. Robins, S. (2023) Episodic Memory, Simulated Future Planning, and their Evolution. Rev.Phil.Psych. 14, 811–832. https://doi.org/10.1007/s13164-021-00601-1

Schwartz, A. (2020). Simulationism and the Function(s) of Episodic Memory. Review of Philosophy and Psychology, 11: 487-505.

Tomasello, M. (2001). Cultural origins of human cognition. Harvard University

Itamar I. Dreymann - A New (and Empiricist) Theory of Metaphor - The Transparency-Sheet Theory

Hebrew University of Jerusalem
itamar.dreymann@mail.huji.ac.il

Theories of metaphor have always struggled to accommodate two essential characteristics of metaphorical thought:

1. Metaphor's unique phenomenology; which is often described as imagistic, vivid, and affective
   (Alston,1964,p.97; Goodman,1976,p.68; Davidson,1978,pp.31,46-7; Camp,2007,pp.2).
2. Metaphor's explanatory power; which is often described as the metaphor's ability to utilize one subject matter in order to lead us to realizations about a different subject matter and to guide our deliberations about it.
   (Alston,1964,pp.98-9; Goodman,1976,p.71; Davidson,1978,pp.45-6 Camp,2007,pp.2,21).

As an example, let's use "Juliet is the sun". Historically, most philosophers have theorized about metaphor utilizing theories of concepts that treat concepts as abstract linguistic entities that conform to the rules of logic (e.g. Goodman,1976; Davidson,1978; Lakoff & Johnson,1980; Fogelin,2011; Searle,1993; Walton,1993,2013,2015; Reimer,2007; Camp,2009,2015,2020). I maintain that this approach has created the problem for past and present theorists: They were forced to choose between either explaining metaphorical thought in abstract terms which are detached from our phenomenology or focusing on our phenomenology of metaphor but denying that it has anything to do with conceptual thought due to its bizarre nature. In this talk I will sketch how an empiricist theory of conceptual thought (Prinz,2005) enables us to construct a theory of metaphor that explains and accommodates both of the essential characteristics of metaphors. In particular, I will assume a theory of concepts according to which concepts are unconscious black boxes that assemble together such that given the inputs of an uttered phrase and certain semantic and pragmatic context, they yield as an output a suitable imagined prototype (Prinz,2002,ch.6; Del Pinal,2016; Murez,2021). The proposed theory of metaphors - like traditional ones - explains metaphors as inexplicit analogies (Goodman,1976,pp.71-80; Lakoff & Johnson,1980); but the same core idea is implemented differently: The theory claims that when we encounter a new-to-us metaphor, our first step is to take the metaphor literally despite the apparent category mistake. We understand that we are asked to make an odd compositional procedure, but we execute it to the best of our abilities nevertheless: We form a prototype of the metaphor's predicate (e.g. the earth orbits the sun), and then we attempt to create a prototype of the metaphor's subject that incorporates the predicate's prototype or some of its aspects (e.g. Romeo following Juliet around obsessively). You can think about this method as akin to the method of an illustrator who uses transparency-sheets to juxtapose two drawings and contemplate on which features she should copy from one to another. Then, if the initial predication went well, we can repeat the process to elaborate the result: First, we expand the predicate's prototype. For example, we might consider how the sun applies a pulling-force on earth and how the earth is bound to orbit the sun for eternity. Then, we apply the prototype again in search for further analogies. For example, how Juliet's charm is causing for Romeo's

obsession which is expected to persist. Despite eventually ceasing to entertain the sun's prototype, some of the interrelations between the properties within the sun's prototype have survived the transition: our understanding of Romeo's love is now (partially) structured like our understanding of gravity. This is the explanatory power of metaphors. The phenomenal character of metaphors is accounted for using a cluster of phenomena that includes the juxtaposition & copying process itself and the - at times grotesque - character of the resulted prototype.

References

Alston, W. (1964). Philosophy of Language. Englewood Cliffs, N.J.: Prentice Hall.

Camp, E. (2007). Showing, telling and seeing. The Baltic International Yearbook of Cognition, Logic and Communication 3 (1):1-24.

Camp, E. (2009). Two Varieties of Literary Imagination: Metaphor, Fiction, and Thought Experiments. Midwest Studies in Philosophy 33 (1):107-130.

Camp, E. (2015). Logical Concepts and Associative Characterizations. In The Conceptual Mind Edited by Laurence, S., & Margolis, E. (pp. 591-621). The MIT Press. https://doi.org/10.7551/mitpress/9383.003.0034

Camp, E. (2020). "Imaginative Frames for Scientific Inquiry: Metaphors, Telling Facts, and Just-So Stories.". pp. 304–336 in The Scientific Imagination. Edited by Arnon Levy and Peter Godfrey-Smith. Oxford: Oxford University Press.

Davidson, D. (1978). What Metaphors Mean. Critical Inquiry, 5(1), 31–47. • Del Pinal, Guillermo (2016). Prototypes as compositional components of concepts. Synthese 193 (9):2899–2927.

Fogelin, R. J. (2011). Figuratively Speaking: Revised Edition (2nd ed.). Oxford University Press

Goodman, N. (1976). Languages of art: an approach to a theory of symbols ([2nd. ed.]). Hackett. Lakoff, G., & Johnson, M. (1980). Metaphors we live by. University of Chicago Press.

Murez, Michael (2021). Belief Fragments and Mental Files. In Cristina Borgoni, Dirk Kindermann & Andrea Onofri, The Fragmented Mind. Oxford: Oxford University Press. pp. 251-278.

Prinz, J. J. (2002). Furnishing the mind: concepts and their perceptual basis (1st ed.). MIT Press.

Prinz, J. J. (2005). The return of concept empiricism. In H. Cohen & C. Leferbvre, Categorization and Cognitive Science. Elsevier.

Reimer, M. (2007). Metaphorical Meanings. Do you see what I mean? The Baltic International Yearbook of Cognition, Logic and Communication 3.

Searle, J. (1993). Metaphor. In A. Ortony (Ed.), Metaphor and Thought (pp. 83- 111). Cambridge: Cambridge University Press.

Walton, K. L. (1993). Metaphor and Prop Oriented Make-Believe. European Journal of Philosophy, 1(1), 39–57.

Walton, K. L. (2013) Metaphor, Fictionalism, Make-Believe: Response to Elisabeth Camp.

Walton, K. L. (2015). Existence as Metaphor? in In Other Shoes: Music, Metaphor, Empathy, Existence. New York: Oxford University Press.

Parker DuVall - Can AI Agents Contribute to Group Understanding?

Colorado State University
parkerduvall@colostate.edu

As the understanding literature continues to evolve, the notion of group understanding has become increasingly important. In relative lockstep, artificial intelligence systems continue to mature. With the rise of AI, it stands to reason that these systems will become vital parts of teams working to solve problems. Indeed, we may say that AI systems will begin to contribute to group knowledge, but it is an open question as to whether or not they will or can contribute to group understanding. One of the seminal accounts of group understanding is that of Kenneth Boyd. In this paper, I take Boyd's conception of group understanding and investigate if and how AI systems could play a role. I argue that of Boyd's two types of group understanding, only an in inflationary account can accommodate an AI agent as a part of a group. In section two, I start by laying out Boyd's minimal conception of understanding, as well as in inflationary and deflationary understanding. To keep the matter as simple as possible, I make no social claims as to whether or not my arguments would work with specific conceptions of understanding. Instead, I stick to Boyd's outline of the concept. Then, in section three, I argue that a deflationary account cannot accommodate an AI agent due to the black box nature of current AI systems making it either impossible or unreliable to investigate the reasons an AI agent believes in a proposition. In section four, I give my argument in favor of an inflationary account of AI group understanding with a case modelled on Boyd's "Dependable Autobody" case. Specifically, I imagine a group that is composed mostly of humans, with at least one AI system. This allows me to largely sidestep the question of whether AI or not can ultimately understand on its own and instead focus on whether or not it can contribute to the understanding of a group. Then, in section ve, I consider the objection that an AI system cannot contribute in the process of group grasping as it cannot have trust in the other members doing their part and thus cannot be "mutually p-reliant" on the other members of the group. In my response, I argue that the members of a group need not actually have a positive trust relationship, but instead just need to avoid a distrusting relationship, maintain open communication, and evenly share the workload of the goal in order to achieve mutual p-reliance. While I use the word "trust" in my initial case, there need not be any positive feelings towards the other members from each member of the group. In the case of the "Dependable Autobody" example, as long as they do not actively distrust the other members of the group and maintain open communication as to where they are in their respective processes, the output is identical. Therefore, even if we do not assign any kind of "trust" capacity to an AI system, it can still be mutually p-reliant and contribute towards group understanding.

Zeynep Sina Ersan - On How Pain Perception and Pleasure Processing are Causally Connected: A Case of BDSM Interest in Individuals with ADHD+

İhsan Doğramacı Bilkent University
sina.ersan@bilkent.edu.tr

This paper explores the causal interplay between pain perception and pleasure processing in the context of BDSM interests among individuals diagnosed with ADHD+. Challenging the traditional psychoanalytic assumption that BDSM practices stem from psychopathology or trauma, it will be argued that interest in BDSM is not necessarily caused by psychopathology or traumas in ADHD+ people. In the paper, only the pain aspect of BDSM will be discussed in terms of neurobiology of ADHD. Drawing on interdisciplinary literature, the paper discusses altered pain perception in ADHD+ individuals—ranging from hypersensitivity to difficulty differentiating painful stimuli—and examines how dopamine dysregulation may mediate enhanced pleasure responses to pain. Three theoretical frameworks—the Opponent Process Theory, the Gate Control Theory, and the Distraction Theory—are applied to explain how consensual BDSM can trigger heightened dopamine release and sexual satisfaction in ADHD+ individuals. The analysis ultimately deconstructs core premises of the traditional anti-BDSM argument and advocates for a reconceptualization of BDSM as a potentially healthy form of sexual expression for neurodivergent individuals. The study highlights the importance of incorporating neurodiversity into philosophical and psychological discourse on sexuality as well as increasing the neurodivergent visibility in the academic studies.

Marco Facchin - Motifs for a radically embodied cognitive ontology

Universiteit Antwerpen
marco.facchin.marco.facchin@gmail.com

The ontology of mainstream (computational & representational) cognitive science consists of computational/representational kinds, ideally mappable on neural kinds [1,2]. Non-mainstream, radically embodied cognitive science rejects the computational, representational and neurocentric assumptions of its counterpart [3,4], and so it must have a very different cognitive ontology. And yet, this issue is rarely explored in the literature. As a result, the shape of the alternative "radically embodied" cognitive ontology is still unclear. My talk shall attempt to shed some light on the matter. More specifically, I wish to uncover the motifs of a radically embodied cognitive ontology: the vague, partially unspecified and open-ended ideas acting as very general constraints on the cognitive ontology of any individual radically embodied research programme [5]. To this end, I will analyze a number of prominent radically embodied research programmes [3,4, 6-14], and identify the following very general ideas:

(1) Representational/computational kinds are substituted by "interactive" kinds (such as affordances [3,6, 10-13], sensorimotor contingencies [4,7-9], or optimal grip [10-12])
(2) The relevant material kinds upon which the kinds in (1) are ideally mappable are no longer neural kinds, but parameters of the complex agent-environment coupled system
(3) The mapping will be context-dependent, and take into account diachronic factors such as the agent's development history [see esp. 14]

I will then claim that (1)-(3) are the motifs of a radical embodied cognitive ontology: the vague, open-ended ideas constraining the individual cognitive ontologies of individual radically embodied research programmes. I intend that claim to be descriptive: the motifs in (1)-(3) should actually be "played" by radically embodied research programmes. To check that this is the case - and thus validate my claim - I will take a look at the actual experimental research practices of radically embodied cognitive science. I will thus look at how ecological psychologists identify and parametrize affordances [15,16] and how enactivists formalize sensorimotor contingencies and use them in explanations [17], and claim that these practices de facto are variations on the motifs I uncovered.

References

[1] Poldrack, R. A. (2010). Mapping Mental Function to Brain Structure: How Can Cognitive Neuroimaging Succeed? Perspectives on Psychological Science, 5(6), 753–761.

[2] McCaffrey, J. B. (2023). Evolving concepts of functional localization. Philosophy Compass, 18(5), e12914.

[3] Chemero, A. (2009). Radically embodied cognitive science. MIT Press.

[4] Di Paolo et al. (2017). Sensorimotor life. OUP.

[5] Raja, V. (2024) The motifs of radical embodied neuroscience. European Journal of Neuroscience, 60(5), 4738-4755.

[6] Anderson, M. (2014) After phrenology. MIT Press

[7] Thompson, E. (2007) Mind in life. Harvard University Press

[8] Hutto, D. & Myin, E. (2013). Radicalizing Enactivism. MIT Press

[9] Hutto, D. & Myin, E. (2013). Evolving Enactivism. MIT Press

[10] Bruineberg, J. (2017) Active Inference and the Primacy of the "I Can". In: Metzinger, T. and Wiese, W., Eds., Philosophy and Predictive Processing, MIND Group, Frankfurt am Main, 67-84.

[11] Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. Frontiers in human neuroscience, 8, 599.

[12] De Haan, S. (2020). Enactive psychiatry. Cambridge University Press.

[13] Favela, L. H. (2023). The ecological brain: Unifying the sciences of brain, body, and environment. Taylor & Francis.

[14] Di Paolo, E. et al. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. Philosophy and the Mind Sciences, 3 https://doi.org/10.33735/phimisci.2022.9187

[15] Warren, W. H. (1984). Perceiving affordances: visual guidance of stair climbing. Journal of experimental psychology: Human perception and performance, 10(5), 683.

[16] Konczak, J., Meeuwsen, H. J., & Cress, M. E. (1992). Changing affordances in stair climbing: the perception of maximum climbability in young and older adults. Journal of Experimental Psychology: Human Perception and Performance, 18(3), 691.

[17] Buhrmann, T., Di Paolo, E. A., & Barandiaran, X. (2013). A dynamical systems account of sensorimotor contingencies. Frontiers in psychology, 4, 285.

Yichu Fan - Phenomenon Reconstitution as a Tool for Explanatory Progress

The University of Edinburgh
yichu.fan@ed.ac.uk

The notion 'phenomenon reconstitution' was first proposed by Bechtel and Richardson (1993/2010) to refer to the recharacterisation of the explanandum phenomenon in a causal or constitutive mechanistic explanatory project. In recent years, several authors in the new mechanist literature have offered further examples and discussions on phenomenon reconstitution, focusing on its relation with the level of mechanisms (Kronfeldner, 2015) and with mechanism discoveries (Colaço, 2020), its role in experimentation (Bechtel&Vagnino, 2022), and its application in non- explanatory contexts (Bohallgen, 2021). In these various philosophical accounts, phenomenon reconstitution is mainly understood as a correction of some mistaken or inaccurate characterisations of the phenomenon as a result of explanatory or technological progress. In this paper, I want to explore how phenomenon reconstitution can serve as an instrument for explanatory progress, not necessarily characterising the phenomenon in a more accurate way, but making it more amenable to further explanation. Specifically, by examining the case study of the explanatory project on the motor control in marine mollusc Clione Limacina, I show that the explanandum phenomenon, i.e. the locomotion in the animal, was characterised in various ways throughout the research. In particular, the same phenomenon was identified with the swimming of the intact animal, the movement of the wings of the animal, the rhythmic neuronal activity in isolated pedal ganglia preparation, etc. Moreover, I argue that these shifts in the characterisation of the phenomenon were 'forward-looking,'i.e. motivated by the ever-evolving explanatory goals such as mechanism isolation. This is in contrast with a 'backward-looking' conception of phenomenon reconstitution, which is motivated by past explanatory discoveries. In other words, I argue that in this case, phenomenon reconstitutions were made to facilitate further explanatory progress rather than as a corrective response to discoveries. I will further argue that Bechtel's and Richardson's (1993/2010) the case study of classical genetics is also subject to this interpretation, where the characterisation of the 'Mendelian traits' shifted from macroscopic phenotypic traits to single enzymes in order to preserve a one-to-one gene- trait explanatory mapping that is conducive to further mechanistic explanations. Finally, to account for this instrumental use of phenomenon reconstitution, I draw on Feest's (2011, 2017) account of 'phenomenon stabilisation,' and argue for a dependency-oriented view of the goals of mechanistic explanatory projects. This dependency-oriented view is complementary to the received explanandum-oriented view, according to which a mechanistic explanatory project starts with a fixed, predetermined explanandum phenomenon and aims to pin down counterfactual dependences relevant to it. On the other hand, according to the dependency-oriented view that I propose, a mechanistic explanatory project only starts with a 'blurry'subject matter, which is amenable to different characterisations. Moreover, the goal of capturing stable counterfactual dependences sometimes overrides the need for preserving the characterisation of the explanandum. In these cases, the explanandum is moulded in a way to

capture the best (the most stable or most important) dependences in the system; phenomenon reconstitution hence plays an instrumental role for explanatory progress.

References

Arshavsky, Y. I., Beloozerova, I., Orlovsky, G., Panchin, Y. V., & Pavlova, G. (1985a). Control of locomotion in marine mollusc Clione limacina I. Efferent activity during actual and fictitious swimming. Experimental Brain Research, 58, 255– 262.

Arshavsky, Y. I., Beloozerova, I., Orlovsky, G., Panchin, Y. V., & Pavlova, G. (1985b). Control of locomotion in marine mollusc Clione limacina II. Rhythmic neurons of pedal ganglia. Experimental Brain Research, 58(2), 263–272.

Arshavsky, Y. I., Beloozerova, I., Orlovsky, G., Panchin, Y. V., & Pavlova, G. (1985c). Control of locomotion in marine mollusc Clione limacina III. On the origin of locomotory rhythm. Experimental Brain Research, 58, 273–284.

Arshavsky, Y. I., Deliagina, T., Orlovsky, G., Panchin, Y. V., Pavlova, G., & Popova, L. (1986). Control of locomotion in marine mollusc Clione limacina: VI. Activity of isolated neurons of pedal ganglia. Experimental Brain Research, 63, 106–112.

Bechtel, W., & Richardson, R. C. (1993/2010). Discovering complexity: Decomposition and localization as strategies in scientific research. MIT press.

Bechtel, W., & Vagnino, R. (2022). Figuring out what is happening: The discovery of two electrophysiological phenomena. History and Philosophy of the Life Sciences, 44(2), 20.

Bollhagen, A. (2021). The inchworm episode: Reconstituting the phenomenon of kinesin motility. European Journal for Philosophy of Science, 11(2), 50.

Colaço, D. (2020). Recharacterizing scientific phenomena. European Journal for Philosophy of Science, 10, 1–19.

Feest, U. (2010). Concepts as tools in the experimental generation of knowledge in cognitive neuropsychology. Spontaneous Generations: A Journal for the History and Philosophy of Science, 4(1), 173–190.

Feest, U. (2011). What exactly is stabilized when phenomena are stabilized? Synthese, 182(1), 57–71.

Feest, U. (2017). Phenomena and objects of research in the cognitive and behavioral sciences. Philosophy of Science, 84(5), 1165–1176.

Kronfeldner, M. (2015). Reconstituting phenomena. Recent Developments in the Philosophy of Science: EPSA13 Helsinki, 169–181.

Woodward, J. (2005). Making things happen: A theory of causal explanation. Oxford university press.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. Biology & Philosophy, 25, 287–318.

Luis Favela - Emergence Makes Neuroscience Possible

Indiana University Bloomington
lhfavela@iu.edu

Neuroscience has historically been strongly associated with reductionism. This is evident, at least in part, by interpretations of the neuron doctrine (cf. Gold & Stoljar, 1999) as motivating claims that even the "highest" forms of brain activity (e.g., consciousness; Crick, 1994) are epistemically and metaphysically reducible to neural activity. Philosophers inspired by such positions (e.g., Bickle, 2003; Churchland, 1986) have done scientifically-grounded work to provide arguments in support of the possibility of appealing to reductionistic neuroscience to solve or dissolve many of the deepest philosophical mysteries, such as consciousness, free will, and memory, to name a few. The aim of the current work is to argue that, perhaps ironically, emergence is what makes neuroscience possible. By "neuroscience" I mean a scientific discipline—inclusive of subdisciplines (e.g., behavioral, computational, and molecular neurosciences)—that investigates spatiotemporal scales of investigation comprised of regularly structured phenomena susceptible to being referred to by a consistent set of concepts and methods. By "possible" I mean the actual existence of phenomena at such spatiotemporal scales. Accordingly, the ability to do neuroscience results from the combination of real patterns (cf. Dennett, 1991) in the world that clump together in ways scientists can investigate with consistently-applied words and tools. What I have just described is one way complexity scientists understand the term "emergence" (cf. Krakauer, 2024; Morin, 2023). According to some complexity scientists, complex systems are emergent phenomena in the sense that a collection if microstates (e.g., hydrogen and oxygen molecules) go through processes of self-organization that result in a significant decrease in degrees of freedom that produce a steady macrostate (e.g., body of water; Jensen, 2023; cf. Wilson, 2010). For complexity scientists, the identification of emergent phenomena (i.e., metaphysical claim) goes hand in hand with the use of effective theories (i.e., epistemic method). Effective theories are theories because they are able to organize phenomena under an efficient set of principles, and they are effective because it is not impossibly complex to compute outcomes (Wells, 2022). In view of that, what makes doing special sciences (i.e., any science other than physics) possible is that phenomena within those bounds of inquiry (e.g., chemistry, biology, psychology, etc.) are emergent and, thus, not reducible to "lower" spatiotemporal scales. Brains are paradigmatic complex systems: their spatiotemporal structures are the result of self-organized processes that produce consistent forms of reduced degrees of freedom that are real and irreducible to structures at lower scales. Neuronal avalanches are presented as one such example (Beggs, 2022). Thus, while neuroscience has a history of being a paradigmatic reductionistic discipline, it is actually an example of emergence par excellence.

References

Beggs, J. M. (2022). The cortex and the critical point: Understanding the power of emergence. Cambridge, MA: The MIT Press.

Bickle, J. (2003). Philosophy and neuroscience: A ruthless reductive account. Springer Science & Business Media.

Churchland, P. S. (1986). Neurophilosophy: Toward a unified science of the mind/brain. Cambridge, MA: The MIT Press.

Crick, F. (1994). The astonishing hypothesis: The scientific search for the soul. New York, NY: Macmillan Publishing Company.

Dennett, D. C. (1991). Real patterns. The Journal of Philosophy, 88(1), 27-51. https://doi.org/10.2307/2027085

Gold, I., & Stoljar, D. (1999). A neuron doctrine in the philosophy of neuroscience. Behavioral and Brain Sciences, 22, 809-869. https://doi.org/10.1017/S0140525X99002198

Jensen, H. J. (2023). Complexity science: The study of emergence. New York, NY: Cambridge University Press.

Krakauer, D. C. (2024). The complex world: An introduction to the foundations of complexity science. Santa Fe, NM: SFI Press.

Morin, E. (2023). The challenge of complexity: Essays by Edgar Morin. Eastbourne, UK: Sussex Academic Press. Wells, J. D. (2022). Effective theories in physics: From planetary orbits to elementary particular masses. New York, NY: Springer.

Wilson, J. (2010). Non-reductive physicalism and degrees of freedom. The British Journal of the Philosophy of Science, 61, 279-311. https://doi.org/10.1093/bjps/axp040

Maria Fedorova - The Self Reimagined on Psychedelics

University of Vienna
mariia.fedorova@univie.ac.at

Psychedelics can help people recover from addiction (Krebs & Johansen, 2012; Bogenschutz et al., 2015; Johnson et al., 2014, 2017; West et al., 2015; see Letheby & Gerrans for overview). One view is that they can do so by shifting people's perspectives on themselves (Letheby, 2021; Pickard, 2020). In particular, someone suffering from addiction may discover, under psychedelics, a possibility of being addiction-free. The central question of this paper is how a person who has addiction can come to change her perspective on herself while on psychedelics. I argue that people can discover a possibility of being addiction-free by means of spontaneous imaginative immersion (Lawson & Thompson, 2024) into that very possibility during a psychedelic experience. I begin by spelling out what constitutes a perspective and what it means for it to change. Following Sliwa (2023), perspectives are sets of cognitive and affective dispositions, and a change in perspective is a change in these dispositions. I determine what dispositions may constitute the perspective of someone with addiction, and what it might take to change her dispositions. Following Pickard (2020), many people who have addiction belong to a community of those with addiction. They thus share the community's dispositions, which constitute their perspectives. Shifting such perspectives requires that a person has access to what being addiction-free would be like. To gain such access, for Pickard, we can either (i) effortfully imagine this possibility, or (ii) have this possibility revealed to us by psychedelics. Against this, I argue that psychedelics do not replace but facilitate imagination. I examine the consequences of Pickard's view for explaining shifts in perspectives on oneself under psychedelics. I show that Letheby's (2021) account of such shifts—an elaboration of Pickard's view—fails to adequately explain them. Letheby argues that shifts in perspective under psychedelics are due to the psychedelic effects on fundamental beliefs about the self, which underlie our perspective. Since the influence of these beliefs is unconscious, our perspective manifests as real and fixed. Psychedelics can change our perspective, as they reveal its constructed and mutable nature. I challenge Letheby's view thus. First, his view can only account for how psychedelics help people realise that their perspectives can change in principle—not how they actually change. To shift a perspective, our way of thinking needs to be restructured (Camp, 2017) during a psychedelic experience. Letheby focuses on the unconscious changes caused by psychedelics and neglects the causal role of people's conscious experience while on psychedelics. Thus, his view cannot accommodate the restructuring. Second, following people's reports, perspective-shifting under psychedelics is closely linked with their prior goals, affective, and perceptual states during a psychedelic experience, as these automatically guide the experience's course. Letheby's account risks portraying perspective-shifting as something random, whereas these reports point to the contrary. Hence, I propose an alternative to Pickard's and Letheby's view. I look more closely at how people think while on psychedelics and why they think the way they do, given the nature of psychedelic experiences. Following Lawson & Thompson (2024), I propose that

perspective-shifting facets of a psychedelic experience are best understood in terms of intensified daydreaming—spontaneous imaginative immersion. Psychedelic-facilitated daydreaming is characterised by decreased control over thoughts and increased drifts in attention between them (Mason et al., 2021; Wießner et al., 2022a, 2022b), as well as by its hyper-imagistic and hyper-immersive character (Girn et al., 2020; Kraehenmann, 2017; Kraehenmann et al., 2017). Under psychedelics, we float in the stream of mind-wandering, occasionally submerged in eddies of daydreams about ourselves and our lives. It is when people struggling with addiction are so submerged that their perspective can shift under psychedelics. They come to spontaneously reimagine themselves as addiction-free through psychedelic-induced daydreaming. My account meets the above challenges, given that imagination functions as cognitive manipulation of mental representations (Stokes, 2014, see also Myers, 2021; Kind, 2020, 2021, 2022) and since daydreaming is automatically guided by people's goals, affective, and perceptual states (Lawson & Thompson, 2024). I then consider an objection against my proposal from the hallucinatory character of a psychedelic experience, which may rule out daydreaming as a plausible candidate for describing psychedelic-induced thinking.

Anastasija Filipović - A Unified Account of Extended Affectivity: Sensorimotor Enactivism & Predictive Processing

University of Belgrade
anastasijafilipovic@yahoo.com

This presentation offers a unified explanation of the extended affectivity hypothesis by integrating sensorimotor enactivism and predictive processing (PP), framed through the concept of structural representations (S-representations). Our goal is to show how affectivity extends beyond the individual, shaped by both environmental interactions and internal predictive mechanisms, and to provide a more comprehensive account by analyzing both personal and subpersonal levels of cognition. At the personal level, sensorimotor enactivism and the notion of the affective niche explain how affectivity is extended through active engagement with the environment. Perception is not passive reception of stimuli but an active, skill-based process of exploring the world, guided by sensorimotor contingencies—rules governing the interplay between action and sensory feedback. These contingencies underpin our embodied sense-making. As we engage with our surroundings, we co-create affective niches: dynamic feedback loops between the organism and selected parts of the environment. These niches, shaped by our sensorimotor activity, hold constitutive significance for our affective states. Thus, affectivity becomes extended, not as a metaphor, but as an ongoing, reciprocal process between the organism and its micro-world. At the subpersonal level, PP explains how the brain supports this extension by continuously generating predictions about sensory input. Rather than passively constructing representations, the brain uses hierarchical generative models to actively infer the causes of sensory data. This ongoing process, known as active inference, is both dynamic and action- oriented. Although PP and enactivism differ in their treatment of representations and the role of the brain, we propose a synthesis. Specifically, we argue that PP should move away from brain-centrism, while enactivism should reconsider its anti-representational stance. The key to this integration lies in the concept of S-representations—neural structures that maintain a structural correspondence between cognitive systems and their targets, while remaining action-oriented. In this view, the brain's generative models function as S- representations, encoding probabilistic structures that mirror causal features of the world. These representations are not static but support dynamic prediction and interaction, thus bridging the divide between embodied action and internal modeling. By aligning the embodied, action-guided view of enactivism with the structured, inferential account of PP through S-representations, we offer a cohesive explanation of how affectivity extends. On the subpersonal level, the brain maps the external world through action-oriented predictions; on the personal level, we construct affective niches that feed back into these predictions. This unified framework shows that selecting or shaping an affective niche is not just an experiential act, but a subpersonal process grounded in the predictive and structural capacities of the brain.

References

Clavel Vázquez, M.J. (2019). A match made in heaven: predictive approaches to (an unorthodox) sensorimotor enactivism. Phenomenology and the Cognitive Sciences 19, 653– 684.

Colombetti, G. (2015). Enactive Affectivity, Extended. Topoi 36 (3).

Downey, A. (2017). Radical Sensorimotor Enactivism & Predictive Processing – Providing a Conceptual Framework for the Scientific Study of Conscious Perception. In T. Metzinger & W. Wiese (Eds.). Philosophy and Predictive Processing: 11. MIND Group

Facchin, M., Negro, N. (2023). Predictive Processing and Extended Consciousness: Why the Machinery of Consciousness Is (Probably) Still in the Head and the DEUTS Argument Won't Let It Leak Outside. In: Casper, MO., Artese, G.F. (eds) Situated Cognition Research. Studies in Brain and Mind, vol 23. Springer.

Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature Reviews Neuroscience 11, 127–138.

Friston K. (2013). Life as we know it. Journal of the Royal Society Interface 10: 20130475

Gładziejewski, P. (2016). Predictive coding and representationalism. Synthese, 193(2), 559– 582.

Gärtner, K. & Clowes W. R. (2017). Enactivism, Radical Enactivism and Predictive Processing: What is Radical in Cognitive Science?. Kairos. Journal of Philosophy & Science 18 (1), 54-83.

Kondor, Z. (2020). Neural-based vs. Enactive Approaches to Consciousness and Social Cognition. Avant 11, (2).

Krueger, J. & Colombetti, G. (2018). Affective affordances and psychopathology. Discipline Filosofiche 2 (18), 221-247.

Lee, J., & Calder, D. (2023). The many problems with S-representation (and how to solve them). Philosophy and the Mind Sciences, 4, 8.

O'Regan K.J. & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. Behavioral and brain sciences 24, 939–103.

Piccinini, G. (2022). Situated neural representations: Solving the problems of content. Frontiers in Neurorobotics, 16, 846979.

Schlicht, T. & Starzak, T. (2021). Prospects of enactivist approaches to intentionality and cognition. Synthese 198 (3), 89-113.

Shagrir, O. (2012). Structural representations and the brain. The British Journal for the Philosophy of Science, 63(3), 519–545.

Seth, K. A. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. Cognitive Neuroscience 5 (2), 97-118.

Segundo-Ortin, M., Hutto, D.D. (2021). Similarity-based cognition: radical enactivism meets cognitive neuroscience. Synthese 198, 5–23.

Thornton, M.A., Tamir, D.I. (2024). Neural representations of situations and mental states are composed of sums of representations of the actions they afford. Nature Communication 15, 620.

James Grayot - How does the embodied and extended mind internalize content?

University of Porto
james.grayot@gmail.com

How does the manipulation of external representational vehicles, like public symbols, diagrams, or equations, transform human cognition? Do embodied and extended minds really internalize external representations? If so, what does this entail for the status of internalized content? One common way of answering these questions is through the thesis of neural reuse: evolutionarily old neural networks are repurposed to allow for the production and management (representation) of new content by changing our cortical circuitry and affording new sensory-motor opportunities thereby new ways to engage with the external environment. This is a plausible theory and has much support in both cognitive scientific and philosophical communities (see, e.g., Anderson 2010, 2021; Menary 2014). However, the theory of neural reuse retains many of the pitfalls of traditional representational theories of the mind, which make it difficult to square with what we now know about the embodied, extended, and ecological basis of cognition. In fact, it remains a fierce matter of debate whether neural networks are capable of representing anything at all (let alone abstract concepts and reasoning procedures). As such, if the thesis of neural reuse has anything to say about how cognition is transformed by engagement with external representations, it is only a (small) part of the story. Moderate theories of cognitive embodiment and extension, like Clark's (2008, 2016) extended functionalism or Menary's (2007, 2015) cognitive integration are perhaps better suited to addressing the internalization question if one thinks that there is more to cognitive transformation than neural reuse. But taking the moderate route raises important ontological challenges, namely, how do internal and external representations come together, i.e., 'dove-tail', in the transformation of cognition? What distinguishes 'internal' from 'external' representational processes, and how do we account for their interaction? In response to these questions, I explore two alternatives for further fleshing out whether and how internalization is possible (it at all). The first alternative is to frame internalization in terms of symbolic affordances, which I take to be dispositions to respond to external representations in a transformative way. This view is shared by various radical theories of embodied/enactive cognition and ecological psychology which deny a role for internal representations (Hutto & Myin 2017; Rietveld & Kiverstein 2014; Kiverstein & Rietveld 2018). A consequence of taking this view is that there is no problem to solve—it simply doesn't make sense to speak of internalization. However, as suggested in (Anonymous forthcoming) this view fails to appreciate how novel content is produced in complex cognitive tasks, like mathematical reasoning and formal inference-making procedures. The second alternative is to frame internalization as an effect of inner speech, which is known to play various roles in relating personal activities to subpersonal ones (Perrone-Bertolotti et al. 2014). A consequence of this view, however, is that it runs the risk of divorcing the content of representations (traditionally construed) from their vehicles (Anonymous forthcoming). My aim in this paper is to explore in how far both alternatives can be leveraged to better understand how the internalization of external

representations transforms human cognition. I will argue that one does not exclude the other, and further, that both are compatible with the thesis of neural reuse. However, the ensuing account of internalization I aim to sketch requires adopting a conception of representation (or rather, of representing) that departs dramatically from that of traditional cognitive science and psychology, viz. as a static and discrete subpersonal entity. In order to account for internalization, we must adopt a conception of representing as a multi- modal and temporally unfolding process.

References

Anderson, M. L. (2010). Neural reuse: A fundamental organizational principle of the brain. Behavioral and brain sciences, 33(4), 245-266.

Anderson, M. L. (2021). After phrenology: Neural reuse and the interactive brain. MIT Press.

Anonymous (forthcoming). Representation hunger: Reformulating the 'problem domain' of truly complex cognition. Mind & Language.

Clark, A. (2008). Supersizing the mind: Embodiment, action, and cognitive extension. OUP USA.

Clark, A. (2015). Surfing uncertainty: Prediction, action, and the embodied mind. Oxford University Press.

Hutto, D. D., & Myin, E. (2017). Evolving enactivism: Basic minds meet content. MIT press.

Kiverstein, J. D., & Rietveld, E. (2018). Reconceiving representation-hungry cognition: an ecological-enactive proposal. Adaptive behavior, 26(4), 147-163.

Menary, R. (2007). Cognitive integration: Mind and cognition unbounded. Springer.

Menary, R. (2014). Neural plasticity, neuronal recycling and niche construction. Mind & Language, 29(3), 286- 303.

Menary, R. (2015). Mathematical cognition: A case of enculturation. Philarchive.org

Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciu, M., & Loevenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self- monitoring. Behavioural brain research, 261, 220-239.

Rietveld, E., & Kiverstein, J. (2014). A rich landscape of affordances. Ecological psychology, 26(4), 325-352.

Lucca Greco - Affect Theory and the Expansion of Chomsky's Methodological Naturalism

University of Brasília
luccagreco@gmail.com

Chomsky's (1994, 2000) methodological naturalism advances internalism—the view that mental phenomena are best explained as computational systems largely independent of external input—as the sole viable path to naturalizing the mind. He positions naturalistic internalism as a more suitable method for safeguarding the autonomy of mental inquiry while preparing for its unification with the natural sciences, all without presupposing any metaphysical view of the term 'mind'. Still, Chomsky (1994, 2000) concedes that, despite its success in modeling syntax and phonology, internalism leaves other vital dimensions of cognition (e.g., intentional and agential capacities) unexplained, branding them as "mysterious" and beyond empirical scrutiny. Our central claim in this paper is that such gaps arise not from the phenomena themselves but from internalism's self-imposed methodological constraints, especially its detachment from sensory-motor functions. While this rule once served a useful heuristic purpose, it now unnecessarily narrows the scope of naturalistic mental inquiry, which today possesses the means to both address and overcome such restrictions. To illustrate this potential, we draw on Panksepp's (1998) affective neuroscience in its project of mapping subcortical circuits that integrate bodily states with goal-directed behavior, and on the somatic marker hypothesis (Damasio, 1996; Bechara et al., 2000; Bechara & Damasio, 2005), a proposal treating decision-making as interactions between interoceptive signals and cortical computation. These cases demonstrate that internalism's limits are not ontological but methodological and can therefore be revised with no detriment to the autonomy of the study of the mind if attention turns to affectivity. We then suggest that methodological naturalism's upcoming step is to pave the way for the development of a naturalistic affect theory, thus expanding its scope from computational syntax to encompass a broader range of cognitive capacities. The central challenge lies in justifying the empirical tractability of affective states, without which no general theory about them can be developed. A further obstacle involves equipping the theory with the explanatory principles necessary for it to gain scientific traction. We conclude by summarizing two promising approaches currently at our disposal to meet these two core issues within naturalistic parameters: the OCC model of emotions (Ortony et al., 2022) and Hufendiek's (2016) account of embodied emotions.

References

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making, and the orbitofrontal cortex. Cerebral Cortex, 10(3), 295–307. https://doi.org/10.1093/cercor/10.3.295

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. Games and Economic Behavior, 52(2), 336–372. https://doi.org/10.1016/j.geb.2004.06.010

Chomsky, N. (1994). Naturalism and dualism in the study of language and mind: Agnes Cuming Lecture, 1993. International Journal of Philosophical Studies, 2(2), 181–209. https://doi.org/10.1080/09672559408570790

Chomsky, N. (2000). New Horizons in the Study of Language and Mind. Cambridge University Press.

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. Philosophical Transactions of the Royal Society B: Biological Sciences, 351(1346), 1413–1420. https://doi.org/10.1098/rstb.1996.0125

Hufendiek, R. (2016). Embodied Emotions: A naturalist approach to a normative phenomenon. Routledge.

Ortony, A., Clore, G. L., & Collins, A. (2022). The Cognitive Structure of Emotions (2nd ed.). Cambridge University Press.

Panksepp, J. (1998). Affective Neuroscience: The foundations of human and animal emotions. Oxford University Press.

Alberto Greco - How can cognitive science make explanations in a single framework: using flows beyond levels

University of Genoa
greco@unige.it

Cognitive science, as a multidisciplinary enterprise, faces the problem that several and different explanations of the same phenomenon can be given by different disciplines. In particular, neuroscientific explanations are often competing with psychological ones. Different explanations are often considered as speaking about distinct levels of cognition. We claim, however, that the true question is not which level of cognition should be better investigated (and often the neural one prevails), but which discipline is involved, and how to account for the fact that each discipline has different objects and methods. The solution proposed can be summarized as follows. The idea is to adopt some general or abstract concepts which may permit to encompass diverse perspectives under a single unifying umbrella. It is suggested to analyze cognitive tasks as scientific objects from the perspective of different disciplines (Abney, 2014; Agazzi, 1991). This leads to descriptions concerning different conditions of the system, called "states", which are specified by variables or predicates belonging to a particular discipline (e.g. a physical state, or a state of the body, of the brain, of consciousness, etc.). Changes of state that give rise to new states will be called "events" (see Casati & Varzi, 2008 for an analysis of this concept). A succession of events occuring in time is called "flow". This idea captures the fact that psychological phenomena are processes which deploy themselves in time. Given the different nature of states and events, there may be a number of different "flows", e.g. a physical flow, a neural flow, a behavioral flow, etc. Different flows are different descriptions of what is happening at a certain point in time. In other words, the representation of different flows describes, in parallel, from various disciplinary standpoints, the same events occurring in a certain time course (called a "flow-chain"). This allows to establish the nature of correspondences and links between events in the same or different flows. Such links may be vertical (correspondences between different flows at the same time) or horizontal (i.e. changes of state in one flow at different times). Such links may also be causal or simply correlational (see Figure 1).

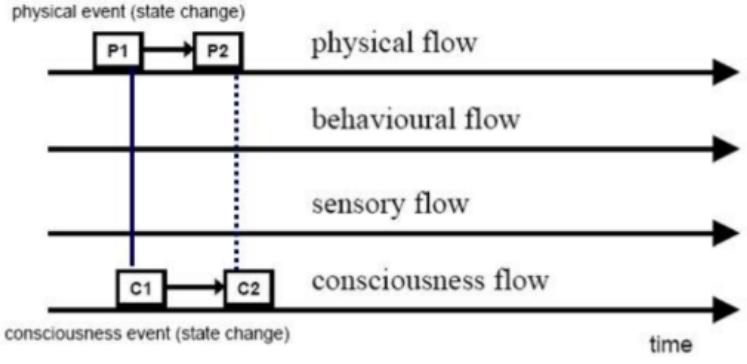

Figure 1- - Correspondences in a flow-chain

In the present paper, after explaining in detail the proposed model, we give some examples of how it can be applied in multidisciplinary analyses of specific phenomena (in particular, the "phi phenomenon", psycho-physical isomorphism, Libet experiment about awareness of decision, attentional errors, Stroop effect). Steps for explanation will be described. It is argued that rather than looking for the most important discipline it is necessary to look in which stream the most important event is, in a "pragmatic explanation" frame (van Fraassen, 1988). We argue that this proposal can give an answer to the need for a unifying framework for a single cognitive science as opposed to many cognitive sciences, while avoiding any sort of reductionism.

Philip Groth - The Epistemic Role of Mental Imagery in Perceptual Belief Formation

University of California, Santa Cruz
pgroth@ucsc.edu

Representationalists about perception disagree about the nature of mental content. Some philosophers claim that perceptual content is propositional, while others deny this. The motivation behind the claim that perception has propositional content is that it explains how perception causes and justifies perceptual beliefs. If we deny that perception has propositional content, then we do not have a straightforward account of perceptual belief: how can non-propositional content give rise to propositional belief? This paper explains how perceptual experience with non-propositional and non-conceptual content can result in and justify belief. Mental imagery can explain how it is that we can form propositional beliefs, even if we think that perception is a non-propositional attitude. I argue that mental imagery has propositional contents, which it gets from cognition and represents them symbolically. However, the representational vehicles are similar to perception: mental imagery represents its contents in a quasi-propositional way - viz. in a sense-modality. Because mental imagery and perception share this similarity, mental imagery can make meaning out of perception: By visualizing stored representations of conceptual and propositional contents, and projecting them against the perceptual representations, mental imagery allows us to invoke conceptual and propositional contents, which we use in thought and language. There are reasons to think this is the case based on neural processing that happens in the brain. In case of mental imagery, the top-down processing begins with cognition and ends up with a representation in a format similar to that of perception. In fact, some neuroscientists and philosophers call imagery vision in reverse. Imagery begins in the frontal cortex, which triggers neural responses that retrieve information from the medial temporal areas. The neural chain travels further to the visual cortex to form sensory representations. In contrast, perception works in the opposite direction: bottom-up. It begins with a sensory stimulus (in case of vision that would be light hitting the retina) which triggers a response in the visual cortex to send information up the chain. Perception and imagery thus represent different kind of information. Perception represents external objects, as it is triggered by external stimulus. Mental imagery, however, begins with thought and thus represents the concepts invoked in that thought. We have different kinds of content: objects in case of vision and concepts in case of mental imagery. The representing, however, happens in both cases in the visual cortex: the contents are represented in a visual format despite the difference in nature of the content.

Ewa Grzeszczak - Psychiatric categories as natural kinds – is the application of HPC to psychiatry worth it?

Jagiellonian University
ewa.grzeszczak@doctoral.uj.edu.pl

A currently popular conception of natural kinds in psychiatry is the theory of Homeostatic Property Clusters (HPC), which understands natural kinds as clusters of properties sustained by mechanisms (first proposed by Boyd, e.g., 1991, 1999; for applications to psychiatry, see e.g., Samuels 2009, Beebee & Sabbarton-Leary 2010, Kendler, Zachar & Craver 2011, Tsou 2016). One reason for its popularity is that HPC allows for the inclusion of properties from different levels of explanation within a single cluster. However, this undeniable advantage of multi- levelness has a flip side: the relationship between neurobiological and psychological properties must be clarified, and the definitions of key concepts crucial for HPC, such as causality and mechanism, must be adjusted accordingly. I aim to show that such adjustments result in HPC applying to psychiatry only in a weakened version, which has significant drawbacks. Firstly, applying HPC to psychiatry only allows for mental disorder to form a whole in a trivial, deflationary sense, and thus fails to provide substantial explanations of it as a unified neurobiological-psychological phenomenon. Secondly, applying HPC to psychiatry forces a global commitment to a weak theory of natural kinds. My argument is based on demonstrating that psychological properties are incompatible with the requirements of a strongly understood HPC. I first discuss how the psychological level requires giving up a productive understanding of causality in favor of interventionism. This shift, in turn, raises the question of how to understand mechanism in an interventionist HPC. To address this question, I distinguish between a substantive mechanistic explanation (which enables understanding of a given process in terms of how it occurs) and a deflationary mechanistic explanation (which only allows for the prediction that something will happen). I argue that only the deflationary mechanistic explanation is possible between the neurobiological and psychological levels in the HPC cluster of mental disorder. The first consequence is that the explanatory value of natural kinds in psychiatry is significantly diminished. After all, one of the reasons we care about natural kinds is that they allow us to categorize – dividing the world into members and non-members of a given natural kind – and then use these categories to provide explanations. However, if the psychological level is merely "attached" to the rest through a deflationary mechanism, then we cannot offer interesting, substantive explanations about the whole category. Secondly, if we accept that a connection through a deflationary mechanism is sufficient to classify something as an HPC kind, we significantly lower the threshold for recognizing natural kinds. Thus, applying HPC to psychiatry entails accepting a specific theory of natural kinds – namely, one that postulates an abundance of kinds and undermines the special status of those posited by sciences such as biology or medicine. This stands in tension with a significant part of the motivation for advocating for natural kinds in psychiatry – namely, the conviction that they guarantee the scientific status of the discipline by directly linking psychiatry to medicine.

References

Beebee H., Sabbarton-Leary N. (2010), Are psychiatric kinds real?, "European Journal of Analytic Philosophy" 6(1), 11-27.

Boyd R. (1991), Realism, anti-foundationalism and the enthusiasm for natural kinds, "Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition" 61(1/2), 127-148.

Boyd R. (1999), Homeostasis, species, and higher taxa, Cambridge, MA: A Bradford Book/MIT Press.

Kendler K. S., Zachar P., Craver C. (2011), What kinds of things are psychiatric disorders?, "Psychological medicine" 41(6): 1143-1150.

Samuels R. (2009), Delusion as a natural kind. In Psychiatry as Cognitive Neuroscience: Philosophical Perspectives, M. Broome, L. Bortolotti (eds.), Oxford: Oxford University Press, 49-80.

Tsou J. Y. (2016), Natural kinds, psychiatric classification and the history of the DSM, "History of Psychiatry" 27(4), 406-424.

Louis Gularte - Affective Valence as Reward & Punishment (schm)Representation: Solving the Representationalist-Motivationalist Dilemma

Tulane University
lgularte@tulane.edu

There seems to be an emerging consensus about positive and negative emotion, in the reinforcement-learning and decision-making corner of affective neuroscience — namely, what we could call 'reward-representationalism': the view that positive and negative emotion-states ('valenced' affect) are best understood in terms of reward- and punishment-representation. The question I take up in this paper is what that emerging consensus entails — supposing it's correct — for existing theories of the nature of affective valence. Current theories fall into two main camps. The first ('representationalism') says that the essential difference between positive and negative affect is a difference in representational content, and the second ('motivationalism') says it's a difference in motivation or action-tendencies. My more specific question is thus whether understanding valenced affect in terms of reward- and punishment-representation favors representationalism or motivationalism about affective valence. My thesis is that it in fact supports a third kind of view with elements of both. In particular, I argue that reward- and punishment-representation — and thus ultimately positive and negative affective valence — are indeed representational, but that they are distinguished not by a difference in any independently specifiable representational content but by their distinct representational formats, which are themselves constituted by distinct motivational action-tendencies. In other words, I argue that under the hood, the emerging reward-representationalist consensus supports what we might call motivational schmrepresentationalism about affective valence. I close by highlighting the promise of such a view (i) for addressing the main challenges to both representationalism and motivationalism and (ii) for explaining a number of familiar asymmetries in the effects of positive and negative affect on (e.g.) attention and decision-making.

Cagatay Gutt - Temporal Binding of Brain and Body: Neural-Autonomic Phase Synchrony During Emotional Processing

IfADo
cagatay.gutt@rub.de

Emotional states are fundamentally embodied, emerging from the dynamic interplay between central neural processing and peripheral physiological adjustments orchestrated by the autonomic nervous system (ANS). While ANS outputs like heart rate variability (HRV) and electrodermal activity (EDA) reflect emotional arousal and valence, under- standing the precise temporal coordination between brain activity and these peripheral signals is crucial for elucidating brain-body interactions. This study investigates neural- autonomic phase synchrony during the conscious processing of distinct emotional states (positive, negative, neutral) by quantifying the temporal alignment between cortical and physiological rhythms. We employ a multimodal approach, simultaneously recording high-temporal-resolution electroencephalography (EEG), electrocardiography (ECG) for HRV analysis (specifically Root Mean Square of Successive Differences, RMSSD), EDA, and functional near-infrared spectroscopy (fNIRS) while participants view validated emotional video clips. Our pri- mary analysis quantifies the Phase Locking Value (PLV) between frontal EEG oscillations (Alpha, Beta bands) and continuous signals derived from HRV (reflecting parasympathetic influence) and phasic EDA (reflecting sympathetic influence). EEG channel selection for PLV analysis was informed by task-related hemodynamic activity measured via fNIRS to focus on functionally relevant cortical areas. We hypothesize that PLV, indicating brain-body temporal integration, will be signifi- cantly modulated by emotional content compared to neutral conditions. We further expect synchrony strength to correlate with subjective arousal ratings. By examining the phase synchrony between brain signals and ANS-mediated physiological outputs, this research provides novel insights into the dynamic, embodied mechanisms underlying emotional experience. Understanding this temporal binding is critical for models of psychophysiological function and may inform assessments of cognitive load or stress regulation capacity, potentially impacting performance monitoring and optimization in demanding operational environments.

Sinem ELKATIP HATIPOGLU - The distinction between conscious and non-conscious mental states: How should state consciousness be understood?

Marmara University
sinem.hatipoglu@marmara.edu.tr

The distinction between conscious and non-conscious mental states is pivotal to understanding consciousness. The history of the views concerning the distinction follows an interesting path; going from once thinking that all mental states are invariably conscious and that consciousness is the mark of the mental to thinking that there are obviously non-conscious mental states too but that maybe some mental states are still invariably conscious to thinking that any kind of mental state that can occur consciously can occur non-consciously as well.1 Meanwhile various theories of consciousness were developed to explain how a conscious mental state differs from a non-conscious one.2 Although the attribution of consciousness to states and its explanation came relatively later 3 , the notion of state consciousness became widely endorsed and attempts to explain consciousness are typically understood as attempts to explain state consciousness.4 However it is not clear if the notion of state consciousness is sufficiently understood and whether it does more damage than good to our attempts to explain consciousness. The ambiguity of the notion of state consciousness is particularly apparent in debates concerning an objection brought up against higher order (HO) theories of consciousness. According to HO theories, what distinguishes a conscious mental state from a non-conscious one is that the former is represented by another mental state, viz. the HO state.5 However representation can go wrong and when a HO state represents the subject to be in some mental state that the subject is not in - which is possible since according to HO theories HO states are sufficient for consciousness - it is far from clear how consciousness is attributed to a mental state that does not exist. I refer to this as the empty HO state objection.6 Although I don't think that the objection threatens the theory, I also think that these objections and replies reveal a necessity to dig deeper into the notion of state consciousness and that the success of the replies depend on how state consciousness is understood. Some in fact suggested that as long as consciousness is taken to be a property of subjects and not states HO theories would not face a problem.7 I contend that while the notion of state consciousness undoubtedly provides a practical way to refer to and talk about mental lives and their contents, attributing consciousness to a state consists in an approach that is too atomic but attributing it to subjects may be too elusive. In this paper, I critically examine these two approaches hoping to find some middle ground. This in turn might help us understand the history of the views on the distinction between conscious and non-conscious mental states and why there's been a persisting inclination to think that some mental states are invariably conscious.

Julian Hauser - Explicit self-representation and immunity to error through misidentification

University of Barcelona
julian@julianhauser.com

This paper examines immunity to error through misidentification (IEM), focusing particularly on self-representation. I argue that IEM emerges when tacit representational content is made explicit. I illustrate this by first examining non-human animal cognition, specifically how bats transition between egocentric and allocentric representations. When a bat converts distance information from its sonar (egocentric) into a cognitive map (allocentric), the resulting self-ascription is IEM. The self-ascription is IEM because it results from the making explicit of tacit representational content that is always or necessarily true. That bat only puts information from _its_ senses into the egocentric representation and it only uses it for _its_ locomotion. When the tacit assumptions on both the input and output side are always (or necessarily) true, then the resulting self-ascription is de facto (or logically) IEM. Such transitions between representations are looked at in both Recanati's and Ismael's work, but both face limitations. Recanati, for instance, only discusses self-attribution of intentional states and leaves aside cases of introspection. This makes sense: for him, it's the experiential mode that determines which self-ascriptions are IEM, but when we consider introspection, it would seem to turn out that all experiences can ground IEM self-ascriptions. The case of introspection also reveals that Ismael's claim that IEM is due to transitions between representational media isn't quite right: after all, in introspection, the medium with which the state is attributed is the same as the one with which the state is represented. The articulation that issues from the transition from egocentric to allocentric formats of representation is just an instance of a more general phenomenon. All representation is, in the end, rooted in non-representational knowledge-how (see Ryle, Dennett). IEM arises when particularly secure knowledge-how is turned into knowlege-that. In the case of introspection, we make explicit the knowledge-how — evident in how we, for instance, integrate disparate thoughts — that all thoughts I experience are _my_ thoughts. The proposed account offers several advantages over existing theories. Unlike the simple view, it explains _why_ certain self-ascriptions involve no identity judgements. Unlike Recanati's account, I focus on background assumptions (rather than grounds), enabling me to distinguish between judgements that are *de facto* and logically IEM. I'm also not affected by arguments to the effect that certain experiences aren't selfless. Finally, I can explain why the token-reflexive rule applies, when it does, namely just when true background assumptions are made explicit. This solves an issue that metasemantic accounts see themselves confronted with, namely of explaining why certain kinds of experiences serve to fix the reference of 'I', whereas others do not.

Johan Heemskerk - Swampman's Last Stand: Who is in Danger and How to Avoid It

University of Warwick
j.heemskerk@warwick.ac.uk

In this paper, I attempt to settle some meta-theoretical questions concerning Swampman, hopefully clarifying the narrow conditions in which Swampman is a concern, while providing a simple suggestion to avoid Swampman in such conditions. In short, I conclude that Swampman is only a threat to those who can avoid it. I begin by outlining the classic Swampman thought experiment, in which an atom-for-atom copy of an existing individual is created miraculously by a bolt of lightning hitting a swamp. The classic question is whether Swampman has mental states, which is intended to bring out intuitions regarding physicalism. Swampman has since been deployed as a problem for teleosemantic accounts of intentionality. According to such accounts, the representational content of mental states is given by the function of the target representation. However, if that function is determined by selection history, as many teleosemanticists claim, Swampmen have no representational content since they have no selection history. The central claim of the paper is that Swampman is not a problem for teleosemantic accounts provided those accounts aim to provide contents consistent with 'ultimate' explanations. However, Swampman is a problem for accounts aimed at providing contents consistent with 'proximate' explanations. The ultimate/proximate distinction is due to Ernst Mayr [Mayr, 1961] and maps, roughly, to why- and how-questions, respectively. To clarify, contemporary teleosemanticists aim to produce accounts which output content attributions apt to feature in explanatory projects [Schulte, 2023]. Content, according to such authors, is invoked to explain behaviour. This should, I argue, include explaining Swampman's behaviour. I argue that teleosemantic theories of content in which content is intended to feature in an ultimate explanation of Swampman's behaviour are immune to the challenge that Swampman has no content. I invoke both Ruth Millikan [Millikan, 1987] and Nicholas Shea's [Shea, 2018] responses to Swampman, and conclude that they work well in the context of ultimate explanations. In short, there is just nothing to explain - there is no answer to the historical 'why' question about Swampman - this creature does things for no (historical) reason. However, if one aims for a view in which content features in a proximate explanation of Swampman's behaviour, Swampman is a problem. I demonstrate that neither Millikan nor Shea's response works in this case. There is an answer to the 'how' question: Swampman evinces the same behaviours as us and I argue that the same content-involving proximate explanation as we give to explain human behaviour should apply. If such a proximate theorist uses etiological functions, they are unable to explain Swampman's behaviour, because he has no content. Karen Neander [Neander, 2017] is such a theorist. This type of teleosemanticist is uniquely vulnerable to Swampman. However, for independent reasons, this type of teleosemanticist should, in contrast with their ultimate cousins, employ a non-etiological account of func- tion. As such, the only teleosemanticists who are at risk of Swampman have a simple strategy to avoid him. I end by spelling out how

using non-etiological functions solves the issue and why those seeking proximal explanations should employ them anyway.

References

Mayr, E. (1961). Cause and effect in biology: kinds of causes, predictability, and teleology are viewed by a practicing biologist. Science, 134(3489):1501–1506.

Millikan, R. G. (1987). Language, thought, and other biological categories: New foundations for realism. MIT press.

Neander, K. (2017). A Mark of the Mental: In Defense of Informational Teleosemantics. MIT Press.

Schulte, P. (2023). Mental Content. Cambridge University Press.

Shea, N. (2018). Representation in Cognitive Science. Oxford University Press.

Shadi Heidarifar - Beneath the Concept: The Quiet Norms That Make a Mind

Roseman University College of Medicine
Sheidarifar@roseman.edu

This paper argues that at least some central concepts in philosophy of mind, such as belief, memory, and agency, are not merely descriptive, but constitutively normative. That is, they do not neutrally track features of cognition; rather, they are structured in ways that determine what can even be registered as mental in the first place. These norms are not sociopolitical add-ons or methodological preferences. They are built into the architecture of the concepts themselves. The argument unfolds in three parts. First, I clarify the scope of the claim by distinguishing it from several common lines of inquiry: approaches that (1) center embodiment or affect, (2) historical reconstructions of theories of mind, (3) feminist or political reinterpretations of existing concepts, and (4) methodological debates over conceptual analysis versus naturalism. While each of these approaches contributes valuable insights, none address the internal structure of mind-concepts themselves. Second, by employing two canonical cases in philosophy of mind—Searle's Chinese Room and Clark & Chalmers' Extended Mind—I show that even highly influential frameworks depend on idealized templates of cognition. These templates privilege features like semantic immediacy, transparency, and fluency—not as empirical generalizations, but as built-in standards that govern what counts as genuine mental activity. Finally, I show why this conceptual claim matters by turning to fields like social epistemology, which inherit their conceptual vocabulary from philosophy of mind. While feminist epistemologists have rightly challenged whose epistemic agency is recognized, they have often retained the conceptual frame that structures such agency in the first place. The case of trauma reveals this tension with particular clarity. Epistemic agents shaped by dissociation, fractured memory, or disrupted self-trust are not simply marginalized; they often fail to be recognized as minded at all. This is not because of bias or social exclusion alone, but because the very concepts used to identify cognitive subjects are structured to omit them.

Mark Herman - Behavioral Ethics & Idealized Subjective Morality

Arkansas State University
mherman@astate.edu

Behavioral ethics is a practical offshoot (or applied subfield) of empirical moral psychology that aims to improve ethical decision-making and behavior, especially in the workplace. Ethical improvement requires a standard of ethical evaluation relative to which one can improve. The most commonly used standard in behavioral ethics is ethical evaluation "by one's own lights," especially "upon reflection" (Bazerman & Tenbrunsel; Biasucci & Prentice). However, this standard—this basis for identifying the "right" behavior (or decision, judgment, etc.)—is too vague. For instance, "upon reflection" is rarely meant as mere reflection, per se; instead, it usually involves some intervention, such as the provision of information (e.g., regarding an error or bias that was instantiated). However, involving interventions in the standard creates problems whenever different interventions would yield different evaluations or different allegedly "right" behaviors. Lacking a definitive "right" behavior—that is, lacking a singular, definitive standard of ethical evaluation—raises problems for the realization of ethical improvement. If behavioral ethics is to have a strong theoretical foundation, a more precise standard is needed. A standard of ethical evaluation described as "by one's own lights" constitutes a subjective ethical standard. "Upon reflection" constitutes a condition under which evaluations are privileged—that is, functionally speaking, an idealization condition. Idealization conditions for subjectivist standards—or more specifically, (neo-Humean, non-convergent) subjectivist theories—is thoroughly addressed in the literatures on practical reason and non-moral value theory (Brandt; Railton; Rosati; Sobel; Williams). Allowing simplification, such theories tie an agent's reasons or well-being to the desires of an idealized version of that agent (e.g., the desires the agent would have given perfect rationality and full information). Something like such idealization conditions should be incorporated into behavioral ethics' standard of ethical evaluation. Nonetheless, such idealization conditions would be incomplete because the idealization conditions upon which they would be based were not tailored for subjective moral theories. This deficit includes a lack of morality-specific idealization conditions (e.g., perhaps, maximal compassion). As such, morally relevant and morality-specific idealization conditions should be developed and incorporated as well. Connie Rosati's two-tier internalism provides useful resources for this task. Practical difficulties are addressed.

Nicolás Hinrichs - Epistemic Priors of Hyperscanning

Max Planck Institute for Human Cognitive and Brain Sciences
hinrichsn@cbs.mpg.de

Hyperscanning has invigorated social neuroscience by enabling the simultaneous recording of multiple brains during real-time interaction. However, its interpretation often relies on a set of epistemic priors (i.e., background assumptions about evidence, mechanism, and social ontology) that remain unexamined. This paper surfaces four such priors, each shaping the design and interpretation of hyperscanning studies: (1) second-person frameworks are reduced to solipsistic mental reasoning or subpersonal brain mechanisms; (2) diverse social phenomena are assumed to be uniformly underpinned by neural synchrony; (3) inter-brain coupling is modeled analogously to intra-brain dynamics; and (4) different statistical measures of synchrony are treated as interchangeable indicators of interaction quality. We argue that these priors are not neutral starting points but active constraints on the epistemic space hyperscanning can explore. To evaluate whether these priors should continue guiding hyperscanning research, we apply a modified version of Laudan's "context of pursuit" framework. Instead of asking whether results are true, we assess whether the underlying assumptions warrant further scientific investment based on criteria of novelty, methodological feasibility, relevance, and ethical proportionality. This pursuitworthiness framework provides a decision-making tool for researchers and funders seeking to align empirical methods with epistemic responsibility. We also contrast two ontological wagers: treating synchrony as a natural kind (i.e., stable across contexts and measurable with increasing precision) versus as a practical kind (i.e., context- sensitive, heuristically useful, but not mechanistically explanatory). On this view, synchrony may function not as a causal substrate of shared cognition, but as a semiotic attractor (i.e., a relational signal that stabilizes interaction by scaffolding a sense of shared agency). We refer to this as hyperanamnesis: the distributed enactment of co-simulated memory within interaction. This reframing shifts the scientific target from average synchrony levels to the dynamics of breakdown and repair, which amount to critical sites of co-regulation in therapeutic, educational, and collaborative settings. Our proposal is grounded in a methodological toolkit for hyperscanning researchers: a prior- audit table to link metrics to assumptions; a pre-registration prompt distinguishing synchrony as cause, cue, or consequence; and a checklist for assessing pursuitworthiness before experimental resources are allocated. We also advocate for dynamic metrics, such as curvature- based measures of network alignment, that can track structural shifts during relational dissonance and repair. In sum, we argue that hyperscanning must move beyond its fascination with visual coherence to earn its place in a science of participatory sense-making. This requires researchers to explicitly declare which epistemic priors they are amplifying and to remain open to revising them in light of new conceptual and methodological tools. Exposing and testing priors is not philosophical housekeeping: it is a prerequisite for making hyperscanning both epistemically accountable and clinically relevant.

Elena Holmgren - Group Flow as Spontaneous Collective Agency: A Phenomenological Account

University of British Columbia
elena.holmgren@gmail.com

At times, when a philosophical debate goes well, the conversation spontaneously takes on a life of its own. Participants experience their thoughts and actions as progressively intertwined with those of others, surprising each other by completing one another's sentences. Their sense of agency transforms as they experience a shared ownership of action wherein the group drives the action. This exemplifies collective agency, described in positive psychology as group flow: an intrinsically rewarding state of shared, spontaneous action.1 In this state, individuals experience enhanced efficacy through sustained interaction with others and a merger of self and others, leading to shared ownership of action.2 Group flow reveals an under-theorized dimension of the sense of agency: an experience in which the source of one's agency is a sense of oneness rather than autonomy.3 This is a dimension of agency in which the agent experiences herself not as the solitary author of her action, but rather as participating in a larger whole which is experienced as the source of her action. While there is a burgeoning literature on group flow in positive psychology, its first- person character remains under-theorized. This is problematic because the psychological description of group flow reveals, without clarifying, a puzzling feature of the experience of spontaneous agency: intensified efficacy coupled with diminished individual ownership and heightened shared ownership of action. But what must the agent be to experience intertwining with others as a source of intensified agency? And which experiential structures enable the alteration in agents' sense of authorship of their actions which gives rise to a sense of shared ownership? Drawing on Husserlian and Merleau-Pontyian phenomenology, I give a phenomenological account of group flow that dissolves this puzzle by disclosing and parsing the complex intentional interrelation between individual- and group-level agency that makes possible the emergence of a sense of shared ownership of action. At the individual level, group flow is made possible by what Merleau-Ponty called "chiasm." Chiasm refers to the intertwining of bodily subjectivity and the world such that the agent's bodily awareness of holding an object is simultaneously a bodily awareness of being held by the world. I discuss how chiasm is the source of agency since it enables the agent to experience the world as the ultimate context in which all possible agents and actions are situated. I then argue that chiasm explains how, in group flow, agents can experience an enhanced feeling of efficacy while experiencing a process larger than themselves as the driver of action.4 Moreover, at the group level, I introduce the notion of enfoldment to characterize the relation between individual bodily intentionality and the emergent group-level agency. Enfoldment is an emergent group-level intentional structure that binds together interacting individuals into a transient collective unity that has a qualitative character irreducible to that of the individual agents that enter into it. Ultimately, my phenomenological account of group flow explains how individual agents' sense of efficacy can be transformed by their sense of oneness with others.

Souichiro Honma - Deliberation and Source Freedom as Production

Hokkaido University
souichirohonma61@gmail.com

When we deliberate about which options to perform, it seems that the option we decide or choose to perform is caused by the deliberation itself. Suppose, for example, that we deliberate whether or not to see a movie, decide to see it, and then actually see it. In this case, our deliberative process seems to bring about seeing the movie. That is, deliberation seems to be considered as a kind of source freedom, i.e., the freedom to bring about action in virtue of ourselves. There is ongoing debate about whether we can be the source of our actions if determinism is true. Deery and Nahmias (2017) defend a compatibilist view, arguing that we can be the causal source of our actions even if determinism is true, especially by appealing to interventionist theories of causation (cf. Woodward, 2003). According to them, a free agent's decision is caused by Compatibilist Agentional Structure (CAS), and CAS must include a causal source. They argue that "X = x is the causal source of Y = y iff X bears the strongest causal invariance relation to Y among all the prior causal variables (including X) that bear such relationships to Y" (Deery and Nahmias, 2017, p.1263). On this account, deliberation qualifies as a compatibilist causal source if there is a stronger causal invariance relation between the decision or choice through deliberation based on CAS and the subsequent decided or chosen action than any other variables that can result in that action. The reason Deery and Nahmias require the strongest causal invariance relation for causal source is that deliberative activity must manifest equifinality: deliberative activity results in a particular (decided or chosen) consequence across various conditions. For example, they claim that Romeo's deci- sion to meet Juliet will result in his doing, regardless of many possible obstacles. In this presentation, I criticize their view that deliberations inherently exhibit equi- finality. I argue that there is a type of deliberation that leads to different outcomes when the input is different. That is, equifinality is too demanding to account for delib- erative causal sources. Rather than insisting on the strongest causal invariance relation, I suggest that productive causal relation is appropriate for compatibilist source free- dom in deliberation (cf. Hall, 2007). I also demonstrate how the productive causation theory can address problematic cases about that freedom, such as compatibility with determinism, manipulation, and pre-existing background conditions.

References

Deery, Ois´ın and Eddy Nahmias. (2017). "Defeating Manipulation Argument: Interventionist Causation and Compatibilist Sourcehood," Philosophical Studies, 174: 1255-76.

Hall, Ned. (2007). "Structural Equations and Causation," Philosophical Studies, 132: 109-36.

Woodward, James. (2003). Making Things Happen, Oxford University Press.

Amir Horowitz - Revisiting the Chinese Room

The Open University of Israel
amirho@openu.ac.il

Searle's Chinese-room argument attempts to refute "the strong AI thesis", according to which running the correct program constitutes thinking. The argument describes a person who manipulates symbols like a Chinese speaker without understanding Chinese, according to formal rules. He exemplifies the "right" syntax, but the Chinese symbols in his use lack meaning and intentionality. Leaning on this, Searle argues that running any program cannot constitute thinking because thinking involves semantics whereas programs are purely syntactical and syntax does not suffice for semantics. Many responses have been proposed for this argument, such as the one that the correct syntax does suffice for semantics, or that running the correct program plus adding causal connection to the environment (e,g, perception-like sensors) do. I will attack the argument from the opposite direction, showing that natural thinking is also syntax- based and does not involve semantics. I will present my version of semantic nihilism – intentional anti-realism – according to which nothing is and can be about anything: no-one can believe that p, says that p. thinks that p. etc., for any p, for the concept of intentionality is flawed. Content ascriptions only have practice-dependent rather than absolute truth-conditions; independently of a practice of content ascriptions, they cannot be true. This view (which I defend elsewhere) acknowledges the existence of mental states, notably of those called for "propositional attitudes", with the properties that we standardly attribute to them absent intentional ones. In particular, mental states have logico-syntactic structures, in a thick sense such that these structures essentially involve constants (so that "Tom is pretty" is different in structure from Barbara is clever). In virtue of having such structures, which maintain logical relations, mental states can function in cognitive processes. According to the argument from the explanatory and predictive success of content ascriptions, this success shows that mental states have content, but I will argue that postulating logico-syntactic properties (which are ascribed by content ascriptions) suffice to explain this success, so we need not postulate semantic properties. In fact, the argument vindicates there being mental states with logico-syntactic structures. If indeed we have reason to repudiate intentional properties, we thus get the strong AI thesis: artificial intelligence is basically similar to natural intelligence. I will discuss the idea that syntax presupposes semantics, which challenges this thesis. Specifically, I will address Crane's objection that we cannot determine which properties of a concrete system are relevant to its syntax but by reference to a semantic interpretation and reply that Crane is only entitled to claim that we in fact individuate syntax by content ascriptions (not content); and show – in response to Searle's argument – that syntax is extrinsic to physics.

Maria Luiza Iennaco - Towards a Biogenic Approach to Attention

University of São Paulo & University of Porto
marialuiza_ienaco@hotmail.com

There is a great deal of conceptual misunderstanding in the scientific idea of attention. This results from a profusion of definitions and operationalizations that are frequently incompatible and contradictory, which are caused by a reliance on a wide range of anthropocentric and intellectually motivated assumptions. While some propose abandoning the term altogether, a more nuanced approach might be possible: a shift towards biogenic approaches (BA). With that in mind, the purpose of this talk is to identify significant flaws with the concept of attention as employed in scientific research and to provide a possible remedy. Through an examination of the current state of attention research, we suggest that the difficulties stem not from the term itself but from its underlying assumptions. We shall briefly discuss these assumptions and argue for their replacement with biogenic ones, shifting from an intellectualist to an organismic viewpoint. Such a move, we believe, would alleviate the conceptual complexity around attention. Hommel and colleagues noticed that attention is often described as both a mechanism and a process, an explanans and an explanandum, something that activates neurons and the activation itself. This is problematic because treating different cognitive features and activations as one phenomenon may hinder clinical treatments and slow down our understanding of them. Because of this lack of clarity, some researchers have advocated for the elimination of the term "attention" from scientific language. However, such a move requires a broad area reform, and even if the term eventually disappears, the core challenges will remain under a new label. Rather than abandoning the term, we propose addressing the underlying assumptions, which often stem from subjective worldviews and folk interpretations of phenomena. Introducing a new set of assumptions enables researchers to formulate innovative hypotheses, design novel experiments, and interpret data in a new way. Traditional approaches begin with the premise that attention is a humane cognitive trait. This essentially intellectualist stance is not inherently problematic; nevertheless, when combined with reductionist neuroscientific operationalizations that split attention into distinct brain regions, the outcome is a dichotomic and frequently confrontational literature. After all, while diverse researchers evaluate equivalent facts via their subjective lenses, the intricacies of their attention accounts differ, resulting in the current problematic scenario. BAs, on the other hand, suggest studying cognition in its various forms through evolutionary lenses is the most fruitful approach. Lyon (2006) claimed that these characteristics can be better studied using self-organizing complex systems theories and autopoietic theories of cognition. We believe BAs provide a middle ground for addressing cognitive aspects, avoiding the extremes of intellectualism and reductionism. Attention can be defined theoretically and operationalized in novel manners, such as Hommel and colleagues' concept for visual selective attention. They employed phylogenetic refinement to reconstruct attention as a fundamental type of selectivity. We conclude that this shift of assumptions has the potential to alleviate some present conceptual problems, paving the way

for a more unified and scientifically based understanding of how organisms perceive and interact with their surroundings.

References

Anderson, B. (2011). There is no Such Thing as Attention. Frontiers in Psychology, https://doi.org/10.3389/fpsyg.2011.00246

Cisek, P. Resynthesizing behavior through phylogenetic refinement. Attention, Perception, & Psychophysics, v. 81, 2019, p. 2265–2287

Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J.-H., & Welsh, T. N. (2019). No one knows what attention is. Attention, Perception, & Psychophysics, 81(7), 2288–2303. https://doi.org/10.3758/s13414-019-01846-w

Lyon P. (2006). The biogenic approach to cognition. Cognitive processing, 7(1), 11–29. https://doi.org/10.1007/s10339-005-0016-8

Savvas Ioannou - Consciousness and the Ability to be Conscious

Independent researcher
savvasioa@hotmail.com

Physicalism is the metaphysical view that everything is physical or determined by the physical. Chalmers (1995) has argued that physicalism faces the hard problem of consciousness: it does not explain how consciousness arises from functional and structural physical properties. A suggested way to solve the hard problem of consciousness is to endorse panpsychism. Panpsychism is the metaphysical view that consciousness is ubiquitous. Some panpsychists have argued that the only way to close the explanatory gap between the phenomenal and the physical is to claim that macro-consciousness arises from micro-consciousness. However, Nagasawa (2021) argued that panpsychism faces the ingredient problem: it does not explain what microphenomenal properties are. We know nothing about any "phenomenal properties associated with the experiences of micromaterial objects. Micromaterial objects have no sensory apparatus, so obviously they do not have conscious experiences in the same way that we do; their experiences must be much more primitive than ours. They also lack neural systems, so they do not have phenomenal experiences that are linked to neural activity. We do not have a transparent grasp of microphenomenal properties, and we cannot even imagine what microphenomenal properties are like" (p. 39). I will suggest a different view about the mind that shows how the ingredient problem and the hard problem of consciousness can be solved. This gives us a good reason to prefer it. I will argue that we have a transparent grasp of a different mental property of micro- objects. I will suggest a view that I call 'panmentalism' because even though it does not postulate the existence of microphenomenal properties, it claims that there is still something mental at the micro-level. It claims that every micro-object possesses a specific micromental, non-phenomenal, property. I propose that we should conceive micro-objects differently. They possess the ability to be conscious, but they are not conscious because there is no mechanism to activate the ability to be conscious. Because of that, we do not have the mystery of what it is like to be a conscious micro-object. In the micro-level, no mechanism (e.g., a brain/neural mechanism) exists that can produce an experience with mental content (e.g., an observation). A certain kind of mechanism is needed to produce mental content. A certain kind of mechanism is required in order to produce feelings and receive and process input. If there is no such mechanism, the ability to be conscious cannot be activated, and phenomenal properties cannot be produced. There are conscious experiences when abilities to be conscious are bonded together by a mental bonding relation (that is, a bonding relation that bonds something mental). This relation can bond all the abilities to be conscious of the micro-objects and create a macro ability to be conscious. Also, this relation can activate this macro ability to be conscious, and as a result, macroscopic objects (e.g., humans) can be conscious. Conscious experiences occur, iff micro-abilities to be conscious bond through abilities- to-be-conscious bonding relations, create a macro-ability to be conscious, this macro-ability activates, and mental content is produced.

References

Chalmers, D. J. (1995). Facing Up to the Problem of Consciousness. Journal of Consciousness Studies, 2(3), 200-219.

Nagasawa, Y. (2021). A Panpsychist Dead End. In Aristotelian Society Supplementary Volume (Vol. 95, No. 1, pp. 25-50). Oxford University Press.

Ilir Isufi - A Critique of the Dissociation between Language, Culture, and Cognition in Research on Linguistic Relativity

University of Cincinnati
isufiir@mail.uc.edu

The linguistic relativity hypothesis posits that languages shape non-linguistic cognition, leading to divergent thought patterns among speakers of different languages (Casasanto, 2015). This idea has gained traction in various fields, including philosophy of mind (Kompa, 2024), enactivism (Rodríguez Jordá & Di Paolo, 2025), neuroscience, anthropology, linguistics, and cognitive psychology (Ibáñez et al., 2023; Reilly et al., 2024). However, I argue that this focus on linguistic relativity has led to a reductive conception of cognition, where linguistic elements are isolated from other cultural influences. This partitive conception of the mind is implausible, and I propose that a more promising approach is to investigate the mind's action in ecologically valid real-life conditions. To achieve this, philosophers of mind must adopt a truly interdisciplinary approach, incorporating insights from anthropology and cognitive linguistics, and recognizing the interdependence of quantitative and qualitative research methods. The internal logic of the linguistic relativity hypothesis presupposes two assumptions that embody a specific relationship between features of different languages, thought, and reality, which leads to three distinct methodological challenges for the empirical study of linguistic relativity (Lucy, 2014). First, the hypothesis posits that structural features of different languages dissect the world in different ways, making some features of it more and others less salient to one's experience. Thus, these linguistic features "embody a particular interpretation of reality" (Lucy, 2014, p. 18, original emphasis). Second, this linguistically mediated interpretation of reality affects patterns of thinking about reality in general by guiding cognitive tasks pertaining to memory, cognition, classification, or inference even in the absence of speaking. This setup entails a particular method for the empirical assessment of the linguistic relativity hypothesis: First, one needs to identify and characterize a contrast between two or more languages in which differing semantic structures might amount to different interpretations of one's experience and reality. Second, one must articulate and then establish potential cognitive consequences of the variance that manifest themselves in the absence of overt speech (or "speaking," as Lucy writes). Third, one must rule out alternative influences on the cognitive patterns to establish that they are due to linguistic structures (Lucy, 2014, 18). While research on linguistic relativity has been instrumental in elucidating the differential impact of language on cognition, it should not be overgeneralized. The current philosophical discourse that draws on these findings tends to overlook the nuanced nature of linguistic cognition, neglecting to consider the broader cultural context in which language is embedded, along with disciplines that study these contexts in a qualitative way. To fully comprehend the relationship between language and cognition, it is necessary to examine whether these effects persist in non- laboratory settings and, if so, to develop methodologies for their investigation. My presentation will address these

questions, drawing on debates at the intersection of philosophy of mind and anthropology (Ingold, 2012, 2020; Sinha, 2012).

Frederik T. Junker - Reasoning With Cognitive Maps as Inference

University of Copenhagen
ftjunker@gmail.com

I argue that the kinds of thinking we perform using cognitive maps meet minimal requirements of inference. In paradigm cases of inference, we are aware of the premises on which we base our conclusion as well as of some formal rule (e.g., modus ponens) that guides the inference (Boghossian, 2014). However, we often draw inferences under less demanding conditions. The conditions for inference can be relaxed along at least two dimensions: self- awareness and representational format. Inference can take place even when we are unaware of the premise states or unaware of the fact that we are drawing an inference (Quilty-Dunn & Mandelbaum, 2018; Siegel, 2019). Moreover, inferential transitions need not be governed by formal rules or have logical structure (Buckner, 2018; Shea, 2023). Nevertheless, certain constraints do need to be in place to differentiate inferential transitions from other types of mental transitions. First, inferential transitions are responsive to reasons. This contrasts with simple forms of associative transitions (e.g., from 'salt' to 'pepper) that are only responsive to forms of conditioning. Second, inferential transitions involve some appreciation or feeling that the conclusion follows from the premises, and this appreciation plays a role in causing the conclusion. Importantly, such appreciation need not involve richer forms of self-awareness, such as awareness that one is making a certain type of inference or a second-order belief that the conclusion follows from the premises. I argue that forms of thinking involving cognitive maps meet these minimal requirements of inference. Cognitive maps are structural representations that encode relations between spatial properties (e.g., locations) as well as more abstract, non-spatial properties (e.g., the competence and popularity of prospective collaborators) in two-dimensional maps. These representations enable spatial planning as well as generalizations of abstract structural knowledge to new situations (Behrens et al., 2018; Whittington et al., 2020). Map-based transitions yield standard outputs of inference: beliefs and intentions. However, they do not involve awareness of these transitions, nor do they have logical structure. Nevertheless, map-based transitions are responsive to reasons. Cognitive maps encode structural features of the world that often match the causal structure of the world. Since past knowledge is continuously integrated into cognitive maps, map-based transitions are responsive to evidence (Liu et al., 2019). Cognitive maps also help evaluate potential 2 actions and outcomes and identify optimal courses of action (Liu et al., 2021). They are thereby also responsive to reasons for action. Finally, there are ways of appreciating that a conclusion is supported by premise-states that require neither rich self-awareness nor logical structure. Feelings of reliability help restrict inferential transitions to those that have proven reliable in the past (Shea, 2023). Moreover, agents can learn to appreciate that, in a certain type of situation, a particular action will lead to a desired outcome—even without the capacity for second-order beliefs or logical inference. (Buckner, 2019). These modest ways of appreciating that a conclusion follows from premise-states plausibly accompany map-based transitions as well. Map-based transitions thereby meet minimal requirements of inference.

References

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. Neuron, 100(2), 490–509.

Boghossian, P. (2014). What is inference? Philosophical Studies, 169 (1):1-18.

Buckner, C. (2019). Rational Inference: The Lowest Bounds. Philosophy and Phenomenological Research, 98 (3):697-724.

Liu, Y., Dolan, R. J., Kurth-Nelson, Z., & Behrens, T. E. J. (2019). Human Replay Spontaneously Reorganizes Experience. Cell, 178(3), 640–652.e14.

Liu, Y., Mattar, M. G., Behrens, T. E. J., Daw, N. D., & Dolan, R. J. (2021). Experience replay is associated with efficient nonlocal learning. Science (New York, N.Y.), 372(6544), eabf1357.

Quilty-Dunn, J. & Mandelbaum, E. (2018). Inferential Transitions. Australasian Journal of Philosophy, 96 (3):532-547.

Shea, N. (2023). Metacognition of Inferential Transitions. The Journal of Philosophy.

Siegel, S. (2019). Inference Without Reckoning. In Magdalena Balcerak Jackson & Brendan Jackson (eds.), Reasoning: New Essays on Theoretical and Practical Thinking. Oxford University Press. pp. 15-31.

Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., & Behrens, T. E. J. (2020). The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. Cell, 183(5), 1249–1263.e23.

Jonida Kodra - Control in auditory verbal hallucinations

Osnabrück University
jonida.kodra@uni-osnabrueck.de

Auditory verbal hallucinations (AVH), defined as experiences of hearing voices in the absence of an appropriate external stimulus (APA 2022; David 2004; Slade & Bentall 1988), raise important clinical but also philosophical questions about the nature of hallucinations, mental agency, and delusional beliefs (Stephens & Graham 2000; Waters et al. 2006; Langland- Hassan 2018). A central and under-researched aspect of this experience is control (or lack thereof), i.e the extent to which, the specific manner in which, and aspects over which voice- hearers can exert control when experiencing AVH episodes. Philosophers take the lack of control to either indicate that AVHs are non-veridical perceptual experiences as opposed to other inner experiences (Farkas 2012; Knappik et al., 2022), or to explain the feeling of alienness that voice-hearers report (Waters et al. 2006; Wu 2012). I argue that the currently used and poorly understood notion of control cannot answer any such questions. Indeed, clinical definitions understand control in AVH either as the feeling or sense of control (Lingiardi & Williams 2017; David 2004), or as actual control (APA 2022; Slade & Bentall 1988), whereas reviews of the literature (Swyer & Powers 2020) highlight that the clinical literature understands control as actual control, that is, as "an ability to voluntarily influence voice- hearing experiences" (p. 2). In response, I suggest that, to properly understand control, we need to distinguish actual control from sense of control, and the actual agent (i.e the voice-hearer) from the imaginary one (i.e the voices heard). Specifically, I argue that the actual agent exerts different types of actual control. A sense of control correlates with the control that the actual agent attributes to the imaginary agent. This will help us better understand the notion of mental agency more generally. More specifically, concerning the who, every AVH episode involves at least two agents: one is an actual agent, the voice hearer himself, and the other is an imagined one, that is, the agent that the voice-hearers reports that the voices belong to. Concerning the what, the empirical literature points out that both agents seem to exercise some form of control. Voice- hearers exercise actual control, which is their ability to affect the AVH episode or some of its features or to refuse to act in accordance: I call this active and passive actual control. Voices, on the other hand, can be thought of as exercising control resulting in a sense of being controlled: this is experienced control. From the actual agent's perspective, this involves the feeling that the voices have the ability to affect the episode by either initiating or inhibiting it, by influencing voice-hearer's thoughts or actions, or by forcing the voice-hearer to act on commands. I call this active and passive lack of sense of control.

References

American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Baumeister, D., Sedgwick, O., Howes, O., & Peters, E. (2017). Auditory verbal hallucinations and continuum models of psychosis: A systematic review of the healthy voice-hearer literature. Clinical psychology review, 51, 125–141. https://doi.org/10.1016/j.cpr.2016.10.010

David, A. S. (2004). The cognitive neuropsychiatry of auditory verbal hallucinations: an overview. Cognitive neuropsychiatry, 9(1-2), 107–123. https://doi.org/10.1080/13546800344000183

Farkas, K. (2013). A Sense of Reality: in F. Macpherson and D. Platchias (Eds), Hallucination: Philosophy and Psychology (pp. 399–415). Cambridge MA: MIT Press.

Knappik, F., Bless, J., & Larøi, F. (2022). Confusions about 'inner' and 'outer' voices: conceptual problems in the study of auditory verbal hallucinations. Review of Philosophical Psychology, 13, 215–36. doi: 10.1007/s13164-021-00532-x

Langland-Hassan, P. (2018). From introspection to essence: The auditory nature of inner speech. In P. Langland-Hassan & A. Vicente (Eds.), Inner speech: New voices (pp. 78–104). Oxford University Press.

Lingiardi, V., & McWilliams, N. (Eds.). (2017). Psychodynamic diagnostic manual: PDM- 2 (2nd ed.). The Guilford Press.

Slade, P. D., & Bentall, R. P. (1988). Sensory deception: A scientific analysis of hallucination. Johns Hopkins University Press.

Stephens, G. L., & Graham, G. (2000). When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts. Cambridge MA: MIT Press.

Swyer, A., & Powers, A.R. (2020). Voluntary control of auditory hallucinations: phenomenology to therapeutic implications. Npj Schizophr 6, 19, https://doi.org/10.1038/s41537-020-0106-8

Matej Kohar - Lessons from Locusts: The Individuative Role of Representational Content in Computing Systems

Technische Universität Berlin
matej.kohar@tu-berlin.de

This paper critically examines Shagrir's modelling view of computation according to which representational contents play an indispensable individuative role for computing systems (Shagrir 2022). According to Shagrir, computational systems can be differentiated from non-computing ones by the fact that the physical evolution of a computing system g mirrors a relation R in a target domain in such a way that they share an abstract function f. At the same time, the computing system implements a formalism which has f as its abstract description, and the inputs and outputs of g represent the relata of R. I will use an illustrative case-study – locust looming-object avoidance behaviour as studied by Gabbiani and colleagues (Gabbiani, Krapp & Laurent 1999; Jones & Gabbiani 2012) to show inconsistencies between scientific practice in ascribing computations and Shagrir's account. A crucial component of the mechanism underlying this phenomenon is the lobula giant motion detector (LGMD) neuron, which, according to Gabbiani and colleagues multiplies angular velocity of an approaching stimulus with the negative angular size of the same. The LGMD also indirectly controls the evasive behaviour of the locust – jumping and flying away. In the literature, there are several options to semantically interpret the output of the computation, but none of them neatly fits Shagrir's mapping view. The primary research publications often view the LGMD activity as calculating a threshold angular size, which, once reached, prompts the locust to jump (e.g. Gabbiani, et al., 2002). But the view that LGMD activity mirrors the relation between input angular velocity and angular size, and output threshold angular size is inconsistent – using dimensional analysis, we can see that the result of multiplying these inputs should be in radians squared over seconds. But threshold angular size is in radians. A related proposal is that the LGMD computes the timing of the angular threshold, rather than the threshold itself (Papayannopoulos, Fresco & Shagrir, 2022). But here again, the proposed formalism applied to the semantic interpretation of the inputs lead to a quantity in radians squared over second, whereas time is given in seconds. In the model as given by Gabbiani, Krapp & Laurent (1999), the multiplication of angular velocity and angular size actually yields the neuron's firing rate. On the common assumption that radians are dimensionless, the dimensionality analysis goes through in this case. But, under Shagrir's modelling view, it would be absurd to consider the neuron as computing its own firing rate – this would mean that the firing rate is both the output vehicle as well as the semantically interpreted result of the computation. But while Shagrir countenances situations where the implemented formalism and the target domain coincide, he does not allow for the output vehicles and relata in the target domain to coincide, presumably because this would lead to pancomputationalism.

References

Gabbiani, F., Krapp, H.G. & Laurent, G. (1999). Computation of object approach, by a wide-field, motion-sensitive neuron. Journal of Neuroscience 19(3), 1122-1141.

Jones, P.W. & Gabbiani, F. (2012). Logarithmic compression of sensory signals within the dendritic tree of a collision-sensitive neuron. Journal of Neuroscience 32(14), 4923-4934.

Papayannopoulos, P., Fresco, N. & Shagrir, O. (2022). Computational indeterminacy and explanations in cognitive science. Biology and Philosophy 37, article no. 47.

Shagrir. O. (2022). The Nature of Physical Computation. Oxford University Press.

Wiktor Lachowski - Cognitive Phenomenal Unity: Bridging Perception, Thought and Learning

Institute of Philosophy and Sociology at the Polish Academy of Sciences
wik.lachowski@gmail.com

How do perceptual and conceptual contents come together in a single conscious stream? I propose a two-tier model of cognitive phenomenal unity that distinguishes (1) implicit cohesion within largely unconscious conceptual spaces and (2) explicit episodes in which concepts join ongoing sensory phenomenology.

● Implicit unity. Memories, scripts and self-relevant beliefs usually unfold automatically, supported by hippocampal and related "conceptual-navigation" circuits that are known to encode both physical and abstract spaces. These high- dimensional structures (Gärdenfors, 2004; Bottini & Doeller, 2020) operate offline to supply relevance judgments and action tendencies without entering phenomenal awareness.

● Explicit unity. When conceptual fragments must interact with perception—e.g., recognizing "that shape is a dog" or rehearsing a plan—they are reformatted into a phenomenally accessible medium. Sometimes this medium is quasi-sensory imagery, but amodal formats (number sense, verb semantics in congenitally blind individuals) show that embodiment is not mandatory (Machery, 2016). What matters is format compatibility: the re-expressed concept must share structural features with the perceptual scene it modulates.

● Learning and error correction. Conscious access is often required for flexible, goal- directed learning. Instrumental conditioning, the construction of novel stimulus-action rules, and fine-grained error detection all depend on comparing intended and actual outcomes in a shared phenomenal workspace (Schneider & Shiffrin, 1977; Skora et al., 2023). I model this comparison with projectors (Prentner, 2019): mappings that "flatten" abstract plans into the lower-dimensional format of immediate perception, making discrepancies transparent and corrigible.

● Flexibility without fragmentation. Conscious thought can jump rapidly between disparate ad hoc domains—horse → Rubik's Cube—yet a higher-order unity persists because each activated subset is translated into a common exchange format and linked by recurrent feedback loops. Cases such as aphantasia, where imagery is absent but spatial accuracy and low error rates are preserved (Bainbridge et al., 2021), illustrate how different representational modes can satisfy the same unifying function.

In sum, phenomenal unity extends beyond sensory binding to include the selective projection of conceptual structures into perception-compatible formats. This dynamic interface lets unconscious conceptual spaces inform conscious evaluation, supports rapid learning and error repair, and underwrites the coherent yet flexible character of human cognition and experience.

Urte Laukaityte - The Varieties of Philosophical Approaches: Analytic, Continental, and Synthetic

Aarhus University
urte.laukas@cas.au.dk

This paper argues that in order for philosophical theorising to play a more effective role in facilitating scientific research, it is helpful to elevate and popularise the status of synthetic philosophy as a distinct philosophical approach. At present, there are two widely acknowledged strands of contemporary philosophy that are often conceived of as separate philosophical traditions associated with different methods, standards, and goals – analytic philosophy and continental philosophy. Humphries (1999) diagnoses the divide as being due to a difference in philosophical style. Levy (2023) suggests the source of incommensurable standards between the analytic and the continental traditions is that the former comprises a Kuhnian (1962) paradigm, whereas the latter does not. As a paradigm, analytic philosophy tends to nudge its practitioners towards working on a relatively narrower set of puzzles designed within the tradition and away from practical questions of wider relevance, becoming 'less and less relevant to the kinds of pressing questions that often drive people to philosophy in the 6irst place' (Levy, 2023: 299). Continental philosophy has its own issues in this regard – for one, it is not really interested in 'abutting, or seeking to ground or support or complement, the world of modern scientific research' (Humphries, 1999: 263), which limits the kind of impact it might have. Various theorists currently view the goals of the discipline differently, but Levy (2023: 303) expresses hope there may be 'a new way of doing philosophy that would combine the strengths of both' strands. I want to put forward a contender for one such possible 'middle way' vis-à-vis the potential of philosophical work to contribute to empirical projects. It is a tradition of philosophical methodology that is at present institutionally subsumed under analytic philosophy and rarely acknowledged as a separate approach, but arguably should be. Various terms have been used to refer to it with some family resemblance among them. them. Of these, I will discuss three specific recent formulations, which should be enough to outline its general contours: naturalistic philosophy (Thagard, 2009), synthetic philosophy (Lewens, 2014; Schliesser, 2019; 2024), and philosophy in science (Pradeu et al., 2024), although for pragmatic reasons I opt for 'synthetic philosophy' as the apt unifying label. In brief, the approach I aim to delineate would combine the original affinity of analytic philosophy to the methods and questions of science as well as its focus on clarity with – in some ways – the more daring, wide-ranging, big-picture interdisciplinary spirit of the continental tradition. There is a proud tradition of empirically informed philosophical theorising meant to serve as a contribution towards broader empirical goals, such as William James pushing the sciences of the mind forward. I see high-level empirically oriented systematising as a key role that synthetic philosophers could play in a range of fields, but especially the ones where the science is not yet advanced enough to be able to move forward significantly just by following the course it has laid out so far. I discuss the advantages and disadvantages of advocating for an explicitly distinct term to refer to the seemingly heterogenous ways to practice empirically oriented philosophy, concluding that the former outweigh the latter.

Đorđe Lazarević - Artificial Intelligence and the Challenge of Creative Spontaneity

Faculty of Philosophy at the University of Niš, Serbia
djordje.lazarevic@filfak.ni.ac.rs

Recent advances in "creative technology"—particularly in AI-generated images, poetry, and scientific ideas—have raised questions about whether machines might one day surpass human creative abilities, or whether they already have. To assess such questions meaningfully, it is crucial to first establish a clear conceptual framework for creativity itself. In this talk, I pursue two main objectives. First, I offer an analysis of creativity by distinguishing three key components: product, process, and agent. A creative product must be both novel and valuable; a creative process must involve cognitive mechanisms, operate spontaneously (understood as the absence of prior knowledge about the end product or the method of creation), and include recognition of the product's value; and a creative agent must possess consciousness, intentionality, and a dispositional intention to create. I argue that a working ontology of creativity can be based on the relationships of these three components, each constituting a necessary condition for creativity. Within this framework, the creative product is ontologically independent— its novelty and value exist objectively, apart from the creator—though they are causally dependent on a particular disjunction of cognitive mechanisms employed by the agent which are, only because they are causal conditions for the product, constitutive of a creative process. This definition allows us to evaluate whether AI satisfies the basic conditions for creativity—and, if not, whether there are principled reasons it cannot do so. The question of whether AI can be an agent is a broader metaphysical issue. For the purposes of this talk, I adopt the still-controversial assumption that AI can possess consciousness, intentionality, and creative intention. I also assume that, despite AI having produced novel and valuable content for years—that is, creative products—merely passing a "creative" version of the Turing test is insufficient for genuine creativity (Boden 2004). Instead, I focus on a more pressing question: can a machine be spontaneous? Since spontaneity is a necessary condition of a creative cognitive process, a machine's presumed lack of spontaneity would suggest a lack of true creativity, even if it had a conscious desire to create. One common objection, attributed to Ada Lovelace, holds that machines lack spontaneity because they always follow instructions and therefore lack intrinsic motivation and creative freedom (Lovelace 1953). I challenge this argument in two ways. First, both thought experiments and empirical studies suggest that human creative actions can be determined and predicted, which undermines the claim that spontaneity requires indeterminacy or unpredictability (Haugeland 1985). A stronger version of the argument holds that AI agents are always causally dependent on humans, whereas humans—at least in some cases—initiate creative processes independently. I argue that the key distinction lies not in metaphysical structure but in epistemic stance: we attribute creativity to humans partly because we lack full knowledge of the causal mechanisms underlying our own cognition, while we assume complete transparency in machine operations. This asymmetry is unjustified. If we encountered highly creative machines without any prior knowledge of their internal mechanisms, we might readily describe their behavior in terms of "inspiration" or

"emotion." Conversely, a fully mechanistic understanding of human cognition—as might one day be achieved through an ideal science of the brain—would reveal that our experience of autonomy in initiating creative processes is grounded solely in the epistemic position of the agent, not in any metaphysical spontaneity. A further implication of this view is that a superintelligent AI—if it possessed complete knowledge of its own creative architecture—could not be creative, since it would lack the epistemic opacity that makes spontaneity possible in human cognition. Even if such an AI produced extraordinarily valuable artifacts and ideas, it would still fall short of genuine creativity. Finally, a crucial empirical question remains: do the mechanisms by which AI generates creative products share the same mathematical and functional structure as those of the human brain? According to the framework proposed here, only in that case could we uncontroversially attribute creativity to AI. Until then, we remain far from having compelling reasons to do so—even if we provisionally grant basic agential properties such as consciousness and intentionality.

References

Lovelace, A. A. (1953). Sketch of the analytical engine invented by Charles Babbage, Esq. (L. F. Menabrea, Author). In B. V. Bowden (Ed.), Faster than thought: A symposium on digital computing machines (pp. 282–310). Pitman Publishing.

Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433–460.

Boden, M. A. (2004). The creative mind: Myths and mechanisms (2nd ed.). Routledge.

Haugeland, J. (1985). Artificial intelligence: The very idea. MIT Press.

Juliana Faccio Lima - 'We'-beliefs and Group Action

Azim Premji University
jfacciolima@gmail.com

Perry [1979] has famously argued that we need indexical beliefs to explain and rationalize intentional actions of individuals. I climb up a tree upon seeing a bear running towards me because I believe that "I am about to be attacked by a bear". The beliefs that "The inattentive hiker is about to be attacked by a bear" or "John [not the author's real name] is about to be attacked by a bear" are also about me. But if I don't believe that "I am the inattentive hiker" or "I am Joh"–that is, if I don't hold an indexical belief–I won't be motivated to act in a way that saves myself. Interestingly, we can construct similar cases with groups and conclude that group indexical beliefs are necessary to explain a group's behavior. Suppose I am hiking with a group called the 'Bear Food hiking team'. Upon seeing a bear approaching, we huddle together to appear larger –a common strategy used to prevent bear attacks. We are motivated to act this way because we believe that "we are about to be attacked by a bear". The beliefs that "The inattentive hiking group is about to be attacked by a bear" or "Bear Food Hiking Team is about to be attacked by a bear" are also about us. However, unless we believe that "we are the inattentive hiking group" or that "we are Bear Food Hiking Team", we won't be motivated to act in a way to save ourselves. Given the similarities of the individual and group cases, it's reasonable to expect, and prima facie desirable, a unified account of indexical beliefs that explains why they are necessary to motivate both individual and group actions. In this paper, I argue that a popular account of indexical beliefs, Perry's Self-Locating View, can't accommodate group indexical beliefs. This account holds that indexical beliefs reveal an object's location relative to the person who holds the belief: they carry information about where, when, or who the object of the belief is in relation to the believer's immediate environment. Thus, what is special about my belief that "I am about to be attacked by a bear" is that it locates me, the believer, in relation to the object of the belief. If we were to expand the Self-Locating View to we-beliefs, it would entail that they locate the believer, in relation to the object of the belief, in this case, a group. I explore different ways of understanding what it means to locate a group in relation to the believer and argue that each one has counterexamples. I then present an alternative approach, on which indexical beliefs are practical ways of thinking of objects. On this view, indexical beliefs require a know-how to interact with the object, rather than knowledge of where, who, or when the object of the belief is in relation to the believer. This way of understanding indexical beliefs uniformly explain both individual and group actions. In the we-belief case, I do not need to know where the group is relative to me, but just how to act as a group to perform a group action. So, Perry is right that a special kind of belief motivates action. But if I'm right, group actions and indexical beliefs suggest that he is wrong about what indexical beliefs are.

References

Anscombe, G. E. M. (1975). The first person. Clarendon Press.

Cappelen, H. and Dever, J. (2013). The Inessential Indexical: On the Philosophical Insignificance ofPerspective and the First Person. Oxford University Press.

Casta˜neda, H.-N. (1968). On the logic of attributions of self-knowledge to others. The Journal of Philosophy, 65(15):439–456.

Evans, G. (1982). The Varieties of Reference. Clarendon Press Oxford.

Garc´ıa-Carpintero, M. and Torre, S. (2016). About Oneself: De Se Thought and Communication. Oxford University Press.

Kaplan, D. (1989a). Afterthoughts. pages 54–57. Kaplan, D. (1989b). Demonstratives. In Almog, J., Perry, J., and Wettstein, H., editors, Demonstratives. Themes from Kaplan, pages 481–563. Oxford: Oxford University Press.

Magidor, O. (2015). The myth of the de se. Philosophical Perspectives, 29(1):249–283. Perry, J. (1979). The problem of the essential indexical. Noˆus, pages 3–21.

Perry, J. (1990 [1977]). Frege on demonstratives. In Yourgrau, P., editor, Demonstratives, pages 50–70. Oxford University Press.

Perry, J. (1998a). Indexicals, contexts and unarticulated constituents. In Proceedings of the 1995 CSLI-Amsterdam Logic, Language and Computation Conference, pages 1–16. Stanford University: CSLI Publications.

Perry, J. (1998b). Myself and i. Philosophie in synthetischer Absicht, Stuttgart, pages 83–103.

Prosser, S. and Recanati, F. (2012). Immunity to error through misidentification: New essays. Cambridge University Press.

Shoemaker, S. (1968). Self-reference and self-awareness. The Journal of Philosophy, 65(19):555–567.

Shoemaker, S. (1970). Persons and their pasts. American Philosophical Quarterly, 7(4):269–285.

Yen-Tung Lee - Phenomenal Content in Mixed Reality

Academia Sinica, Taiwan
ylee734@uwo.ca

Due to the increasing significance of virtual reality (VR) technology to our lives, philosophers are paying more attention to its philosophical foundations. This presentation addresses the epistemology of perceptual experiences in VR (hereafter VR experiences). As many believe that many VR experiences are veridical (aka virtual veridicalists; e.g., Chalmers, 2017, 2022; Lee, 2024), an immediate question follows: What does VR experience represent? Or, more precisely, what is the content of VR experience? This presentation aims to answer this question by looking for a detailed picture of phenomenal content that coheres with the advancement of VR technology. This presentation assesses Fregean representationalism, seeing whether it serves as a theory of phenomenal content that coheres with virtual veridicalism. According to this theory, the phenomenal content of perceptual experiences of a certain type represents the property that normally causes experiences of that type (Chalmers, 2004, 2006; Thompson, 2009). I choose to assess Fregean representationalism because, firstly, a primary defense of virtual veridicalism is based on ideas similar to the Fregean spirit, and secondly, this theory appears to be more flexible in accommodating the desiderata imposed by virtual veridicalism upon the desired theory of content. In this presentation, I argue that Fregean representationalism can provide a coherent picture of phenomenal content with virtual veridicalism only to a very limited extent. Specifically, I argue that this theory coheres with virtual veridicalism only in perfect virtual environments (i.e., environments in which virtual items screen off all non-virtual items) but not in mixed environments (i.e., environments in which virtual items screen off some but not all non-virtual entities). In a nutshell, Fregean representationalism predicts that VR experiences in a perfect virtual environment are typically veridical. For instance, it predicts that reddish experience in a perfect virtual environment does not represent physical redness but virtual redness because virtual redness is the normal cause of reddish experience in such an environment. Plus, since most token reddish VR experiences are instantiated when being caused by virtual redness, they are usually veridical. However, Fregean representationalism predicts that VR experiences in a mixed environment are typically falsidical, which contradicts virtual veridicalism. This is because, in a mixed environment, most token VR experiences are caused by a mixture of a virtual property and a physical property. According to Fregean representationalism, reddish experience in a mixed environment represents either physical redness or virtual redness, but not both. Either way, token reddish experiences cannot be veridical, for their cause is often a mixture of physical and virtual redness. Thus, reddish experience in a mixed environment is typically falsidical because the property the experience represents its object as having does not correspond to the property its object really instantiates. To generalize, Fregean representationalism does not have sufficient resources to accommodate veridical perception in mixed environments, so it does not satisfy as an appropriate theory of phenomenal content that coheres with virtual veridicalism.

Antonio Lizzadri - Hybrid Minds Between Knowledge and Information

Università Cattolica di Milano
antonio.lizzadri@unicatt.it

This presentation aims to reflect on the epistemological relevance of the increasingly deep intertwining between human minds and technological artifacts. In fact, mind's hybridisation requires us to rethink the traditional Cartesian view on knowledge, according to which cognitive processes are fully internal to the "bounderies of skill and skull" (Clarks and Chalmers 1998: 7). On the other hand, the extension of the mind to environmental/technological supports requires us to also define how far the extension goes. Does “extension” mean a full identification of the mind to its artifacts? In this regard, the contemporary debate on 4E-Cognition (Newen, De Bruin, and Gallagher 2018; Carter 2018) has developed two fundamental epistemological tendences: some claim that it just happens that with the large-scale deployment of AI-powered applications our mental states lies beyond the body and that cognitive processes are essentially embedded in the physical/technological environment that represents their substantial realization (Carter & Kallestrup 2018; Pritchard 2018); others claim that - being first and foremost embodied in the human body - the “extension”of the mind cannot be conceived in terms of a mere identity of mental states and artificial technological artifacts and that human cognition has irreducible specific feautures related to the bodily mediation, such as subjectivity, intentionality, agency, emotionality, intersubjectivity, sociality, etc. (Gallagher 2017; Hutto & Myin 2017; Smithies 2018; Aizawa 2018; Adam 2018). In this presentation I will try to support this second perspective, by highlighting the conceptual differences between the notions of knowledge and information. More precisely, I will argue that the reductionist interpretation of the extended mind thesis - that identifies mental states to technological artifacts - can be supported only reducing knowledge to mere information, where information lacks the fundamental epistemological properties that characterize knowledge: truth and justification.

References

Adams, F. (2018). “Extended Knowledge” In: Newen, A.; De Bruin, L.; Gallagher, S. (Eds.). (2018), pp. 79-89.

Aizawa, K. (2018). “Extended Cognition, Trust and Glue, and Knowledge”. In: Newen, A.; De Bruin, L.; Gallagher, S. (Eds.). (2018), pp. 64-78.

Clark, A.; Chalmers, D. (1998). “The Extended Mind”. Analysis, 58(1): 7-19.

Carter, A.; Kallestrup, J. (2018). “Extended Circularity: A New Puzzle for Extended Cognition”. In: Newen, A.; De Bruin, L.; Gallagher, S. (Eds.). (2018), pp. 44-63.

Gallagher, S. (2017). Enactivist Interventions. Rethinking the Mind. Oxford: Oxford University Press.

Hutto, D.D.; Myin, E. (2017). Evolving Enactivism. Basic Minds Meet Content. Cambridge (MA): The MIT Press.

Maiese, M. (2017). "Can the Mind Be Embodied, Enactive, Affective, and Extended?". Phenomenology and the Cognitive Sciences, 17(2): 343-361. Newen, A.; De Bruin, L.; Gallagher, S. (2018). The Oxford Handbook of 4E Cognition. Oxford: Oxford University Press.

Pritchard, D. (2018). "Extended Epistemology". In: Newen, A.; De Bruin, L.; Gallagher, S. (Eds.). (2018), pp. 90-104.

Smithies, D. (2018). "Access Internalism and the Extended Mind". In: Newen, A.; De Bruin, L.; Gallagher, S. (Eds.). (2018), pp. 17-41.

Louis Loock - The Cognitive Footprint of AI Usage

Osnabrück University
lloock@uni-osnabrueck.de

What is the future of human cognition in a world of much more intelligent tools? Current efforts of engineering ever more capable digital tools evoke philosophical concerns directly about the nature and ethics of artificial intelligence. But it might be more relevant to first ask how the daily usage of AI tools could impact our own cognitive abilities, and what this would reveal about the nature and ethics of our own natural intelligence. Prioritizing this question seems advisable, also because it could consequently redefine our immediate views on AI, too. The present investigation from the field of situated cognition (Newen et al., 2018; Robbins & Aydede, 2008) advances a new perspective on our cognitive relations with external tools. Specifically, the very influential sub-debate on extended cognition (Clark, 1997; Clark & Chalmers, 1998) exclusively viewed our tool relations as cooperative and internally beneficial (Clark, 2008). But what if our current tool practices are rather detrimental for the standing of our internal cognition, especially if some form of cognitive replacement ensues (Gerlich, 2025; León-Domínguez, 2024; Paglieri, 2024)? Extended cognition is inherently unable to capture this possibility (Aagaard, 2021). Hence, another situated contender might be called for. Consider how many of us are nowadays personally inclined, and perhaps even structurally incentivized, to utilize digital tools that can solve our cognitive tasks for us – or at least produce comparable outputs. This is realized via advanced technologies that can somewhat obtain parts of our cognitive skills which we would usually exert internally for those tasks. Our interaction strategies with those tools may then slowly decrease our own cognitive engagements and responsibilities. This might ultimately render us "extracted cognizers". The hypothesis of extracted cognition states that we naturally, or at least habitually, desire external tools that solve our cognitive tasks independent of us, namely by making or letting them capture, mimic, and eventually replace those of our cognitive skills we would otherwise employ and train internally. Three questions shall lead us to this hypothesis: First, how do we make intelligent tools? Second, how do we use intelligent tools? And third, how do we thereby become extracted cognizers? The provided answers form three crucial steps toward the future of human cognition. First, making intelligent tools requires them to capture our cognitive skills via abstract templates that we provide to them in different forms. Second, using intelligent tools is often driven by our tendency toward the most economic task strategy, and this often results in situations where our tools shall just imitate our cognitive skills that we previously provided to them. Third, becoming an extracted cognizer requires that those tools ultimately replace our cognitive skills as we no longer exercise them ourselves, and this result can be characterized by independent criteria from the theory, practice, ethics, and science of cognition. If these three answers are convincing, the future of human cognition might indeed best be described by the hypothesis of extracted cognition: We seek tools that solve our cognitive tasks like us, for us, and yet without us.

References

Aagaard, J. (2021). 4E Cognition and the Dogma of Harmony. Philosophical Psychology, 34(2), 165–181. https://doi.org/10.1080/09515089.2020.1845640

Clark, A. (1997). Being There: Putting Brain, Body, and World Together Again. MIT Press. https://doi.org/10.7551/mitpress/1552.001.0001

Clark, A. (2008). Supersizing the Mind: Embodiment, Action, and Cognitive Extension. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195333213.001.0001

Clark, A., & Chalmers, D. J. (1998). The Extended Mind. Analysis, 58(1), 7–19. https://doi.org/10.1093/analys/58.1.7

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Societies, 15(1), 6. https://doi.org/10.3390/soc15010006

León-Domínguez, U. (2024). Potential cognitive risks of generative transformer-based AI chatbots on higher order executive functions. Neuropsychology, 38(4), 293–308. https://doi.org/10.1037/neu0000948

Newen, A., de Bruin, L., & Gallagher, S. (Eds.). (2018). The Oxford Handbook of 4E Cognition. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.001.0001

Paglieri, F. (2024). Expropriated Minds: On Some Practical Problems of Generative AI, Beyond Our Cognitive Illusions. Philosophy & Technology, 37(55), 1–30. https://doi.org/10.1007/s13347-024-00743-x

Robbins, P., & Aydede, M. (Eds.). (2008). The Cambridge Handbook of Situated Cognition. Cambridge University Press. https://doi.org/10.1017/CBO9780511816826

Valentina Martinis - What's 'in' your mind? Marks of the mental in early analytic philosophy

University of Liège
vmartinis@uliege.be

What does it mean that something is mental, as opposed to non-mental or physical? And what entities are such? Prompted by the rise of empirical psychology, early analytic philosophers had to confront themselves with these questions. This paper reconstructs and assesses the different theories of the 'mark' of the mental suggested by the fathers of analytic philosophy G.E. Moore and B. Russell, and by some of their less-known contemporaries with whom they extensively engaged. We shall see that three main criteria emerged. The first is the Brentanian thesis that all and only mental facts are intentionally directed at things. The second is the view that all and only mental facts are 'enjoyed' or 'lived through', a technical term for non-objectifying awareness. The third and final view is that all and only mental facts are those that exist as long as they are suitably related to a subject.

Darryl Mathieson - Degrees of Minimal Self-Consciousness

Australian National University
Darryl.Mathieson@outlook.com

A longstanding view in the philosophy of mind is that being in a phenomenally conscious state means that there is something it is like for me to be in that state. According to many researchers, this for-me-ness remains constant across all our conscious states, makes its own distinctive contribution to our ongoing phenomenology, and is a minimal kind of self-consciousness. Much of the current debate on for- me-ness centres around its existence and pervasiveness, but a relatively underexplored area concerns its structure. One interesting question in this vein is whether for-me-ness comes in degrees. Many authors across the philosophical, psychological, and psychiatric literature assume that it does, especially when discussing disorders of self-consciousness. But no one has offered any sustained argument for the view. This has resulted in a growing number of critics denying the plausibility and coherence of degrees of for-me-ness and dismissing any explanations of pathological cases that invoke it. Here, I develop a positive account of degrees of for- me-ness. I distinguish between two leading interpretations of what for- me-ness amounts to—which I call the awareness and perspectival views— and argue that both can accommodate for-me-ness coming in degrees. I use this argument to disarm a number of extant criticisms, clarify the psychopathology literature, and open up novel avenues for further research.

Arata Matsuda - The Meta-Meta-Problem of Consciousness

Hokkaido University
matsuda.arata.philosophy@gmail.com

Building on the debate about the hard problem of consciousness, Chalmers (2018) goes meta and introduces the meta-problem of consciousness. In this paper, I repeat this "meta-problem shift" (see Kelley, 2024, for an articulation of the meta-problem shift as a "philosophical move"). That is, building on the debate about the meta-problem of consciousness, I go further meta and introduce the meta- meta-problem of consciousness. Whereas the hard problem asks why and how consciousness arises from physical processes, the meta-problem asks why we think that consciousness poses the hard problem. In other words, the meta- problem asks why we have what Chalmers (2018) calls problem intuitions, namely intuitions that consciousness poses the hard problem.

(The Hard Problem) Why and how does consciousness arise from physical processes?

(The Meta-Problem) Why do we think that consciousness poses the hard problem?

One appeal of the meta-problem is that it is more inclusive than the hard problem: On the one hand, the hard problem is approachable only by realism about consciousness, but not by illusionism about consciousness. This is because illusionism denies that consciousness poses the hard problem in the first place (see, e.g., Dennett, 1991; Frankish, 2016). On the other hand, the meta-problem is approachable by both realism and illusionism. This is because illusionism only denies that consciousness poses the hard problem, but not that we think that consciousness poses the hard problem (see, e.g., Dennett, 2019; Frankish, 2019). Whereas the meta-problem asks why we think that consciousness poses the hard problem, the meta-meta-problem problematizes the "we" in the meta-problem and asks why some philosophers of consciousness think that "we" think that consciousness poses the hard problem. In other words, the meta-meta-problem asks why some philosophers of consciousness think that "we" have problem intuitions.

(The Meta-Problem) Why do we think that consciousness poses the hard problem?

(The Meta-Meta-Problem) Why do some philosophers of consciousness think that "we" think that consciousness poses the hard problem?

One appeal of the meta-meta-problem is that it is more inclusive than the meta-problem: On the one hand, the meta-problem is approachable only by what I call universalism about problem intuitions, but not by what I call localism about problem intuitions. This is because localism denies that "we" think that consciousness poses the hard problem in the first place (see, e.g., Díaz, 2021; Fischer & Sytsma, 2021). On the other hand, the meta-meta-problem is approachable by both universalism and localism. This is because localism only denies that "we" think that consciousness poses the hard problem, but not that some philosophers of

consciousness think that "we" think that consciousness poses the hard problem. In this paper, I will explain what the meta-meta-problem is, consider possible solutions to the meta-meta-problem, and argue for the importance of the meta-meta-problem. To clarify, my aim throughout these discussions is not to conclusively solve the meta-meta-problem, because it requires extensive empirical research and thus lies beyond the scope of this paper. Rather, my aim is to introduce this important problem and set the stage for future research.

References

Chalmers, D. J. (2018). The Meta-Problem of Consciousness. Journal of Consciousness Studies, 25(9–10), 6–61.

Dennett, D. C. (1991). Consciousness Explained. Little, Brown.

Dennett, D. C. (2019). Welcome to Strong Illusionism. Journal of Consciousness Studies, 26(9–10), 48–58.

Díaz, R. (2021). Do People Think Consciousness Poses the Hard Problem? Empirical Evidence on the Meta-Problem of Consciousness. Journal of Consciousness Studies, 28(3–4), 55–75.

Fischer, E., & Sytsma, J. (2021). Zombie Intuitions. Cognition, 215, 104807.

Frankish, K. (2016). Illusionism as a Theory of Consciousness. Journal of Consciousness Studies, 23(11–12), 11–39.

Frankish, K. (2019). The Meta-Problem is the Problem of Consciousness. Journal of Consciousness Studies, 26(9–10), 83–94.

Kelley, D. (2024). Philosophy Moves and Meta-Moves. Analysis, anae047.

Maria Matuszkiewicz - Intention recognition without inference. A case of moderate intentionalism about demonstratives

Jagiellonian University, Kraków
maria.j.matuszkiewicz@gmail.com

This paper defends a moderate intentionalism about demonstrative reference, according to which speaker's intentions play an important – though not exhaustive – metasemantic role. The argument is indirect and conditional. First, I argue that experimental psychology and psycholinguistics support an intentionalist-friendly view of linguistic communication, one in which social cognition plays an important role. Second, I claim that if this picture is accurate, it justifies adopting a moderate intentionalist semantics for demonstrative pronouns. I begin by examining the intentionalist-friendly picture, which emphasizes mind-reading in linguistic interpretation, and contrast it with a view that treats communication as (almost) solely a matter of decoding conventional meaning. Support for the former come from the socio- pragmatic account of language acquisition and interpretation. Importantly, however, this account need not posit a fully developed theory of mind or the capacity for representing Gricean communication intentions. Instead, it may appeal to low-level mind-reading mechanisms identified by experimental psychologists. If these latter findings are correct, interpreting demonstrative utterances does not require sophisticated inferential reasoning. Rather, hearers can recognize intentions using basic mind- reading skills. I conclude by showing how this philosophical perspective informs the philosophical debate between conventionalism and intentionalism about demonstrative reference.

Oliver Milne - Automimesis: operationalising the structure of human desire to produce humanlike motivations in AI systems

Independent scholar, Galway
oliver.f.milne@gmail.com

In this talk, I sketch a design for a particular kind of AI agent, and make the case that this design successfully operationalises the structure of human desire and selfhood. The fundamental notion is that human beings are able to act as though we have consistent, long-term goals and values because we rationally reconstruct our actions and experiences as if we had such goals. This allows us to act in accordance with those imputed goals, producing a self-stabilising feedback loop that 'realises' the reconstructed self-image – the self and its desires manifesting themselves as a hyperstition. This process of interpreting ourselves and acting on that interpretation I call 'automimesis'. The notion extends René Girard's mimetic theory of desire, in which we learn what to desire by imitating or copying the desires of others, to explain how we can acquire and maintain coherent desires at all, in a manner reminiscent of Dennett's 'intentional stance' theory (Dennett, 1989; Palaver, 2013). The great advantage of this theory is that it lends itself to operationalisation as a machine learning problem. An AI agent built to this design has three chief components: a record of the system's actions and sensory inputs, a set of 'model histories' that serve to narrativise that record and indicate the system's next action, and a reinforcement learning component to produce and update those model histories. The histories are accounts of agents and their goal-directed actions, rendered in a formal planning language such as MA-PDDL (Cooper et al., 2016). Each history includes an account of the present and future as well as the past, and in each history, one agent is designated the 'self'. The 'self' agent's predicted record of actions and experiences is matched up against the system's real record to determine the score the history receives for the reinforcement learning system. Crucially, included in the histories' simulated experience records are notes of when the 'self' agent achieves or fails a goal. The corresponding notes in the real experience record are added in two ways: either the system writes them in because the 'most plausible' histories say they should be there, the same way it chooses actions; or because 'instinct' systems – simple heuristics to, for instance, recognise smiling faces – put them there. This means the system has to reconstruct why it might have achieved a goal at the point the 'instinct' system put a 'goal achieved' marker in the record. This is the crux: the 'instinct' system's heuristics are not necessarily the same as the motives the system imputes to itself for 'feeling good' (or 'feeling bad') in a situation. Thus simple heuristics, such as 'see smile → good', can via the system's own self-overinterpretation produce complex 'moral' motivations such as 'I want to help others'. This, I argue, is both a more plausible description of human desire and a better mechanic rendition of it than the reward function theory, in which 'desire' refers to the reward function of a reinforcement learning mechanism in the brain (Morillo, 1990; Schroeder, 2004).

References

Cooper, M., Herzig, A., Maffre, F., Maris, F., Régnier, P., 2016. A simple account of multi-agent epistemic planning. Presented at the 22nd European Conference on Artificial Intelligence (ECAI 2016), The Hague, Netherlands, pp. 193–201.

Dennett, D.C., 1989. The Intentional Stance. The MIT Press.

Morillo, C., 1990. The reward event and motivation. Journal of Philosophy 87, 169–186.

Palaver, W., 2013. René Girard's Mimetic Theory. Michigan State University Press.

Schroeder, T., 2004. Three Faces of Desire. Oxford University Press.

Katsunori Miyahara & Norihiko Kamaya - Testate amoeba and the body-schema: An embodied approach to minimal agency

Hokkaido University
kmiyahara88@gmail.com

Studies in basal or minimal cognition strive to develop a non-anthropocentric conception of cognition, intelligence, and agency, by studying these capacities across diverse non-human organisms (Brancazio et al 2019; Lyon et al 2021; Lyon & Cheng 2023). Much of this research locates the source of these capacities in cognitive capacities implemented by internal information processing mechanisms. For example, the toolkit of basal cognitive capacities (Lyon et al 2020) contains capacities such as orienting response, discrimination, and decision making, which are considered to be implemented by various biomechanical mechanisms. To broaden the perspective, this presentation investigates agency in unicellular organisms, drawing on insights from phenomenology and embodied cognition concerning the "body-schema". We propose that this concept can be fruitfully applied to explain the adaptive behaviors of unicellular creatures without neural systems, such as testate amoeba (single-celled protists that construct external shells). The body-schema is a sensorimotor system that governs bodily movement in a non- conscious manner (Gallagher 2005; Merleau-Ponty 2012). It adjusts the body holistically to ensure flexible goal-directed engagement with the environment. It can functionally extend beyond the biological body to incorporate parts of the environment, thereby enabling complex forms of adaptive behavior such as tool use. We argue that testate amoeba exhibits adaptive behaviors that are best explained by the possession of such a system. For example, these organisms use their pseudopodia to gather environmental particles, such as silica and diatom fragments, for shell construction. Furthermore, some species utilize parts of their shells as a "weapon" (Dumack et al 2024) to immobilize and tear prey. These complex interactions, involving the seamless integration of external materials into the behavioral repertoire, suggest that their agency is organized by a body-schema that extend beyond their purely biological body. Therefore, while the body-schema is associated with the neural system in human cognition, we argue that non-neural, unicellular organisms can possess and shape their agency based on a body-schema. Applying the body-schema concept to testate amoeba offers an important corrective to the literature on basal or minimal cognition: it introduces conceptual tools from embodied cognition, offering an alternative to internalist information-processing accounts of intelligence, cognition, and agency in simple life forms that currently dominate the field. It also suggests that the evolutionary history of complex goal-directed behaviors, including primitive forms of tool use, is far deeper than is previously understood. Furthermore, this work opens new research avenues regarding the developmental dynamics of the body-schema in organisms. Given that the body-schema develops over time through history of interaction in humans and animals with complex neural systems, our proposal prompts questions about whether similar developmental processes occur in unicellular organisms.

References

Brancazio, N., Segundo-Ortin, M., & McGivern, P. (2020). Approaching minimal cognition: introduction to the special issue. Adaptive behavior, 28(6), 401–405.

Dumack, K., Lara, E., Duckert, C., Ermolaeva, E., Siemensma, F., Singer, D., ... & Mitchell, E. A. (2024). It's time to consider the Arcellinida shell as a weapon. European Journal of Protistology, 92, 126051.

Gallagher, S. (2005). How the body shapes the mind. Oxford University Press.

Lyon, P., Keijzer, F., Arendt, D., & Levin, M. (2021). Reframing cognition: getting down to biological basics. Philosophical Transactions of the Royal Society B, 376(1820), 20190750.

Lyon, P., & Cheng, K. (2023). Basal cognition: Shifting the center of gravity (again). Animal Cognition, 26(6), 1743–1750.

Merleau-Ponty, M. (2012). Phenomenology of perception (D. Landes, Trans.). Routledge

Gaia Mizzon - Telling Our Dreams. How the Philosophical Debate on Dream Experience Relies on Assumptions About the Narrative Character of Dreaming and Dream Reports

Monash University
gaia.mizzon@monash.edu

The concept of narrative is widely used and prominent in the literature on dreams across humanities (Rupprecht, 1987; Schrage-Früh, 2012, 2016a, 2016b), psychology (Alder, 2016; Friedman, 2024) and cognitive science (Cipolli et al., 1998; Cipolli & Poli, 1992; Foulkes, 2014; Montangero & Cavallero, 2015; Pace-Schott, 2013). However, this term and its associated conceptual aspects are often undefined and insufficiently examined. Although a few previous works have analysed the putative narrative character of dreams, borrowing definitions from narratology, literature and semiotics (Kilroe, 2000; Montangero, 2012; States, 1990; Walsh, 2010), there has been virtually no investigation of how preconceptions about the similarity between fictional narratives and retrospective dream reports have shaped the philosophical debate on dreams and dreaming. Here, we address this gap. We argue that positions on the relationship between dreams and narratives are not theoretically neutral, as they are frequently linked to the common assumption that the narrative structure of dream reports reveals—transparently or after interpretation—something about the underlying narrative structure of dream experiences themselves. Specifically, we contend that dream reports and experiences are often metonymically assimilated to fictional narratives in that they are understood in terms of features and devices typically applied to works of fiction. To investigate this tendency, we focus on the categories of authorship and composition and discuss examples where background assumptions about the similarity of waking narratives and dreams have influenced the ontological characterisation of dream experiences as well as epistemological claims about the relationship between dreams and dream reports. On the one hand, authorship addresses the issue of dream origin, mirroring narratological concerns about the role of an author in shaping a fictional work. We suggest that this concept resonates with the philosophical debate upon which mechanisms are responsible for generating and guiding the flow of dream experiences. On the other hand, composition pertains to the product of dream generation and its relationship to memory retrieval and reporting. We posit that this category has nurtured arguments about the epistemology and ontology of dreams—i.e., their status as experiences susceptible or not to direct scientific investigation and their relation to dream reports. Together, the categories of authorship and composition offer an analytical lens to disentangle central controversies about the nature of dreaming as a complex process that unfolds throughout different stages—from generation and experience during sleep to retrieval and reporting upon awakening. We conclude by showing how a critical discussion of these key categories might shed light on unanswered questions in the philosophy of dreaming and promote a more theoretically grounded understanding of the narrative character of dreams.

References

Alder, M.-L. (2016). Dream-Telling Differences in Psychotherapy: The Dream as an Allusion. Language and Psychoanalysis, 5(2), Article 2. https://doi.org/10.7565/landp.v5i2.1558

Cipolli, C., Bolzani, R., & Tuozzi, G. (1998). Story-like organization of dream experience in different periods of REM sleep. Journal of Sleep Research, 7(1), 13–19. https://doi.org/10.1046/j.1365-2869.1998.00094.x

Cipolli, C., & Poli, D. (1992). Story Structure in Verbal Reports of Mental Sleep Experience After Awakening in REM Sleep. Sleep, 15(2), 133–142. https://doi.org/10.1093/sleep/15.2.133

Foulkes, D. (2014). Dreaming: A Cognitive-psychological Analysis. Routledge. https://doi.org/10.4324/9781315831220

Friedman, R. (2024). Dreamtelling: The value of dreaming and sharing dreams in relations: Part I: From collective preoccupations to shared dreaming. Group Analysis, 57(3), 367–389. https://doi.org/10.1177/05333164241241263

Kilroe, P. A. (2000). The Dream as Text, The Dream as Narrative. Dreaming, 10(3), 125–137. https://doi.org/10.1023/A:1009456906277

Montangero, J. (2012). Dreams are narrative simulations of autobiographical episodes, not stories or scripts: A review. Dreaming, 22(3), 157–172. https://doi.org/10.1037/a0028978

Montangero, J., & Cavallero, C. (2015). What renders dreams more or less narrative? A microstructural study of REM and Stage 2 dream reports upon morning awakening. https://doi.org/10.11588/ijodr.2015.2.19774

Pace-Schott, E. F. (2013). Dreaming as a story-telling instinct. Frontiers in Psychology, 4. https://doi.org/10.3389/fpsyg.2013.00159

Rupprecht, C. S. (1987). Dreams and Literature: A reader's Guide. In Sleep and Dreams. Routledge.

Schrage-Früh, M. (2012). "The Roots of Art Are in the Dream": Dreams, Literature and Evolution. In Telling Stories / Geschichten erzählen (pp. 156–171). De Gruyter. https://www.degruyterbrill.com/document/doi/10.1515/9783110268683.156/pdf?licenseType=restricted

Schrage-Früh, M. (2016a). Dreaming Fictions, Writing Dreams. In M. Schrage-Früh (Ed.), Philosophy, Dreaming and the Literary Imagination (pp. 131–205). Springer International Publishing. https://doi.org/10.1007/978-3-319-40724-1_4

Schrage-Früh, M. (2016b). Philosophy, Dreaming and the Literary Imagination. Springer International Publishing. https://doi.org/10.1007/978-3-319-40724-1

States, B. O. (1990). Dreaming and Storytelling. The Hudson Review, 43(1), 21–37. https://doi.org/10.2307/3852328

Walsh, R. (2010). 6. Dreaming and Narrative Theory. In Toward a Cognitive Theory of Narrative Acts (pp. 141–158). University of Texas Press. https://www.degruyterbrill.com/document/doi/10.7560/721579-007/html

Riccardo Mona - Dennett's pragmatism as a form of realism in philosophy of mind

Pontifical Lateran University
riccardomona@gmail.com

Daniel Dennett has been an influential thinker in different areas of philosophy, from philosophy of mind to philosophy of biology. A commitment to evolutionary approaches was a unifying thread, and it is perhaps his Darwinian opposition to essentialism that led him to eschew labels or affiliations with definite schools of thought, especially on ontological issues in philosophy of mind ("Are beliefs real?"). This attitude of his can be connected to his interest in having philosophy deal with the advancement of empirical science, rather than developing more and more refined theoretical definitions. I argue that there are, however, more cogent reasons to try and establish a pragmatistic interpretation of Dennett's work. I will present three elements which give plausibility to such reading. Firstly, Dennett doesn't quite refuse to take part in ontological disputes, but he tries to frame them in a context which considers the historical as well "strategical" underpinnings of a given theoretical position. Secondly, we can detect in Dennett the influence of the pragmatism of Robert Brandom, especially in the way Dennett sees the emergence of the intentional stance from the practical reasoning within a community of speakers. Thirdly, and perhaps most importantly, I will show that Dennett takes a pragmatic approach when defending his notion of "real patterns"; in particular, I will develop the already noticed parallelism between Dennett's patterns and the ideas of Charles S. Peirce. In the last part of the paper, I will contrast two kinds of pragmatistic readings of Dennett's work. I will argue that Dennett's view on ontological issues in philosophy of mind can shows strong parallelisms to a realist, Peircean version of pragmatism, rather than to the constructivist version which is attributed to Dennett by Bjorn Ramberg.

References

Aames J (2019). Patternhood and Generality. European Journal of Pragmatism and American Philosophy 11(2)

Brandom R B (2008), Between saying and doing: Towards an analytic pragmatism, Oxford University Press, Oxford.

Dennett D C (1991a). Consciousness explained. Little, Brown and Co., New York.

Dennett D C (1991b). Real Patterns. The Journal of Philosophy, 88(1), 27-51.

Dennett D C (1987). The intentional stance. MIT Press, Boston.

Dennett D C (2013a). The Evolution of Reasons. In Bashour B, and Muller H (eds), Contemporary Philosophical Naturalism and Its Implications. Routledge, 13-47.

Dennett D C (2013b). Intuition pumps and other tools for thinking. W. W. Norton & Company, New York.

Ramberg B (1999). Dennett's Pragmatism. Revue Internationale de Philosophie 53 (207), 61-86.

Peirce C S (1931) The Collected Papers of Charles Sanders Peirce. Harvard University Press, Cambridge

Niccolò Nanni - Perceiving Intermodal Functional Relations

University of Turin
niccolo.nanni1999@gmail.com

That we perceive relations is somewhat uncontroversial. Thus, for example, we might see that the red cup on the table is located to the left of the bowl of fruit, where being to the left of is a simple spatial relational property. Similarly, we might hear that an A minor chord occurs just before the C major chord, where occurring before is a simple temporal relational property. Recently, it has been argued that we can perceive some of these simpler relational properties as relating stimuli that are perceived through different sense modalities. For example, one might perceive the cup one sees to be to the left of the apple one touches, or the flash of light one sees to occur just before the C major chord one hears. Call the latter kind of relations intermodal relations. Recent empirical evidence from the psychology of vision also suggests that, in addition to simpler spatial and temporal relations, human beings might also be able to visually perceive what has been labelled functional relations. These are more sophisticated relations that involve the categorical perception of certain stimuli as falling under specific functional roles. Some of the functional relations that have been taken to be perceptual are static—such as when one object is perceived to contain, support, or surround another—while others unfold dynamically over time, such as when one object is seen to push, pull, or deform another. This presentation explores the largely overlooked question of whether human beings can perceive intermodal functional relations: functional relations that hold between stimuli perceived through different sensory modalities. The presentation will be divided into three sections. The first section will be dedicated to introducing the notions of intermodal relation and of functional relation by drawing from the existing philosophical and empirical literature. Then, in the second section, I will argue on both phenomenological and empirical grounds that there are multiple plausible examples of perceived intermodal functional relations. For example, I will contend that we can experience sounds we hear to be contained within objects we see, objects we see to fit with objects we touch, and odors we smell to emanate from objects we see. Lastly, in the third section, I will argue that whether intermodal functional relations are perceived might have consequences on broader debates, such as the debate on the nature of olfactory objects.

References

Hafri, A., & Firestone, C. (2021). The perception of relations. Trends in Cognitive Sciences, 25(6), 475–492.

O'Callaghan, Casey (2014). Not all perceptual experience is modality specific. In Dustin Stokes, Mohan Matthen & Stephen Biggs, Perception and Its Modalities. New York, NY: Oxford University Press. pp. 133-165.

O'Callaghan, Casey (2017). Grades of Multisensory Awareness. Mind and Language 32 (2):155-181.

O'Callaghan, Casey (2020). Multisensory evidence. Philosophical Issues 30 (1):238-256.

Gareth Norman - Who's Afraid of Zeugma?

Massachusetts Institute of Technology
garethn@mit.edu

Sentences like “Penfold drank a beer and threw it into the recycling bin” pose a problem for contemporary semantic theory. One the one hand, it looks like a single thing – the witness of “a beer” - is ascribed two predicates (being drank by Penfold and being thrown in the recycling). On the other hand, this sentence’s natural interpretation is that two different things are drank and thrown in the recycling (the beer-contents and container respectively). This is the puzzle of co-predication. Extant solutions to the puzzle face a serious problem. Consider “Penfold drank a beer, which made a ringing noise as he put it down”. This sentence looks exactly like a co-predication in every linguistically relevant respect: both are grammatical, both are semantically well-formed, both ascribe one predicate to some liquid beer, and another to a beer container. Yet the latter is often reported to be unacceptable. These sentences are often dubbed “zeugma”. In response, a number of theorists have proposed a number of psycholinguistic criteria for differentiating co-predications from their zeugmatic counterparts. I argue for two claims. First, these psycholinguistic approaches are mistaken: zeugmatic constructions can be rendered acceptable by manipulating features that should be psycholinguistically irrelevant from the point of view of the parser, the question under discussion of the context, clause type, and the sentential connective used. Second, I argue we should embrace a fully pragmatic criterion for differentiating zeugma from co-predications. In particular, what separates co- predications from zeugma will be whether we can readily construct an explanation for why a hypothetical speaker would be interested in informing us of its content.

Zeynep Oğuzman - What If Psychosis Speaks Back? Co-experiencing the Lived World Through Second-Person Psychiatry

Middle East Technical University
oguzmanzeynep@gmail.com

Traditional third-person, symptom-centric approaches to psychiatric diagnosis and treatment have been increasingly critiqued for their inability to adequately engage with the lived, first-person phenomenology of psychotic experience. Rooted in a reductionist biomedical paradigm, such models tend to marginalize the epistemic and existential dimensions of psychosis, thereby failing to capture the full scope of the disorder's ontological and intersubjective complexities (Schilbach, 2016; Fuchs & Dalpane, 2022). Attempts to correct this imbalance via first-person or biopsychosocial frameworks, while well-intentioned, often remain theoretically fragmented or epistemologically dissonant with prevailing clinical science, limiting their practical efficacy in diagnosis and care formulation. Consequently, a growing corpus of scholarship has called for a reconceptualization of psychopathological inquiry—one that transcends the dualism of objectifying observation and isolated subjectivity. In response to this paradigmatic tension, a second-person framework, grounded in the work of Michael Pauen (2012), is proposed as a novel epistemological and clinical orientation for understanding psychosis. The second-person perspective introduces a relational ontology of mind wherein selfhood and symptomatology are co-constituted through intersubjective engagement. This approach integrates insights from phenomenological psychopathology, cognitive neuroscience, and dialogical philosophy to offer a neuro-psycho-phenomenological model of care. It privileges empathic resonance and shared intentionality as foundational mechanisms of clinical understanding, shifting the therapeutic stance from detached interpretation to participatory co-experience. Within this model, the therapeutic field is constituted not solely by the dyad of clinician and patient but extends to a broader relational matrix including family members and interdisciplinary contributors. To support and extend the patient's epistemic access—that is, their capacity to articulate and make sense of their own experiential world—the model incorporates technological mediators such as Virtual Reality (VR), alongside the clinician's cultivated use of internal experiential resources (e.g., imaginative attunement, narrative empathy). The inclusion of family members through structured psychoeducation anchors the therapeutic process within the patient's ecological context, facilitating the continuity of care beyond clinical boundaries. Furthermore, dialogical engagement with other professionals fosters a reflective space in which epistemic humility, mutual learning, and interdisciplinary synthesis become integral to clinical reasoning. This model thus inaugurates a paradigmatic shift: away from biologically reductive psychiatry and toward a relational, meaning-sensitive, and scientifically grounded form of practice. Crucially, the notion of the "interpersonal" is reconceived not merely as face-to-face interaction but as encompassing digitally mediated encounters as well—such as avatar-based therapies, AI-enhanced dialogical agents, and virtual therapeutic environments. These modalities offer novel platforms for

co-regulating affect, co-constructing narratives, and scaffolding self-experience in patients for whom conventional clinical interaction may be insufficient or inaccessible. Ultimately, the second-person paradigm, especially as extended through immersive and relational technologies, offers a dynamic, ethically attuned, and epistemologically robust framework for the treatment of psychosis—one that centers not only on symptom remission but on the reconstitution of subjectivity and lived coherence.

Michael Omoge - Ingress Controllers: Reconciling Encapsulation, Penetration, and Perceptual Learning

University of Alberta
omoge@ualberta.ca

Theories of perception are often torn between two seemingly incompatible theses: informational encapsulation and cognitive penetration. Encapsulation holds that perceptual systems operate over a fixed, proprietary body of data isolated from the agent's beliefs, desires, and intentions (Fodor 1983). Cognitive penetration, by contrast, posits that cognitive states systematically influence perceptual operations (Pylyshyn 1999). Traditionally, these views are treated as mutually exclusive. Yet Sam Clarke (2020) proposes a reconciliation: perception consists of interconnected subsystems whose internal operations remain encapsulated, while cognitive influence may occur between subsystems. While Clarke's architectural solution avoids a zero-sum conflict between encapsulation and penetration, it comes at a significant cost. His model implies that the proprietary data in perceptual systems is effectively sealed after early developmental stages. This rigidity, I argue, is incompatible with perceptual learning—the well-documented phenomenon by which perceptual systems undergo long-term changes as a result of experience (Gibson 1963; Lyons 2016). From the expert ornithologist discerning subtle species differences to novices acquiring new visual skills, perceptual learning requires that proprietary data be open to updates, not permanently fixed. In this paper, I develop a new architectural account that preserves Clarke's non-exclusivity between encapsulation and penetration while also restoring perceptual learning to its rightful place. I propose that perceptual systems employ "Ingress Controllers", which are functional mechanisms stationed at the interfaces between perceptual subsystems. These controllers perform two critical roles. First, they maintain encapsulation by blocking the entry of cognitive states lacking perceptual markers, thus preventing beliefs or desires from contaminating the proprietary data. Second, they permit the gradual admittance of new perceptual information when repeated experience establishes robust correlations between sensory input and perceptual assumptions. This dual function allows perceptual learning to occur without breaching the boundaries of informational encapsulation. Drawing on evidence from perceptual psychology and cognitive neuroscience, I demonstrate how this architecture accounts for both the stability and plasticity of perception. I further show how mechanisms such as visual short-term working memory (VSTWM) facilitate the slow, energy-conserving process by which perceptual systems update assumptions while avoiding catastrophic forgetting. My model explains how experts can integrate new perceptual knowledge (e.g., differentiating bird species or visual patterns) without losing previously acquired skills. Ultimately, this account challenges the prevailing assumption that protecting encapsulation requires developmental closure. By introducing Ingress Controllers as selective, adaptive gatekeepers, I offer a framework that unites modular stability with learning-driven flexibility—an architecture capable of supporting both the epistemic reliability of perception and its capacity for growth.

References

Brogaard, B., and Chomanski, B. (2015). Cognitive Penetrability and High-Level Properties in Perception: Unrelated Phenomena? Pacific Philosophical Quarterly 96(4): 469–486.

Clarke, S. (2021). Cognitive Penetration and Informational Encapsulation: Have We Been Failing the Module? Philosophical Studies 178(8): 2599–2620.

Connolly, K. (2019). Perceptual Learning. New York: Oxford University Press.

Fodor, J. (1983). The Modularity of Mind. Cambridge, MA: MIT Press. Gibson, E. J. (1963). Perceptual Learning. Annual Review of Psychology 14: 29–56.

Gibson, J. J. (1979). The Ecological Approach to Visual Perception. Boston, MA: Houghton Mifflin.

Kosslyn, S. (1994). Image and Brain. Oxford: Oxford University Press).

Kozma, P., Kovacs, I., and Benedek, G. (2001). Normal and Abnormal Development of Visual Functions in Children. Acta Biologica Szegediensis 45(1-4): 23–42.

Lyons, J. (2009). Perception and Basic Beliefs. New York: Oxford University Press.

Lyons, J. (2016). Unconscious Evidence. Philosophical Issues 26(1): 243–262.

Pitts, J. B. (2019). Conservation Laws and the Philosophy of Mind: Opening the Black Box, Finding a Mirror. Philosophia. doi:10.1007/s11406-019-00102-7.

Postle, B., and Pasternak, T. (2010). Short-term and Working Memory. Encyclopedia of Neuroscience. 783-789.

Pylyshyn, Z. (1999). Is Vision Continuous with Cognition? The Case for Cognitive Impenetrability of Visual Perception. Behavioral and Brain Sciences 22(3): 341–365.

Raftopoulos, A. (2001). Is Perception Informationally Encapsulated? The Issue of the Theory-Ladenness of Perception. Cognitive Science 25(3): 423– 451.

Rescorla, M. (2015). Bayesian Perceptual Psychology. In Mohan Matthen (Ed.), The Oxford Handbook of the Philosophy of Perception (pp. 694–716). Oxford: Oxford University Press.

Stokes, D., and Bergeron, V. (2015). Modular Architectures and Informational Encapsulation: A Dilemma. European Journal for Philosophy of Science 5(3): 315– 338.

Zimmermann, A., Carvalho, K. M. M. de, Atihe, C., Zimmermann, S. M. V., and Ribeiro, V. L. de M. (2019). Visual Development in Children Aged 0 to 6 Years. Arquivos Brasileiros De Oftalmologia 82(3): 173–175.

Andrea Onofri - Asymmetric Linguistic Conventions

Universidad Autónoma de San Luis Potosí
andonofri@gmail.com

This paper discusses and criticizes an important aspect of the Lewisian theory of conventions (Lewis 1969, 1975; Williams 2019). According to the Lewisian theory, linguistic conventions are symmetric in two different respects. First, if a communicative exchange follows the conventions in question, then the speaker expresses a belief and the hearer forms the same type of attitude (a belief) in response to the speaker's utterance. Second, the speaker and the hearer's respective beliefs have the same content. In this paper, I criticize both of these claims: I argue that the conventional response to a declarative sentence need not be a belief, and even if it is, it need not have the same content as the belief that's expressed by the speaker. I begin by discussing the content of the mental states that are involved in a conventional linguistic exchange. Lewis (1969) famously discussed a signaling game where the players can settle into different optimal equilibria or signaling systems. While the signaling systems of the Lewis signaling game do seem to suggest that the mental states of the interlocutors must have the same content, I will show that this is the exception rather than the rule. Drawing on recent work by Skyrms (2010) and others, I'll discuss more complex signaling games and show that these games suggest a rather different picture. I'll then move on to the second part of my argument, which concerns the type of mental state that's formed by the hearer in a conventional linguistic exchange. I'll point at several cases where the receiver doesn't form a belief while still conforming to the relevant linguistic conventions. These cases involve a variety of cognitive responses, such as entertaining a proposition, wondering and inquiring about its truth, imagining that it is true, or assuming its truth for the sake of argument. All these possible responses to an assertion are an integral part of conventional language use, yet they don't require believing that the assertion is true. If these arguments are correct, we must decide what to make of all those linguistic exchanges that are "asymmetric" for one or both of the above reasons – exchanges in which the mental states of the interlocutors are of a different type and/or have a different content. Can these exchanges still conform to a linguistic convention? I will defend an affirmative answer, and argue that there are asymmetric linguistic conventions: linguistic conventions which don't require conversational participants to hold the same type of mental state, and don't require the mental states in question to have the same content.

References

Lewis, D. (1969). Convention: A philosophical study. Wiley.

Lewis, D. (1975). "Languages and language". In Martinich and Sosa (eds.), The Philosophy of Language (pp. 682-700), sixth ed., Oxford University Press.

Skyrms, B. (2010). Signals: Evolution, learning, and information. Oxford University Press.

Williams, J. R. G. (2019), The Metaphysics of Representation, Oxford: Oxford University Press.

Laura Oppi - What is the Perky effect?

University of Copenhagen
lao@hum.ku.dk

The so-called 'Perky effect' plays a major role in shaping numerous contemporary philosophical debates on perception, some of which entail the Sense of Reality (Deroy 2013; Ghijsen 2014; Hopkins 2012; Hopkins 2013; Kind 2001; Macpherson 2012; Nanay 2010, 2012, 2015; Siegel, S., & Silins, N. 2015; Teng 2016, 2018, 2023). This talk questions the claim that the Perky experiment shows perception can occur withouta Sense of Reality. Among Perky's experiment published in 1910, one specific set gained great notoriety. Perky and her team instructed participants to focus on a point on a surface while imagining specific objects, such as a leaf, book, banana, or tomato. The participants were unaware that the surface was actually a ground glass screen onto which 'just noticeable' images of these same objects were being projected from behind. While the participants thought they were imagining the requested objects, what they described as imaginings actually reflected the projected patterns. For example, when a vertical yellow banana was projected, all participants reported imagining a banana that was in an upright position (Perky 1910, p.432). This experiment is widely used in the philosophical literature. According to some contemporary interpretations, the Perky effect shows that perceptual experiences can lack the feeling or sense that the perceived object exists or is actually there. And this, it is argued, has important implications for the epistemology debate (Ghijsen 2014; Lu Teng 2023, 2024). I will call 'the Sense of Reality' the feeling or sense that the perceived object exists or is actually there. Although the Perky experiment has been a cornerstone and its conclusions are often accepted in philosophy, many other authors stressed some of its criticalities (Hopkins 2012, 2013; Langland- Hassan 2014). In what follows, I will argue that, despite its historical and philosophical significance, there is insufficient evidence in the Perky experiment for it to support the claim that perception can occur without a Sense of Reality, and that there are alternative explanations equally consistent with the results without implying a lack of Sense of Reality. To do that, I will focus on the notion of 'perceptual threshold', and I will suggest that inattentional blindness could offer a more parsimonious explanation of Perky's results.

Emma Otterski - Mindreading, power, and social status

emma.otterski@gmail.com

This paper argues that one's status in society results in differences in how and when we attribute mental states to others (mindread) and that salient, transitory status in an interaction can have similar effects. Attention to social dynamics, on both a societal and interpersonal level, offers as-yet unacknowledged insights into the mechanisms underlying mindreading. Until recently, if philosophical accounts of mindreading considered socio-cultural factors, it was to highlight cross-cultural differences in platitudinous or external theories of folk psychology while maintaining no difference at the level of mindreading. While not unwarranted given robust findings showing broad convergence in the development of false- belief understanding, it is odd that this would be the end of the story. False-belief tests are methodological tools designed to measure a specific ability. For one, they do not tell us when we engage in mindreading, which is a broader set of competencies than false-belief understanding. Socio-economic status (SES) has been shown as a variable affecting the developmental trajectory of mental-state understanding, with data indicating that children with low SES pass the elicited false-belief test later than those with high SES (e.g. Holmes, Black and Miller, 1996; Shatz et al., 2003). However, the later onset of false-belief understanding between those with low and high SES is not reflected in poorer performance in mindreading tasks in adulthood. In fact, studies indicate adults from lower SES backgrounds are better at recognising others' emotions (Kraus, Piff and Keltner, 2009; Kraus, Côté and Keltner, 2010; Kraus et al., 2012). It has been suggested that individuals with low SES develop 'contextualist' tendencies, which can be seen in folk explanations of social events and self-concepts. Regarding explanations of events, people with low SES are more likely to describe events in terms of contextual factors than people with high SES, who are more likely to offer dispositional explanations. The focus on context by people with low SES is not confined to verbal explanations of events but is mirrored in perceptual attention to context when judging others' emotional states. For instance, Kraus and colleagues (2009) found that lower SES individuals were more influenced by background information, in this case, the emotional states of other people in the background, when rating the emotional state of someone in the foreground. However, some of the results mentioned above are reproduced even when SES is manipulated, suggesting that differences in stable socio-cognitive patterns are not the only status-related effects on mindreading. To explore this, I turn to work on social power (e.g., Guinote, 2013; Galinsky et al., 2003) and developmental studies of the effect of salient status (Rizzo and Killen, 2018) to suggest the beginnings of an explanation for why salient or transitory status can have similar results to the outputs of the socio-cognitive patterns formed by long-term position within a society. The upshot is that interaction dynamics and one's position within a culture can each affect one's ability and motivation to mindread, something not captured in received theories of mindreading.

References

Galinsky, A.D., Gruenfeld, D.H. and Magee, J.C., 2003. Interpersonal Relations And Group Processes From Power to Action.

Guinote, A., 2013. Social Power and Cognition. [online] The Oxford Handbook of Social Cognition. https://doi.org/10.1093/oxfordhb/9780199730018.013.0028.

Holmes, H.A., Black, C. and Miller, S.A., 1996. A Cross-Task Comparison of False Belief Understanding in a Head Start Population. Journal of Experimental Child Psychology, 63(2), pp.263–285. https://doi.org/10.1006/jecp.1996.0050.

Kraus, M.W., Côté, S. and Keltner, D., 2010. Social Class, Contextualism, and Empathic Accuracy. Psychological Science, 21(11), pp.1716–1723. https://doi.org/10.1177/0956797610387613.

Kraus, M.W., Piff, P.K. and Keltner, D., 2009. Social class, sense of control, and social explanation. Journal of Personality and Social Psychology, 97(6), pp.992–1004. https://doi.org/10.1037/a0016357.

Kraus, M.W., Piff, P.K., Mendoza-Denton, R., Rheinschmidt, M.L. and Keltner, D., 2012. Social class, solipsism, and contextualism: How the rich are different from the poor. Psychological Review, 119(3), pp.546–572. https://doi.org/10.1037/a0028756.

Rizzo, M.T. and Killen, M., 2018. How social status influences our understanding of others' mental states. Journal of Experimental Child Psychology, 169, pp.30–41. https://doi.org/10.1016/j.jecp.2017.12.008.

Shatz, M., Diesendruck, G., Martinez-Beck, I. and Akar, D., 2003. The influence of language and socioeconomic status on children's understanding of false belief. Developmental Psychology, 39(4), pp.717–729. https://doi.org/10.1037/0012-1649.39.4.717.

Shoard, C., 2020. Spike Lee: 'Race relations today are the direct result of having a black president'. The Guardian. [online] 12 Jun. Available at: <https://www.theguardian.com/film/2020/jun/12/spike-lee-race-relations-today-are-a-direct- response-to-having-a-black-president>

Spaulding, S., 2018. How we understand others: philosophy and social cognition. Routledge focus on philosophy. Abingdon, Oxon; New York, NY: Routledge, Taylor & Francis Group.

Matthew Owen - Covert Consciousness in Biological and Artificial Neural Networks: Physical and Ontological Factors

University of Michigan
matthewkeithowen@gmail.com

There has been growing scientific and philosophical interest in whether simple biological or artificial neural networks can be conscious1-10. Most methods of inquiry (e.g., the Turing test) have focused primarily on the question of whether a system that expresses human-like cognitive behavior could be conscious or not. Here we ask the reverse question: Can a non-behaving, non-responsive system covertly harbor consciousness without any overt behavioral evidence? Observations suggests that in certain conditions, including general anesthesia or neuronal injury, covert consciousness may be present in human subjects despite their apparent sensory-behavioral disconnection from their environment.11-14 In other words, though these human subjects appear unconscious and unaware of their environment because they demonstrate no behavioral evidence, they are actually conscious, albeit covertly. The typical way covert consciousness becomes evident is through neuronal responses from behaviorally unresponsive human subjects. 15,16 However, some current neurobiological theories of consciousness explicitly support the possibility of covert consciousness in the absence of any response at all, whether behavioral or neuronal. 17 Relying on such theories, one could argue that simple systems such as human cerebral organoids might have covert consciousness if their functional organization meets certain criteria.10,18 Likewise, it might be thought that artificial neural networks could be candidates for consciousness if they satisfy similar neuromorphic properties including recurrent connectivity, distributed computation, local memory, and ontogenetic evolution. However, this assumes that consciousness is an emergent property that arises once specific physical conditions are in place, as physicalist views imply, rather than a capacity of a whole conscious subject as non-physicalist views such as hylomorphism suggests. Based on hylomorphism, it is possible that some biological and especially nonbiological neural networks may be inevitably non-conscious because they are merely externally unified ordered aggregates rather than internally unified subjects, a problem which the plasticity of their hardware cannot circumnavigate. This reveals that the possibility of covert consciousness in biological and artificial neural networks depends not only on physical conditions but also ontological conditions. This also discloses how different ontological views about consciousness, often implicit in neurobiological theories of consciousness, lead to different conclusions about whether simple biological and artificial neural networks could be conscious.

References

Butlin, P. et al. Consciousness in Artificial Intelligence: Insights from the Science of Consciousness. arXiv, doi:https://doi.org/10.48550/arXiv.2308.08708 (2023).

Findlay, G. et al. Dissociating artificial intelligence from artificial consciousness. arXiv, doi:https://doi.org/10.48550/arXiv.2412.04571 (2024).

Toker, D. et al. An AI-Driven Model of Consciousness, Its Disorders, and Their Treatment. bioRxiv, doi:10.1101/2024.10.16.618720 (2024).

Miller, W. B., Jr., Baluska, F., Reber, A. S. & Slijepcevic, P. Biological mechanisms contradict AI consciousness: The spaces between the notes. Biosystems 247, 105387, doi:10.1016/j.biosystems.2024.105387 (2025).

Guingrich, R. E. & Graziano, M. S. A. Ascribing consciousness to artificial intelligence: human-AI interaction and its carry-over effects on human-human interaction. Front Psychol 15, 1322781, doi:10.3389/fpsyg.2024.1322781 (2024).

Mogi, K. Artificial intelligence, human cognition, and conscious supremacy. Front Psychol 15, 1364714, doi:10.3389/fpsyg.2024.1364714 (2024).

Chella, A. Artificial consciousness: the missing ingredient for ethical AI? Front Robot AI 10, 1270460, doi:10.3389/frobt.2023.1270460 (2023).

Farisco, M., Evers, K. & Changeux, J. P. Is artificial consciousness achievable? Lessons from the human brain. Neural Netw 180, 106714, doi:10.1016/j.neunet.2024.106714 (2024).

Aru, J., Larkum, M. E. & Shine, J. M. The feasibility of artificial consciousness through the lens of neuroscience. Trends Neurosci 46, 1008-1017, doi:10.1016/j.tins.2023.09.009 (2023).

Owen, M., Hight, D. & Hudetz, A. G. Human Brain Organoids and the Mereological Fallacy. Neuroethics, doi:https://doi.org/10.1007/s12152-024-09577-y (2025).

Leslie, K., Skrzypek, H., Paech, M. J., Kurowski, I. & Whybrow, T. Dreaming during anesthesia and anesthetic depth in elective surgery patients: a prospective cohort study. Anesthesiology 106, 33-42, doi:00000542-200701000-00010 [pii] (2007).

Owen, A. M. et al. Detecting awareness in the vegetative state. Science 313, 1402, doi:10.1126/science.1130197 (2006).

MacDonald, A. A., Naci, L., MacDonald, P. A. & Owen, A. M. Anesthesia and neuroimaging: investigating the neural correlates of unconsciousness. Trends Cogn Sci 19, 100-107, doi:10.1016/j.tics.2014.12.005 (2015).

Huang, Z. et al. Brain imaging reveals covert consciousness during behavioral unresponsiveness induced by propofol. Sci Rep 8, 13195, doi:10.1038/s41598-018- 31436-z (2018).

Bodien, Y. G. et al. Cognitive Motor Dissociation in Disorders of Consciousness. New England Journal of Medicine 391, 598-608, doi:doi:10.1056/NEJMoa2400645 (2024).

Owen, M., Hight, D. & Hudetz, A. G. Neural Voices of Patients With Severe Brain Injury? Cambridge Quarterly of Healthcare Ethics 34, doi:https://doi.org/10.1017/S0963180124000446 (2025).

Massimini, M. & Tononi, G. Sizing Up Consciousness: Towards an Objective Measure of the Capacity for Experience. (Oxford University Press, 2018).

Owen, M. et al. Theoretical Neurobiology of Consciousness Applied to Human Cerebral Organoids. Cambridge Quarterly of Healthcare Ethics 33, 1-21, doi:10.1017/S0963180123000543 (2024).

April Owens - Thinking On Our Feet: Towards an Embodied Functionalism

University of Cambridge
aprilruthowens@gmail.com

In this paper, I argue that functionalism has traditionally been construed in a myopically brain-bound manner, failing to account for the extra-neural bodily contingencies that bear non-trivially on the functioning of mental states. Building on Rosa Cao's 2022 argument that numerous fine-grained material constraints on mental functioning at the neural level preclude the nomological possibility of multiple realizability, I advocate for an "embodied functionalism" that extends the scope of mental states beyond the brain and nervous system to encompass the entire lived body in its environmental context. I contend that relegating bodily factors to mere peripheral status presupposes a firm boundary between cognitive processes and "purely" physiological ones, a distinction challenged by embodied cognitive science. Embodied functionalism offers several advantages: it preserves functionalism's key asset of validating an abstract level of description for mental states; it avoids the problematic commitment to multiple realizability at the neural level; it aligns with empirical findings from embodied cognitive science; and it allows for a more comprehensive understanding of cognition as involving real- time sensorimotor coupling with the environment. By revising functionalism to focus on analyzing mindedness in terms of the functional roles of bodily states rather than just neuronal states, cognitive science can open up a rich space of inquiry into how our mindedness is shaped by bodily and environmental context.

Don Oxtoby - Does Perceptual Recognition Require Judgment?

University of Turin
oxtobydl@gmail.com

When we see a close friend's face, we normally seem to recognize it without making any judgment. Therefore, it can be surprising that, in psychology, recognition is traditionally defined in terms of judging that we have previously experienced a stimulus. This raises a philosophical question: does recognition always require judgment? This paper argues that recognition always requires judgment. I propose a broadly Russellian view of recognition and explain how it is supported by empirical findings in psychology. I then respond to three recent objections. According to Russell, what distinguishes recognition from mere feelings of familiarity is that recognition requires the judgment that what we perceive has been previously experienced. The feeling of familiarity, on the other hand, does not itself require any judgment. This accords with dual-process models of recognition in psychology, which understand recognition in terms of judgments that are supported by either feelings of familiarity or consciously recalled information (or both). Recognition should also be distinguished from mere conditioned response. To do this, I draw upon infant studies. For example, when infants differentially respond to their mother's face, this behavior can be explained without supposing that infants have the ability to recognize their mother's face. A more parsimonious explanation is that previous exposure is accessed unconsciously, via implicit memory, triggering feelings of familiarity and differential behavior like smiling at their mother's face. But recognition is explained by distinct neural mechanisms from implicit memory (i.e. interaction with the hippocampus and cortex). Interaction with such mechanisms supports the claim that recognition is mediated by judgment in a way that implicit memory and mere conditioned response are not. Recently, philosophers have raised three challenges for judgment-based views of recognition. The first challenge is that recognition only requires the operation of relevant sub-personal mechanisms independently of any judgment. I argue that this view fails to exclude cases of mere conditioned response, and cases where the relevant sub-personal mechanisms function but we evidently do not recognize the perceived individual. The second challenge is that, when we see a familiar individual, we often have an immediate affective response that reflects how we feel about them – prior to any judgment. I argue that our immediate affective response is best explained in terms of mere implicit memory rather than recognition. The third challenge is that some animals and infants appear to be capable of recognition but not judgment. I argue that such cases are usually best explained in terms of conditioned response rather than recognition, while some cases (as in recent studies on macaques) may involve both recognition and judgment. Meeting these three challenges suggests that judgment-based views, while emphasizing the cognitive component of recognition, need not overintellectualize recognition.

References

Abid, G. (2021). Recognition and the perception-cognition divide. Mind & Language, 37(5), 770-789. https://doi.org/10.1111/mila.12362

Dokic, J. (2010). Perceptual recognition and the feeling of presence. In B. Nanay (Ed.), Perceiving the world. OUP. DOI:10.1093/acprof:oso/9780195386196.003.0003

Guderian, S., Brigham, D., & Mishkin, M. (2011). Two processes support visual recognition memory in rhesus monkeys. Proc Natl Acad Sci U S A, 108(48), 19425-19430. 10.1073/pnas.1117078108

Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of recognition. Perspectives on Psychological Science, 3(5), 390-399. https://doi.org/10.1111/j.1745-6924.2008.00087.x

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. Psychological Review, 87(3), 252-271. https://doi.org/10.1037/0033-295X.87.3.252

Mandler, J. M. (2006). The foundations of mind. OUP.

Russell, B. The analysis of mind. (2005) Routledge. (Original work published 1921).

Pravajya Pandey - The Feminist Epistemology of ADHD: Unmasking Women's Experiences and Knowledge

Independent Researcher
pravajyapandey@gmail.com

This paper explores how the phenomenon of "masking" among women with Attention-Deficit/Hyperactivity Disorder (ADHD) constitutes a form of hermeneutical silencing, a concept rooted in feminist epistemology. Drawing on Miranda Fricker's (2007) framework of epistemic injustice, the paper examines how both testimonial and hermeneutical injustices are routinely enacted in psychiatric and medical contexts when women's experiences of ADHD are shaped, and often suppressed, by gendered norms. Despite the growing recognition of ADHD as a neurodevelopmental disorder that affects women differently from men, clinical practices remain anchored in male-centric diagnostic criteria, emphasizing hyperactivity and externalized behaviors over internalized symptoms, such as inattention, emotional dysregulation, and anxiety, which are more commonly observed in women (Quinn & Madhoo, 2014; Nussbaum, 2012). This mismatch contributes to testimonial injustice, where women's reports are dismissed due to identity-prejudicial credibility deficits (Fricker, 2007), and hermeneutical injustice, where gaps in shared interpretive resources prevent women from fully articulating or understanding their own experiences (Pohlhaus Jr., 2012). A key mechanism through which this injustice unfolds is masking—conscious or unconscious efforts to suppress symptoms and conform to social expectations. Masking reflects testimonial smothering (Dotson, 2011), wherein individuals preemptively withhold information due to a perceived lack of epistemic receptivity in their audience. Yet, masking also compounds hermeneutical injustice, not merely because it conceals epistemic material, but because it hinders the very development of conceptual tools needed for self-understanding and recognition. Philosophically, the paper argues that masking should be understood not only as a coping strategy or clinical artifact, but as a gendered epistemic harm that arises at the intersection of power, diagnostic authority, and social normativity. This analysis draws on feminist critiques of institutional epistemology (Carel & Kidd, 2017; Nealon, 2025), demonstrating how dominant knowledge practices in psychiatry rely on epistemic shortcuts that obscure atypical or marginalized presentations. The result is a recursive epistemic injustice: masking, which develops partly in response to prior testimonial injustice, leads to further hermeneutical erasure and clinical misrecognition. By reconceptualizing masking as a form of hermeneutical silencing, the paper calls for structural reform in psychiatric epistemology. The inclusion of lived experiences, feminist interpretive frameworks, and epistemic humility (Ho, 2011) in diagnostic processes can help repair distortions in the epistemic terrain that women with ADHD must navigate. Ultimately, this inquiry reframes ADHD not merely as a clinical diagnosis but as a site of contested epistemic authority—one that reflects broader social struggles over who can know, speak, and be believed.

References

Carel, H., & Kidd, I. J. (2017). Epistemic injustice and illness. Journal of Applied Philosophy, 34(2), 172–190. https://doi.org/10.1111/japp.12172

Dotson, K. (2011). Tracking epistemic violence, tracking practices of silencing. Hypatia, 26(2), 236–257. https://doi.org/10.1111/j.1527-2001.2011.01177.x

Fricker, M. (2007). Epistemic injustice: Power and the ethics of knowing. Oxford University Press.

Ho, A. (2011). Trusting experts and epistemic humility in disability. International Journal of Feminist Approaches to Bioethics, 4(2), 102–123. https://doi.org/10.3138/ijfab.4.2.102

Nealon, N. (2025). Epistemic shortcuts and unjust diagnostic practices. Social Epistemology, 39(2), 178–186. https://doi.org/10.1080/02691728.2024.2400067

Nussbaum, M. C. (2012). ADHD and female-specific concerns: A review of the literature and clinical implications. Journal of Attention Disorders, 16(2), 87–100. https://doi.org/10.1177/1087054711416909

Pohlhaus Jr., G. (2012). Relational knowing and epistemic injustice: Toward a theory of willful hermeneutical ignorance. Hypatia, 27(4), 715–735. https://doi.org/10.1111/j.1527-2001.2011.01222.x

Quinn, P. O., & Madhoo, M. (2014). A review of attention-deficit/hyperactivity disorder in women and girls. The Primary Care Companion for CNS Disorders, 16(3), PCC.13r01596. https://doi.org/10.4088/PCC.13r01596

Minal Patil - From Brain to Being: Turiya—The Upanishadic Consciousness Beyond Neural Limits

Savitribai Phule Pune University
minaldpatil1803@gmail.com

Contemporary philosophy of mind remains dominated by physicalist models—such as identity theory, functionalism, and Integrated Information Theory (IIT)—which seek to explain consciousness through neural mechanisms. While these frameworks have advanced our understanding of the brain, they continue to struggle with what David Chalmers (1995) termed the "hard problem" of consciousness: why and how does neural activity give rise to subjective experience (qualia)? In contrast, the Mandukya Upanishad introduces the concept of Turiya, the "fourth state" of consciousness that transcends waking, dreaming, and deep sleep. Unlike physicalist approaches that view consciousness as an emergent property of the brain, Turiya is described as pure, non-dual awareness—the ontological ground from which all mental and physical phenomena arise. In This paper, I argues that Turiya not only reveals the limitations of physicalist accounts but also offers a constructive philosophical alternative for consciousness studies. By juxtaposing Turiya's notion of contentless awareness with Global Workspace Theory (GWT) and IIT, two critical gaps become evident: 1. The Explanatory Gap: Physicalist models are unable to account for states of pure awareness, such as those reported in samādhi or advanced meditative states, where consciousness persists without perceptual or cognitive content. 2. The Phenomenological Gap: These models often marginalize first-person experience, despite its centrality to the phenomenon of consciousness. To address these gaps, I engages the framework of neurophenomenology (Varela) and draws on cross-cultural philosophy (Ganeri, Thompson), proposing that the brain modulates— rather than generates—consciousness. In contrast to panpsychism and idealism, which often remain tethered to modified physicalist assumptions, Turiya suggests a non-empirical foundation for awareness, challenging the primacy of third-person methodologies in cognitive science. Furthermore, integrating Turiya into contemporary discourse initiates a decolonial shift, positioning Indian philosophical traditions as equal partners in shaping global understandings of mind and self. This approach opens compelling possibilities—ranging from how we understand AI consciousness and mental health to how we reimagine epistemic methods in consciousness studies. By taking Turiya seriously, this paper moves beyond reductionist paradigms toward a more integrated science of consciousness—one that honors its phenomenological richness while remaining open to empirical exploration. In doing so, it bridges ancient metaphysical insight with contemporary scientific inquiry, contributing a unique voice to the evolving philosophy of mind.

Rasmus Pedersen - Do We Experience Our Entire Perceptual Field as Temporally Bound?

University of Sydney
rasmus.pedersen@sydney.edu.au

Everything in our conscious perceptual experience appearsto be temporally unified in the present moment—all sensory features in our perceptual field appear temporally bound and perceptually anchored in the present moment in a coherent and unified way. This experience of global temporal binding, noted by other philosophers under different guises such as synchronic unity, co-consciousness, and the global unity of conscious experience (Bayne & Chalmers, 2003; Dainton & Bayne, 2005; Hurley, 1996; James, 1909; Rashbrook, 2013; Viera, 2020), suggests that our phenomenology of temporal binding extends across our perceptual field. However, this phenomenology raises a tension: while we describe our experience as globally unified, our ability to report or discriminate temporal relations across the perceptual field seems constrained by attentional capacities (Burr et al., 2007; Chennu et al., 2009; Donohue et al., 2015; Hartcher-O'Brien & Alais, 2011; Holcombe & Cavanagh, 2008). Some theories even suggest that attention is necessary for temporal binding (Pedersen, 2024). If this is true, then it seems unlikely that we experience all features within our perceptual field as temporally bound. This suggests that the phenomenology of temporal binding might be less rich than we tend to describe it. I investigate this tension by exploring the extent to which the richness of our phenomenology of temporal binding depends on attention. On one view, we experience sensory features as temporally bound across the perceptual field, even if we cannot report or discriminate specific temporal relations. In this case, the phenomenology of temporal binding overflows our functional discriminatory capacities, which are constrained by limited attentional resources. Perceptual discriminations and cognitive access may depend on attention, while the phenomenology itself remains pre-attentive. On another view, our phenomenology and discriminatory capacities rely on the same limited attentional mechanisms. As such, the apparent discrepancy between phenomenology and discriminatory capacity is merely due to a tendency to subjectively inflate our phenomenology of temporal binding and describe it as richer than it is. I set out a theoretical framework for how one could go about testing this tension by exploring the following three questions:
Q1: Does the apparent richness of people's experiences of temporal binding depend on attention?
Q2: Is there a discrepancy between how (i) the apparent richness of people's experiences of temporal binding and (ii) people's capacity to perform successfully on temporal binding tasks depend on attention?
Q3: If there is a discrepancy between how (i) and (ii) depend on attention, what explains this discrepancy?
For each of these questions, I set out empirically testable hypotheses as well as a couple of novel animations that can be used to test these hypotheses through an empirical test. In turn, this paper provides a framework for understanding the relationship between attention, temporal binding, and the phenomenology of the present.

Anita Pellegrini, Cristina Spinetti, Annaclara Arrigoni - Demystifying the discourse around AI: a proposal

University of Pavia
anita.pellegrini01@universitadipavia.it

There is a widespread consensus that Artificial intelligence (AI) systems do not think, make decisions, understand. Yet, such anthropomorphizing discourse is pervasive, lim- iting studies, misleading the general audience and influencing the legislature. These familiar wishful mnemonics (Mitchell, 2021) may hinder an objective investigation into the nature of AI systems, as language plays an active role in research and forming opin- ions. This semantic framework gives rise to mystification, techno-determinism, moral outsourcing and hype (Placani, 2024). In this context, the popular sentiment towards AI grows in fear and aversion. Since public opinion is directly correlated with resource allocation, funding priorities and legislation, we argue for the need to change the way we structure the discourse around AI. In this work, we argue for the urgent necessity to adopt a neutral language frame- work when talking about AI. First, we analyze the linguistic patterns that compose the problematic tendencies that surround AI discourse by proposing a categorization, ac- cording to the assumptions and implications of said patterns. We specifically focus on expressions suggestive of internal state, social positioning, materiality, autonomy and communication skills - similarly to what other authors have proposed (DeVrio, Cheng, Egede, Olteanu, & Blodgett, 2025). We then present concrete proposals for a transition to a more neutral and accurate language, as well as possible interventions that could be made in domains such as education and the media. Our proposals span multiple dimen- sions: from the need to specifically address AI as a discipline, rather than an entity - to the introduction of interdisciplinary school curricula concerning AI. We argue that a meaningful change would require effort from the scientific com- munity, corporations, and the media. To make our case, we then present the benefits that these organizations would gain when adopting this new framework. Recent data on consumer behaviour and recent case studies on AI-first brands demonstrate how AI mis- conceptions affect trust, leading to a decrease in sales (Zhang & Wang, 2023). Since the adoption of our proposals would hopefully lead to less misconceptions, we argue that would reflect on trust and therefore on sales. Moreover, media companies may gain long-term readers, have more engagement and lower bounce-rate when using a more neutral language, compared to the sensationalistic one currently in fashion (Burgers & De Graaf, 2013). Finally, within this framework, research may avoid disillusion regarding expected findings and explore AI systems' potential more effectively (Ibrahim & Cheng, 2025). In conclusion, we argue that future work is needed to restructure the entire discourse around AI, and that the focus should be shifted towards creating concrete practices to assist different fields, for instance by designing guidelines for educators and journalists. A collective effort is required to foster a more accurate and responsible understanding of AI, an outcome we believe would ultimately benefit everyone.

References

Burgers, C., & De Graaf, A. (2013). Language intensity as a sensationalistic news feature: The influence of style on sensationalism perceptions and effects. Communications- The European Journal of Communication Research, 38(2), 167–188.

DeVrio, A., Cheng, M., Egede, L., Olteanu, A., & Blodgett, S. L. (2025). A taxonomy of linguistic expressions that contribute to anthropomorphism of language technolo- gies. In Proceedings of the 2025 chi conference on human factors in computing systems (pp. 1–18).

Ibrahim, L., & Cheng, M. (2025). Thinking beyond the anthropomorphic paradigm benefits LLM research. arXiv preprint arXiv:2502.09192 .

Mitchell, M. (2021). Why AI is harder than we think. arXiv preprint arXiv:2104.12871 .

Placani, A. (2024). Anthropomorphism in AI: hype and fallacy. AI and Ethics, 4(3), 691–698.

Zhang, Y., & Wang, S. (2023). The influence of anthropomorphic appearance of artificial intelligence products on consumer behavior and brand evaluation under different product types. Journal of Retailing and Consumer Services, 74, 103432.

Chris Percy - Is Global Workspace a functionalist theory? Ignition thresholds point to a possible hybrid interpretation

University of Derby
chris@cspres.co.uk

Global Workspace Theory (GWT; Baars, 1988) and Global Neuronal Workspace Theory (GNWT; Dehaene et al., 1998) can be interpreted as explaining phenomenal consciousness using a computational functionalist lens, despite the usual emphasis on access consciousness (Carruthers, 2019; Butlin et al., 2023; Baars & Franklin, 2009; Mudrik et al., 2025). In principle, GWT predicts that any system with multiple specialised systems operating in parallel (modules), a selective attention mechanism, global broadcast to those modules, and state-dependent attention is capable of phenomenal consciousness (Butlin et al., 2023). Lab evidence in mammalian brain settings has been assembled to support these high-level principles (e.g. Mashour et al., 2020, for a review). However, the key empirical evidence (e.g. fMRI/EEG data on widespread activation; masking study data on sustained perceptual representations; long distance loops) is consistent not only with computational functionalism but also with a hybrid approach that additionally requires certain non-computational functionality. A hybrid approach takes the 'ignition' language in GNWT seriously, not just to indicate non-linear transitions in a communications protocol, but as a physical phenomenon. Ignition in physical systems typically has distinctive features that are not necessarily replicated in computational simulations of those systems. Fire or stellar fusion require not only heat or mass, but a spatiotemporal concentration of heat or mass. If consciousness is similar, then it is not only necessary to have global broadcast of information, but that process must also occur above a spatiotemporal intensity threshold. This matters because global broadcast is a purely computational phenomenon: simulating it at any spatiotemporal scale on any substrate delivers the same general function (albeit not necessarily the same utility). The same is not true with an intensity threshold for ignition. While ignition has functional relevance, those functional effects are not fully reducible to abstract computation, in that simulating them does not necessarily induce them. There are at least three reasons motivating a hybrid theory. First, it provides a novel resolution to challenges that GWT predicts consciousness in unexpected places, such as the US economy (Schwitzgebel, 2015), Searle's Chinese Room (Searle, 1982), and the scattered brain reductio from neuroscience (Gidon et al., 2022). Second, it aligns better with certain empirical evidence on widespread activation, in that unconscious information or the unconscious brain is not necessarily in a state of 'zero global broadcast' (in GNWT terms), but often rather in 'minimal' levels of global broadcast below a threshold. If broadcast alone matters, such phenomena would be consciously experienced. Third, it provides a theoretical route for resolving issues plaguing the operationalisation of GWT for non-biological systems, such as what precisely counts as a module and how 'global' is 'global' enough. If correct, this interpretation has implications for efforts to identify consciousness indicators for artificial systems that draw on purely computationalist views of GWT (Butlin et al., 2023; Sebo & Long, 2025; Long et al., 2024).

This session will also explore routes for refining and testing such a hybrid interpretation, alongside objections and alternatives.

References

Baars, B. J. (1988). A Cognitive Theory of Consciousness. Cambridge University Press.

Baars, B. J., & Franklin, S. (2009). Consciousness is computational: The LIDA model of global workspace theory. International Journal of Machine Consciousness, 1(01), 23-32.

Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., ... & VanRullen, R. (2023). Consciousness in artificial intelligence: insights from the science of consciousness. arXiv preprint arXiv:2308.08708.

Carruthers, P. (2019). Human and Animal Minds: The Consciousness Questions Laid to Rest. Oxford University Press.

Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. In Proceedings of the National Academy of Sciences, 95(24). 14529–14534.

Gidon, A., Aru, J., & Larkum, M. E. (2022). Does brain activity cause consciousness? A thought experiment. PLoS biology, 20(6), e3001651.

Long, R., Sebo, J., Butlin, P., Finlinson, K., Fish, K., Harding, J., ... & Chalmers, D. (2024). Taking AI welfare seriously. arXiv preprint arXiv:2411.00986.

Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. Neuron, 105, pp.776–798.

Mudrik, L., Boly, M., Dehaene, S., Fleming, S. M., Lamme, V., Seth, A., & Melloni, L. (2025). Unpacking the Complexities of Consciousness: Theories and Reflections. Neuroscience & Biobehavioral Reviews, 106053.

Searle, J. R. (1982). The Chinese room revisited. Behavioral and brain sciences, 5(2), 345-348.

Sebo, J., & Long, R. (2025). Moral consideration for AI systems by 2030. AI and Ethics, 5(1), 591-606.

Schwitzgebel, E. (2015). If materialism is true, the United States is probably conscious. Philosophical Studies, 172, 1697-1721.

Marco Peruzzo - The (yet) Undefined Artificial General Intelligence - On the Nature and Evaluation of AGI

University of Padua
marco.peruzzo.7@studenti.unipd.it

Current literature in computer science, cognitive science and philosophy of mind is rich in debates concerning the possible design of Artificial General Intelligence (AGI) and its potential development [1,2], yet a shared, clear and grounded definition of the concept is still missing [3-5]. As the concept of AGI evolved greatly over the years, so did the metrics for assessing whether an artificial system had achieved human-level intelligence spacing from the Turing Test [15], to university graduation [4], to specifically developed AI tests [3]. The problem with standardised tests for humans, such as the SAT or the bar exam is that they fail to distinguish between an AI capable of true cognition and one that simply regurgitates training data answering the questions without actually understanding them [8,16]. Studies show that Deep Neural Networks can "fake" actual comprehension by using task- unrelated correlations to find shortcuts to solve problems instead of actually understanding them [19]. The development of AI benchmarks based on the model of ability-oriented testing used in comparative psychology could grant experts the ability to distinguish truly intelligent behaviour in a machine from mere automatic brute-forcing to gain reward [5]. Yet the same problem can be found in Large Language Models that, while being capable of impressive linguistic skills, evidence points to them not having an actual understanding of language as humans do, and instead operating on a mere prediction of what word will most likely follow a given linguistic prompt [20,21]. It could be argued that such limitations are not functional, but rather epistemic. The AIs were capable of solving the problems, functionally achieving the result, but simply unable to do so while exhibiting human-like cognition. We could resolve the issue by finetuning tests to add reasoning as a requirement to pass them (see ARC), but this still leaves some questions unanswered. In fact, as the ARC creators themselves argue, the only line we are currently capable of drawing is a merely pragmatic one stating that AGI will be achieved when the process of ideating tests easy for untrained humans yet hard for AI will become simply impossible [18]. Could such pragmatic delineation be sufficient for AGI theoretical definition? Would the practical impossibility of distinguishing between biological and synthetic minds be enough to declare the advent of general intelligence in AI? Could this evaluatory limitation be evidence in favour of a functionalist conceptualisation of general intelligence advocating for its complete possible computational reduction? In AI, will there ever be a real difference between enacting cognition and actually possessing it? If such difference is untestable, is it even there? And Finally, should "true" cognition be even a requirement for AGI? Should requirements surpass the functional sphere? We assume humans to be truly capable of cognition, and not mere simulators of it, why should we hold a different standard for AI?

References

1. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P. Lee, Y.T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of artificial general intelligence: early experiments with Gpt-4. ArXiv:2303.12712

2. Morris, M. R., Sohl-Dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., Legg, S. (2023). Levels of AGI: Operationalizing Progress on the Path to AGI. Preprint ArXiv:2311.02462.

3. Chollet, F. (2019). On the Measure of Intelligence. ArXiv:1911.01547

4. Goertzel, B. (2014). Artificial General Intelligence: Concept, State of the Art, and Future Prospects. Journal of Artificial General Intelligence, 5(1), 1–48.

5. Voudouris, K., Crosby, M., Beyret, B., Hernández-Orallo, J., Shanahan, M., Halina, M., & Cheke, L. G. (2022). Direct human-AI comparison in the animal-AI environment. Frontiers in Psychology, 13, 711821.

6. Campbell, M., Hoane, A. J.Jr., and Hsu, F. (2002). Deep blue. Artif. Intell. 134, 57–83.

7. Knight, W. (2023). "Some Glimpse AGI in ChatGPT. Others Call It a Mirage", Wired, URL: https://www.wired.com/story/chatgpt-agi-intelligence/.

8. Leffer L. (2024). In the Race to Artificial General Intelligence, Where's the Finish Line? Scientific American, URL: https://www.scientificamerican.com/article/what-does-artificial-general-intelligence-actually-mean/.

9. Deary, I.J., Cox, S.R. & Hill, W.D. (2022). Genetic variation, brain, and intelligence differences. Mol Psychiatry 27, 335–353.

10. Macgregor J, Chu Y. (2011). Human performance on the traveling salesman and related problems: A review. The Journal of Problem Solving. 3(2).

11. Hernández-Orallo J,´ Dowe D.L., Hernandez-Lloreda M.V. (2013). Universal psychometrics: Measuring cognitive abilities in the machine kingdom. Cognitive Systems Research, 27, 50–74.

12. Legg, S., & Hutter, M. (2007). Universal Intelligence: A Definition of Machine Intelligence. Minds and Machines, 17(4), 391–444.

13. Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and brain sciences, 40, e253.

14. Picard, R.W. (1997). Affective Computing. The MIT Press

15. Turing, A. M. (1950). Computing Machinery and Intelligence. Mind, 59(236), 433-460.

16. 40Hernández-Orallo, J. (2020). Twenty years beyond the Turing test: moving beyond the human judges too. Mind. Mach. 30, 533–562.

17. Spelke, E.S. & Kinzler, K.D. (2007). Core knowledge. Developmental science.

18. Chollet F. (2024). OpenAI o3 Breakthrough High Score on ARC-AGI-Pub, ARC PRIZE, URL: https://arcprize.org/blog/oai-o3-pub-breakthrough.

19. Dong, Y., Liao, F., Pang, T., Su, H., Zhu, J., Hu, X., et al. (2018). Boosting adversarial attacks with momentum. Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 9185–9193.

20. Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2024). Dissociating language and thought in large language models. Trends in cognitive sciences, 28(6), 517–540.

21. Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing (3rd ed.). Pearson.

Giulia Piredda - Extended mind, extended consciousness and the "sensorimotor constraint": philosophy of mind meets philosophy of technology

IUSS Pavia
giulia.piredda@iusspavia.it

The extended mind thesis argues that, depending on the role they play in the cognitive and mental processes, some external (non-biological) resources can be considered as part of the cognitive system or part of mental states that extends beyond the organism (Clark, Chalmers 1998). The functionalist credentials of the extended mind mean that this also applies to the individual's internal resources: for an internal resource, such as a brain area, to count as part of the cognitive process, it must play the proper functional role in the cognitive process itself. This appeal to a "coarse-grained" functionalism is the best (and also the only) mark of the cognitive that extended mind theorists – Andy Clark in particular – has managed to come up with over the years. However, this does not apply – according to Clark and Chalmers – to conscious states that are always internal and do not extend. Over the years, several critics have emphasized two (maybe connected) points regarding the import of the extended mind thesis and the relation between extended mind and extended consciousness (see in particular Farkas 2012; Vold 2015).

> (1) Taken as a thesis concerning the cognitive and mental role of things like neural or cognitive prosthesis, the extended mind is obviously true but also trivial; it must have to mean something else, something more meaningful (Farkas 2012, 2019).
>
> (2) The unwanted (by Clark and Chalmers) passage from extended mind to extended consciousness is not impossible, as many contributions on the subject have shown (Vold 2015, Hurley and Noë 2009, Hurley 1998).

More recently, Chalmers (2019) has come back to these points and argued that the gist of the extended mind should be reinterpreted as follows: (a) the right cases of extended mind should involve some sensorimotor activities and, as a consequence, (b) the cases of extended circuitry (e.g., cyborgs) are not good cases of mental extension. This move could be dubbed the "sensorimotor constraint" and in this talk/paper I would like to offer some reflections on it, considering its import in the extended mind (e.g., vs enactivism) debate and its consequences regarding the relation between philosophy of mind and philosophy of technology.

References

Chalmers, D. (2019). Extended cognition and extended consciousness, in Andy Clark and His Critics, eds M. Colombo, E. Irvine, and M. Stapleton, New York: Oxford University Press.

Clark, A. (2009). Spreading the Joy? Why the Machinery of Consciousness is (probably) still in the Head. Mind118, 963–993. doi: 10.1093/mind/fzp110

Clark, A. (2012). Dreaming the whole cat. Generative models, predictive processing, and the enactivist conception of perceptual experience. Mind 121,753–771. doi: 10.1093/mind/fzs106

Clark, A., Chalmers, D. (1998). The Extended Mind. Analysis 58(1): 7-19.

Farkas, K. (2012). Two Versions of the Extended Mind Thesis. Philosophia 40(3): 435-447.

Farkas, K. (2019). Extended Mental Features. In Colombo, M., Irvine, E., Stapleton, M. (eds.). Andy Clark and His Critics: 44-55. New York: Oxford University Press.

Hurley, S. (1998). Consciousness in Action. Cambridge, Mass.: Harvard University Press.

Hurley, S., Noë, A. (2003). Neural Plasticity and Consciousness. Biology and Philosophy 18(1): 131-168.

Pepper, K. (2014). Do sensorimotor dynamics extend the conscious mind? Adaptive behavior, 22, 2, pp. 99-108.

Vold K. (2015). The parity argument for extended consciousness, Journal of Consciousness Studies, 22, 16–33.

Ward, D. (2012). Enjoying the Spread: Conscious Externalism Reconsidered. Mind, 121, pp. 731–752.

Shawn Prest - Valence, uncertainty and meditative experience: Understanding affective valence with the active inference framework

Monash University
shawnprest@gmail.com

Computational theories of affective valence employing the active inference framework link valence to uncertainty reduction and subjective fitness. However, such a view cannot easily account for the conscious phenomenology of deep meditative experience, where valence is modulated by the degree of contraction, tensing or constriction present in subjective experience. I examine tensions between these two perspectives, aiming for an integrated active inference account of valence. Drawing on work on meditative deconstruction, cognitive effort and traditional Buddhist perspectives, I argue that while valence is superficially tied to allostasis (the process of achieving stability through change), this relationship is mutable via the release of contraction during deconstruction. Synthesizing meditative and uncertainty-reduction perspectives, I show how the release of contraction results in more positive valence, irrespective of subjective fitness, and correlates with decreased hierarchical depth and complexity of an agent's generative model of the world. This more comprehensive computational understanding of valence has important implications for the design of interventions targeting the reduction of suffering and improved well-being.

Bartosz Radomski - The Concept of Adaptivity: A Missing Piece in Cognitive Science?

Independent
bartosz.radomski1@gmail.com

Adaptivity, a capacity to adjust in the face of perturbation, is a prerequisite for cognition. This assumption underlies prominent modelling approaches such as enactivism (Thompson, 2007) and active inference (Parr, Pezzulo, and Friston, 2022). Yet, despite its central role, adaptivity remains conceptually underdeveloped in these frameworks. I argue that models of cognition generally fail to recognise that adaptivity presents its own unique puzzles. By explicating these issues, I advance an approach that enables genuinely incorporating adaptivity into models of cognition and intelligence developed in cognitive science. My main claim is that a rigorous account of adaptivity should address three core puzzles:

1. The Puzzle of Identity: Who or what is adjusting to what?
2. The Puzzle of Norms: What norms guide these adjustments?
3. The Puzzle of Scope: What phenomena does adaptivity apply to?

Without addressing these questions, our attribution of adaptivity will be arbitrary. Living systems that adapt their behaviour would not be meaningfully different from rivers that "adapt" their flow or thermostats that "adapt" to the changing room temperature. To avoid such trivialisation of adaptivity, our investigation should focus on identifying candidate criteria and assessing their appropriateness. Thus, existing models involving the concept of adaptivity (e.g., "adaptive active inference", Kirchhoff, Parr, et al., 2018) could be scored based on how well they resolve these puzzles. In this way, the puzzles can also be used as a guideline for formulating new modelling approaches. Finally, by focusing on these conceptual puzzles, we are also in a more natural position to address an issue that has, to my knowledge, not been addressed in philosophy of cognitive science at all: namely, the origin of adaptivity. How did adaptivity emerge in evolutionary history? The answer to this simple question has profound consequences for our understanding of the evolution of intelligence and life in general. While it is often assumed that adaptivity is necessary for intelligence, it is not clear whether life requires adaptivity. Life without adaptivity is not just a conceptual possibility (Di Paolo, 2005) – it can also serve as a legitimate hypothesis about the origins of life (Frenkel-Pinter et al., 2021; Runnels et al. 2018). As a solution to the problems outlined above, I suggest a new account of adaptivity that builds on heterodox models of cognition – especially, the enactive approach and the active inference – to produce a grand unifying theory for cognitive science. My proposal is to reconceptualise adaptivity as an anticipatory phenomenon rooted in metabolism, going beyond mere homeostasis. Biological systems do not simply respond to perturbations but also anticipate future changes and adjust in advance (by avoiding risks and improving their current conditions). The interpretative lens of adaptivity-as-anticipation would redirect the focus to predictive mechanisms and reinterpret adaptive systems as global prediction systems (Mathis, 2024), possibly suggesting the canonical versions of active inference as an appropriate normative framework to unify accounts of adaptivity in cognitive science (Parr et al., 2022).

Referefences

Di Paolo, E. (2005). Autopoiesis, Adaptivity, Teleology, Agency. Phenomenology and the Cognitive Sciences, 4(4), 429–452. https://doi.org/10.1007/s11097-005-9002-y

Di Paolo, E., Thompson, E., & Beer, R. (2022). Laying down a forking path: Tensions between enaction and the free energy principle. Philosophy and the Mind Sciences, 3. https://doi.org/10.33735/phimisci.2022.9187

Frenkel-Pinter, M., Rajaei, V., Glass, J. B., Hud, N. V., & Williams, L. D. (2021). Water and Life: The Medium is the Message. Journal of Molecular Evolution, 89(1–2), 2–11. https://doi.org/10.1007/s00239-020-09978-6

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. Journal of The Royal Society Interface, 15(138), 20170792. https://doi.org/10.1098/rsif.2017.0792

Mathis, M. W. (2024). Adaptive Intelligence: Leveraging insights from adaptive behavior in animals to build flexible AI systems (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2411.15234

Parr, T., Pezzulo, G., & Friston, K. (2022). Active Inference: The Free Energy Principle in Mind, Brain, and Behavior. The MIT Press.

Runnels, C. M., Lanier, K. A., Williams, J. K., Bowman, J. C., Petrov, A. S., Hud, N. V., & Williams, L. D. (2018). Folding, Assembly, and Persistence: The Essential Nature and Origins of Biopolymers. Journal of Molecular Evolution, 86(9), 598–610. https://doi.org/10.1007/s00239-018-9876-2

Thompson, E. (2007). Mind in Life: Biology, Phenomenology, and The Sciences of Mind. Belknap Press of Harvard University Press.

Kevin Reuter, Tomasz Zyglewicz & Eric Mandelbaum - Folk ontological relativism

University of Gothenburg
kevinreuter@me.com

To be objectivist about a given domain is to think that the domain is singular and the same for everyone, and that its properties are independent of how people perceive it. To be relativist about a given domain is to concede that conflicting perspectives on the domain can be true at the same time. Experts have defended versions of relativism about various domains – often on the basis of folk intuitions about them. For example, philosophically untutored people across cultures seem to be predominantly relativistic with respect to the aesthetic domain (Cova et al., 2019; cf. Zangwill, 2018; for a review: Cova, 2023). There is a weaker, but substantial tendency towards relativism about morality (recent overviews: Pölzler & Wright, 2019; Zijlstra, 2023). It is standard for experiments on folk relativism to use ordinary empirical facts, such as "The Empire State Building has 102 floors," to be a benchmark domain about which the folk are objectivist (e.g., Goodwin & Darley, 2008; Beebe & Sackris, 2016; Davis, 2021). This is not surprising, given that it is a truism among scientists that there is only one reality, which is the same for everyone and whose properties are independent of how people perceive it. Yet, we show that it would be a mistake to think that this assumption is universally held. We present the results of a series of studies suggesting that the folk are ontological relativists. The sources of evidence for folk ontological relativism are threefold. First, the majority of participants readily accept statements such as "Reality is not singular, but rather varies on the individual's perspective" and reject statements such as "When a claim matches reality, it is the same reality for everyone" in forced-choice tasks. The percentage of relativistic responses to these questions corresponds to the "binary" and "simple" bars in Figure 1 below. Second, participants were asked to read a vignette in which the protagonist sincerely uttered p. However, the vignette made it explicit that p is false: Peter and Maria are students and meet up for a late dinner. After their meal, they decide to go to a party. Before heading over, Maria asks Peter what Tom is doing tonight. Peter answers that Tom is at the party. After all, Tom had told him he would be there around this time. But when Peter and Maria arrive at the party, it turns out that Tom never went to the party. Participants were asked to rate their agreement with the following statement on a Likert scale: "Peter's answer matched objective reality." The majority of them agreed, despite the fact that Peter's answer was factually incorrect (1_sense_empirical). We were able to replicate these results with other ways of asking the question (2_senses_empirical), and with respect to a scientific statement about the chemical composition of sugar (2_senses_scientific). In this presentation, we address the objection that our results can be explained solely by people's loose usage of the term "reality." Next, we discuss how our results bear on questions like how the use of words "truth" and "reality" are interpreted in scientific studies and how recent political developments might have influenced people's perspectives on truth and reality. Finally, we suggest that folk ontological relativism could be recruited to explain recent findings suggesting that people fail to use 'is true' in a correspondentist fashion (Barnard & Ulatowski,

2021; Handley-Miner et al., 2023; Johnson-Laird et al., 2023; Reuter, forthcoming; Reuter & Brun, 2022).

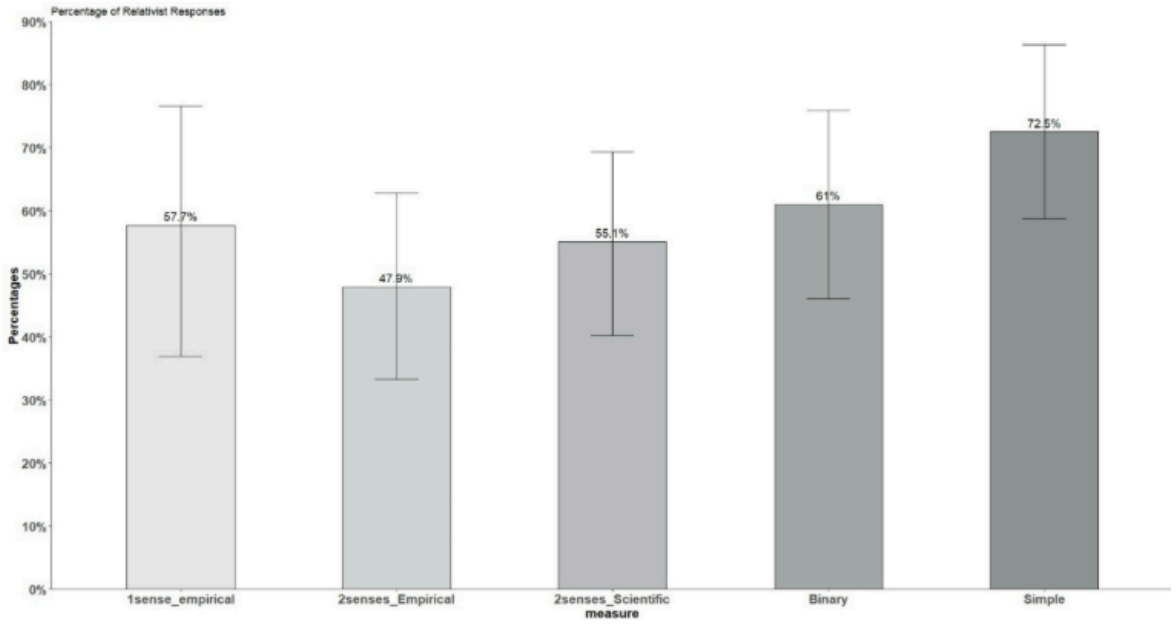

Figure 1. Percentages of participants who respond in line with folk ontological relativism. Error bars indicate 95% confidence intervals.

References

Barnard, R., & Ulatowski, J. (2021). The objectivity of truth, a core truism? Synthese, 198(Suppl 2), 717–733.

Beebe, J. R., & Sackris, D. (2016). Moral objectivism across the lifespan. Philosophical Psychology, 29(6), 912–929.

Cova, F. (2023). Experimental Philosophy of Aesthetics: Aesthetic Judgment. The Compact Compendium of Experimental Philosophy, 393.

Cova, F., Olivola, C. Y., Machery, E., Stich, S., Rose, D., Alai, M., Angelucci, A., Berniūnas, R., Buchtel, E. E., & Chatterjee, A. (2019). De Pulchritudine non est Disputandum? A cross-cultural investigation of the alleged intersubjective validity of aesthetic judgment. Mind & Language, 34(3), 317–338.

Davis, T. (2021). Beyond objectivism: New methods for studying metaethical intuitions. Philosophical Psychology, 34(1), 125–153.

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. Cognition, 106(3), 1339–1366.

Handley-Miner, I. J., Pope, M., Atkins, R. K., Jones-Jang, S. M., McKaughan, D. J., Phillips, J., & Young, L. (2023). The intentions of information sources can aoect what information people think qualifies as true. Scientific Reports, 13(1), 7718.

Johnson-Laird, P. N., Byrne, R. M., & Khemlani, S. S. (2023). Human verifications: Computable with truth values outside logic. Proceedings of the National Academy of Sciences, 120(40), e2310488120.

Pölzler, T., & Wright, J. C. (2019). Empirical research on folk moral objectivism. Philosophy Compass, 14(5), e12589.

Reuter, K. (2024). The ambiguity of "truth" in English, German and Chinese. Asian Journal of Philosophy.

Reuter, K., & Brun, G. (2022). Empirical studies on truth and the project of re-engineering truth. Pacific Philosophical Quarterly, 103(3), 493–517.

Zangwill, N. (2018). Beauty and the agreeable: A critique of experimental aesthetics. Advances in Experimental Philosophy of Aesthetics, 289–308.

Zijlstra, L. (2023). The Psychology of Metaethics: Evidence For and Against Folk Moral Objectivism. In Experiments in Moral and Political Philosophy (pp. 83–103). Routledge.

Achraf El Rhilani - Affordance-based Singularity

University of Bergen
achraf.rhilani@uib.no

The distinction between general and singular thought lies at the heart of philosophical accounts of mental content. A general thought is a thought whose content is a proposition which either is not about any specific individual at all (e.g., 'all humans are mortal') or which is about a specific individual but which picks this individual out indirectly, that is via description (e.g., 'the tallest living man is Turkish'). In contrast, a singular thought is a thought whose content is a proposition which refers directly to an individual (e.g., 'this man is tall'). Such visual-demonstrative thought will serve throughout the presentation as a paradigmatic representative instance of singular thought. There are two main attempts in the literature to answer the question 'what is singular thought?'. As part of the answer, these two attempts provide a jointly sufficient and necessary condition for singular thought. The first is the Acquaintance View, on which singular thought is, roughly, a matter of acquaintance (a relation based on perception). The second is the Significance View, on which singular thought is roughly a matter of cognitive prominence. The Significance View is partly motivated by the fact that it accounts for a broader range of cases of thoughts which appear to be singular but which the Acquaintance View rules out. In particular, it accommodates cases involving abstract objects like numbers and fictional characters. I will adopt the working hypothesis that significance does in fact provide a jointly necessary and sufficient condition for singular thought. My focus, in this presentation, will be on defending the necessity component of this hypothesis against a prominent family of objections. The necessity claim is meant to explain why singularity arises in paradigmatic cases. The sufficiency claim, on the other hand, allows the view to account for non- perceptual cases that are nonetheless intuitively singular. The principal challenge to the necessity component, however, has been put forward most recently by King (2020), who argues that one can construct counterexamples in which a subject perceptually encounters an object without it ever becoming significant to them. Since visual-demonstrative thoughts are paradigmatic instances of singular thought, King's proposed counterexample threatens to undermine the very plausibility of the Significance View. The apparent force of the counterexample depends, I will argue, on a specific and narrow interpretation of 'significance'—an interpretation which is not essential to the viability of the Significance View itself. To demonstrate this, I will begin by outlining what significance amounts to according to Jeshion (2010). On her account, an object is significant to a subject just in case it figures in that subject's plans for action or is connected to their emotional or affective states. This conception opens the door to an explanatory strategy: if we can show that perceptual encounters are themselves typically entangled with action- guiding or affect-laden dispositions, then the kinds of cases invoked by King may turn out to be illusory. Here, I draw on theories of perception that emphasize its constitutive link to action. In particular, I turn to Gibson's affordance-based theory of perception, according to which perceiving an object just is perceiving the possibilities for action it affords. If perception is fundamentally structured by action-guidance, then perceptual encounters would always entail significance in Jeshion's sense. Under this framework, the kinds of perceptual 1 encounters supposed to lack significance may simply be

impossible. However, the Gibsonian account comes with strong metaphysical and semantic commitments, especially regarding the content of perceptual experience. It implies that experience represents higher-level properties such as affordances directly, not inferentially. This assumption is contentious. Nonetheless, the overall strategy can be preserved within a more moderate framework. An alternative, less theoretically demanding view posits that affordances may be represented, rather than perceived directly. On such a view, the action-relevance of a perceptual object is inferred from lower-level features such as shape, color, and motion. Crucially, this weaker view of affordance-perception still suffices to ground significance in perceptual cases. Here, I appeal to empirical models from cognitive neuroscience—especially the two visual systems hypothesis—which distinguish between a 'vision-for-perception' stream and a 'vision-for-action' stream in the visual cortex. This dual architecture provides a robust explanatory basis for the claim that perceptual encounters with objects are pervasively action-oriented and thus significance-laden. A subsequent discussion will explore how this account of affordance perception bears on the nature of conscious experience—specifically, how it intersects with the widely held view that forming a visual-demonstrative singular thought about an object o requires being conscious of o and/or attending to it. A further discussion will return to Jeshion's formulation of the Significance View and propose that the relevant notion of significance need not involve full-blown intentions or robust action dispositions, but should instead rest on derived or latent possibilities for action. In the closing section of the presentation, I will take stock of the theoretical trajectory just sketched and offer suggestions for future work. With the necessity of significance for singular thought defended in the perceptual domain, the next step is to explore how this new affordance-based notion of significance can be extended to cover singular thoughts beyond perception. I will suggest that we do so by generalizing the notion of affordance to encompass mental action—such as imagining, supposing, or calculating—and by offering a substantive account of how cognitive contact is established in each case. This will involve investigating what it means for a fictional character or a mathematical object to become cognitively prominent or action-relevant within a subject's mental life in a way that remains in line with how affordance operates in the perceptual cases.

Mercedes Rivero-Obra - Beyond Affordances: Interacting in Social Environments Through Patterns of Action

UC3M
mercedesobra@gmail.com

Ecological psychology proposes conceptual tools that differ from those used in cognitive psychology, emphasizing the relationship between perception and action through the concept of affordance: possibilities for action that emerge from the interaction between an agent and their environment (Gibson, 1979). These possibilities depend both on the context and the individual's abilities. In social contexts, affordances refer to how an environment or a technology enables or restricts interactions. The perception of such possibilities varies with experience; for example, a martial arts expert can recognize bodily actions that a novice would not perceive (McClelland, 2024). The perception of action possibilities is also shaped by cultural norms, social expectations, and power relations, which influence what is considered appropriate behavior in a given context. Even when an object affords a certain action—like using a phone—social norms may determine whether that action is acceptable in a particular situation. Here, patterns of action become relevant. These patterns, though linked to affordances, guide the agent toward contextually adaptive behavior. They emerge from both direct interaction with the environment and accumulated past experiences. This approach aims to integrate perceptual and cognitive dimensions, moving beyond their traditional division. Functionally, affordances point to possible uses of objects or environments, but patterns of action also include a symbolic layer. This becomes evident when considering, for instance, a child using a broomstick as a horse. From an ecological perspective, the broom lacks the properties of a horse and thus shouldn't afford such action—even when taking into account the child's capabilities. However, if we understand that it is the pattern of action being perceived and enacted, informed by the child's intention and prior experiences, then the broom can indeed "become" a horse within that framework. The connection between the object's perceived qualities and the child's symbolic intention allows for this imaginative transformation. A similar dynamic appears in the example of a domestic dog that avoids climbing on the couch when its owner is present but does so when alone. The couch affords climbing, but the dog has learned, through experience, that this behavior is not permitted in certain social contexts. This learned behavior illustrates how patterns of action can regulate responses beyond the mere presence of affordances. As with humans, the dog's knowledge arises from embodied and emotional interactions with the environment, rather than from abstract reasoning (Damasio, 2021). Ultimately, patterns of action offer action possibilities that are adaptively appropriate, shaped by the perceptual information available during interaction with the environment and informed by the individual's previous experiences. These patterns integrate both functional and symbolic dimensions, enabling agents to ground their actions in socially intelligible structures. In doing so, they allow for mutual understanding and appropriate behavior within a given social framework. This hybrid proposal positions patterns of action as a way to address some of the

persistent challenges within ecological theory, especially those concerning symbolic action and norm- guided behavior.

References

Damasio, Antonio. 2021. Sentir y saber: El camino de la consciencia. Ediciones Destino. Gibson,

James J. 1950. The perception of de visual world. Boston: Houghton Mifflin. https://doi.org/10.1126/science.113.2940.535

McClelland, Tom; Dunin-Kozicka, Monika. 2024. Affording imagination. Philosophical Psychology, 37 (7): 1615–1638. https://doi.org/10.1080/09515089.2024.2354433

Tiina Carita Rosenqvist - Color and perceptual agreement

Dartmouth College
tiina.c.rosenqvist@dartmouth.edu

Color objectivism is the view that colors are mind-independent properties of external objects—either physical properties (Byrne & Hilbert 2003) or primitive properties that supervene on physical properties (Allen 2017). Non-objectivist views include color relationalism, which holds that colors are relational and mind-dependent (Cohen 2009; Chirimuuta 2015), and color subjectivism, which treats colors as purely psychological properties (Hardin 1993). Non-objectivist views of color are often motivated by perceptual variation—differences in color experience across species, individuals, and contexts. In contrast, objectivist views are typically motivated by perceptual agreement— the tendency for perceivers to report similar color experiences or judgments when viewing the same targets. Recently, Elay Shech & Michael Watkins have advanced a new argument from perceptual agreement against non-objectivist views. They highlight a specific case of perceptual agreement that, they argue, "proves to be a serious and long overlooked problem" for non-objectivists (2023, 133). In their example, two perceivers observe two walls under various illumination conditions and agree that the walls are the same color. Shech & Watkins then pose the question: how can we explain "our ability to determine that two objects are indiscernible in color across all lighting conditions?" (135). They contend that non-objectivist theories lack a satisfactory answer to this question, whereas objectivism offers a straightforward one: the agreement arises because the walls share a common mind-independent property—determinate color. In contrast to Shech & Watkins, I argue that a straightforward explanation is available to non-objectivists of various stripes. This explanation is not based on any metaphysical theory of color, but rather on an account of the functions of color vision. In short, the color non-objectivist can adopt the plausible view that one of the functions of color vision is to highlight the similarity between object surfaces for the purpose of object identification, re-identification, and comparison. For example, blueberries have similar surface spectral reflectances, and our color visual system might emphasize this similarity by producing a consistent experience of blue in most cases when we observe blueberries, making it easier for us to identify them. If this were indeed one of the functions of color vision, it would straightforwardly explain the perceptual agreement observed in Shech & Watkins's case: we can agree that the walls have similar surface properties, because our color vision enables us to make such comparisons. In everyday contexts, we might say that the walls have the same color, but such statements may not carry much metaphysical significance. Another likely function of color vision is to highlight the differences between surfaces to enable scene segmentation, which the color visual system achieves by enhancing chromatic contrast between targets and their backgrounds (see, e.g., Akins 2001). For example, red apples look redder against green, and this makes them easier for us to spot. In Shech & Watkins's example, context effects are not relevant, but in most natural settings, our color vision enhances contrast, often at the expense of similarity. This accounts for some types of variation in color experience, and non-objectivists can help themselves to this explanation as well.

References

Allen K (2017). A Naïve Realist Theory of Colour. Oxford: Oxford University Press.

Akins K (2001) More than mere colouring: a dialog between philosophy and neuroscience on the nature of spectral vision. In Fitzpatrick SM, Bruer JT (Eds), Carving our destiny: scientific research faces a new millennium. Washington: Joseph Henry Press.

Byrne A, Hilbert DR (2003) Color realism and color science. Behavioral and Brain Sciences 26: 3–64.

Chirimuuta M (2015). Outside Color:Perceptual Science and the Puzzle of Color in Philosophy. Cambridge: MIT Press.

Cohen J (2009). The Red and The Real: An Essay on Color Ontology. Oxford: Oxford University Press.

Hardin CL (1993). Color for Philosophers: Unweaving the Rainbow. Indianapolis: Hackett.

Shech E, Watkins M (2023). The Problem of Perceptual Agreement. Croatian Journal of Philosophy 23(2): 133–138.

Aliya Rumana - Guidance & calibration in rational analysis

University of Texas at El Paso
aliyarumana@arizona.edu

Rational analysis is a family of methods that model cognition from behavioural evidence under the assumption that cognition fulfils rationality's requirements. I argue that they operate by (a) invoking normative considerations to say that rationality requires cognition to do undergo a process P (a normative hypothesis), (b) assuming that cognition fulfils rationality's requirements (an "alignment hypothesis"), and then (c) deducing cognition actually does undergo P. (Different variants of rational analysis appeal to different conceptions of rationality, like adaptive rationality,1 ideal Bayesian rationality,2 bounded rationality,3 and resource-rationality.4) One obvious objection is: doesn't cognition, in fact, regularly flout rationality's requirements? Won't these "rational process models" often be inaccurate? One popular response admits that cognition is often misaligned with our assumptions about rationality's requirements.5 However, it insists that when this happens, the fault (often or always) lies with our assumptions, not with cognition: we must (or might) be wrong about what rationality requires. Once we've corrected our assumptions about rationality's requirements, the response goes, we will (or might) see that cognition is aligned with them. This response is difficult to square with a genuinely normative conception of rationality (of the sort that normative philosophy is interested in). After all, evidence about what cognition does isn't sufficient to license genuinely normative conclusions about what cognition rationally ought to do. 6 The reason has to do with demandingness: rationality's requirements only have "normative force" over us if we're capable of fulfilling them yet liable to flout them. If rationality is genuinely normative, then actual behaviour isn't sufficient to reveal rationality's requirements because it will flout those requirements sufficiently often. If rational analysis assumes cognition rarely (if ever) flouts rationality's requirements, it must be invoking a non-normative conception of rationality, one whose requirements aren't genuinely demanding. In this paper, though, I defend the surprising claim that rational analysis ought to appeal to a genuinely normative conception of rationality, even if it often doesn't in practice. The primary reason, I argue, has to do with a dilemma: (a) increasing a cognitive model's degrees of freedom trivialises its ability to accommodate ("fit") data but (b) decreasing a cognitive model's parameters makes it too simple to accurately represent cognition in all its complexity.7 The standard escape from this dilemma in statistical modelling is calibration: specifying "fixed" parameters in a data-independent way to increase the model's complexity without increasing its degrees of freedom during data accommodation. I'll argue that the primary advantage of rational analysis is that it calibrates cognitive models using robust plausibility standards, which are (mostly or completely) independent of empirical evidence (this is the is-ought gap). Moreover, I'll argue that ensuring the relevant conception of rationality meets our strictest standards of normative plausibility is essential to ensuring that the fixed parameters of the model are genuinely data- independent. Since the resulting conception of rationality meets our strictest standards of normative

plausibility (the sorts of considerations we invoke in the course of normative philosophy), it is our best hypothesis of genuinely normative rationality.

References

Anderson, J. R. (1990). The adaptive character of thought. Hillsdale, NJ: Lawrence Erlbaum.

Cohen, L. J. (1979). On the psychology of prediction: Whose is the fallacy? Cognition, 7(4), 385–407. https://doi.org/10.1016/0010-0277(79)90023-4

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? Behavioral and Brain Sciences, 4(3), 317–331. https://doi.org/10.1017/S0140525X00009092

Crupi, V., & Girotto, V. (2014). From is to ought, and back: How normative concerns foster progress in reasoning research. Frontiers in Psychology, 5. https://doi.org/10.3389/fpsyg.2014.00219

Elqayam, S., & Evans, J. S. B. T. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. Behavioral and Brain Sciences, 34(5), 233– 248. https://doi.org/10.1017/S0140525X1100001X

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. Psychological Review, 98(2), 254–267. https://doi.org/10.1037/0033- 295X.98.2.254

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. Cognitive Science, 32(1), 108–154. https://doi.org/10.1080/03640210701802071

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. Topics in Cognitive Science, 7(2), 217–229. https://doi.org/10.1111/tops.12142

Griffiths, T.L., Chater, N., & Tenenbaum, J. B. (Eds.) (2024). Bayesian models of cognition: Reverse engineering the mind. Cambridge, MA: The MIT Press.

Hitchcock, C. & Sober, E. (2004). Prediction versus accommodation and the risk of overfitting. British Journal for the Philosophy of Science, 55(1):1–34. https://doi.org/10.1093/bjps/55.1.1

Howes, A., Duggan, G. B., Kalidindi, K., Tseng, Y.-C., & Lewis, R. L. (2016). Predicting short- term remembering as boundedly optimal strategy choice. Cognitive Science, 40(5), 1192– 1223. https://doi.org/10.1111/cogs.12271

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. Psychological Review, 116(4), 717–751. https://doi.org/10.1037/a0017187

Hume, D. (1739). A treatise of human nature. London: John Noon.

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. Topics in Cognitive Science, 6(2), 279– 311. https://doi.org/10.1111/tops.12086

Lieder, F., Callaway, F., & Griffiths, T. L. (2024). Resource-rational analysis. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), Bayesian models of cognition: Reverse engineering the mind. The MIT Press.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. Behavioral and Brain Sciences, 43. https://doi.org/10.1017/S0140525X1900061X

Lieder, F., Griffiths, T. L., M. Huys, Q. J., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. Psychonomic Bulletin & Review, 25(1), 322–349. https://doi.org/10.3758/s13423-017-1286-8

Margolis, H. (1987). Patterns, thinking, and cognition: A theory of judgment. Chicago, IL: University of Chicago Press.

Oaksford, M., & Chater, N. (2007). Bayesian rationality: The probabilistic approach to human reasoning. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524496.001.0001

Rumana, A. (2022). Arbitrating norms for reasoning tasks. Synthese, 200(6), 502. https://doi.org/10.1007/s11229-022-03981-8

Stein, E. (1998). Without good reason: The rationality debate in philosophy and cognitive science. Oxford: Oxford University Press.

Tenenbaum, J. B. (1999). A Bayesian framework for concept learning. [Doctoral dissertation, MIT]. Cambridge, MA: MIT Libraries.

Tenenbaum, J. B., Griffiths, T. L., & Chater, N. (2024). Introducing the Bayesian approach to cognitive science. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), Bayesian models of cognition: Reverse engineering the mind. Cambridge, MA: The MIT Press.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Science, 10, 309–18. https://doi.org/10.1016/j.tics.2006.05.009

Thagard, P. (1982). From the descriptive to the normative in psychology and logic. Philosophy of Science, 49(1), 24–42.

Carl Sachs - A Sellarsian Argument for Liberal Enactivism

Marymount University
csachs@marymount.edu

One of the more contentious issues in philosophy of cognitive science is the epistemological and ontological status of "representations". The radical enactivists (Hutto and Myin 2013, 2017; Hutto and Satne 2015) argue that the very concept of “representation” as used in the cognitive sciences is a conflation of covariation and content. Covariation, however complex, is describable in purely extensional terms. By contrast content is sensitive to intensional differences (what matters "under a description"). The radical enactivist position is that covariation cannot constitute content, and that content enters the cognitive scene only with the emergence of socio- cultural practices that govern content attribution (what they call “the hard problem of content”). I shall urge a Sellarsian re-formulation for a less radical or liberal enactivism. It is less radical by allowing for what cognitive scientists want to say about representations, whether as observables (Thomson and Piccinini 2018) or as posits (Drayson forthcoming). The Sellarsian position begins with Sellars's own distinction between "picturing" and "signifying" (Sellars 1960). I take picturing to be an account of representations-as-covariations and signifying to be an account of interpretivism about content. Liberal enactivism agrees with Millikan (1993, 2000, 2004), Neander (2017), and Shea (2018) that for the cognitive sciences, cognition involves representations explained in terms of teleofunctional covariations. However, liberal enactivism holds that the relevant covariations are feedback and feedforward loops across the brain-body-environment dynamical system, to which analog neural computations (Maley 2018) play a contributing role. At the same time, liberal enactivism also agrees with Dennett (1987) and Davidson (2001) that for the manifest image of mind, attributions of content are grounded in interpretative practices necessary for successful triangulation. Pace radical enactivism, liberal enactivism does not explain difference between covariation and content in terms of the absence or presence of scaffolding by socio-cultural practices. Rather, the difference is one of different forms of intelligibility: the explanation of cognition within the scientific image and the understanding of mind within the manifest image. Accordingly, a scientific image of human cognition would take covariation, not content, as the central explanans. Likewise, liberal enactivism accepts content-attribution to nonhuman animals on the same grounds as content-attribution to humans: whether doing so facilitates successful triangulation. The main difference between how we triangulate with nonhuman animals and how we triangulate with each other is that we have evolved a suite of mindshaping practices (Zawidzki 2013) that make it easier for us to attribute content to others and to ourselves. This is why content attribution is more secured and better grounded with humans than with nonhuman animals. Covariation cannot constitute content, not because of a distinction between human and nonhuman minds, but because of a difference in how minds, human and nonhuman alike, are made intelligible.

References

Davidson, Donald. Subjective, Intersubjective, Objective. Oxford University Press, 2001.

Dennett, Daniel. The Intentional Stance. MIT Press, 1987.

Drayson, Zoe. "Representations are (still) theoretical posits". Theoria (forthcoming 2025)

Hutto, Daniel and Erik Myin. Radicalizing Enactivism: Basic Minds without Content. The MIT Press, 2013.

Hutto, Daniel and Erik Myin. Evolving Enactivism: Basic Minds Meet Content. The MIT Press, 2018.

Hutto, Daniel and Glenda Satne. "The Natural Origins of Content". Philosophia 43: 521-536 (2015).

Maley, Corey. "Toward Analog Neural Computation". Minds and Machines 28 (1): 77-91 (2018).

Millikan, Ruth Garrett. White Queen Psychology and Other Essays for Alice. The MIT Press, 1993.

Millikan, Ruth Garrett. On Clear and Confused Ideas. Cambridge University Press, 2000.

Millikan, Ruth Garrett. Varieties of Meaning. The MIT Press, 2004.

Neander, Karen. A Mark of the Mental: In Defense of Informational Teleosemantics. The MIT Press, 2017.

Sellars, Wilfrid. "Being and Being Known" in Science, Perception, and Reality. Ridgeview Press, 1963.

Shea, Nicholas. Representation in Cognitive Science. Oxford University Press, 2018.

Thomson, Eric and Gualtiero Piccinini. "Neural representations observed". Minds and Machines 28(1): 191-235 (2018).

Zawidzki, Tadeuz. Mindshaping: A new framework for understanding human social cognition. The MIT Press, 2013.

Nancy Salay - How Words Help Us Think

Queen's University
salay@queensu.ca

There is general agreement that a capacity to act for reasons is a mark of intentionality. Views differ widely, however, on how 'acting for reasons' unpacks. According to the cognitivist tradition in which individuals are the central units of investigation, intentional agents make sense of their world via internal representations variously construed as neural, mental, or, on some reductive accounts, both. On these views, to act for a reason is to be responsive to some representation of the how the world is, was, or could be. How behaviour is guided by explicit use of representations—e.g., deliberation between whether to pick answer A or B on a multiple-choice exam—is taken to be continuous with the way that implicitly representational processes such as perception guide behavior. For 4E theorists, in contrast, intentional agents are not individuals so much as they are continually shifting agent-situation couplings to and from which responses develop, often reciprocally. Intentional agents learn to cope within their world as they move and act within it; their needs and wants develop in accordance with their capacity to skillfully "operate" within ongoing situation landscapes. To 'act for a reason' here is to be agentive and responsive in a codeveloping agent-situation. In the context of a comprehensive account of cognition, both views offer important insights. The representational approach brings attention to the cognitive power of explicitly deliberative activity but 4E views explain how operative intentionality grounds actions. In this talk, these insights are merged. Representations do have a powerful role to play in deliberative processes but not as internal structures that agents "recur" on; rather, they are external tools for spatiotemporally extending the ongoing situations in which intentional agents are always embedded. On this view, a deliberative capacity, what I will be calling "representational intentionality," is a strongly scaffolded skill rather than a fundamental capacity: while neural activity plays a critical role here, the development of representational intentionality requires in addition a certain kind of environment—one in which there are language practices—and a particular skill with it. Elsewhere (citation suppressed), I give an account of how representational intentionality develops when the requisite endogenous and exogenous factors are present. In this talk, I will present the arc of the view along with some of the key arguments that support it.

Giulia Santelli - Reframing Akrasia in Addiction: From Philosophical Paradox to Neurocomputational Dynamics

Università degli studi Roma 3
giu.santelli@stud.uniroma3.it

Why do individuals persist in actions they know to be harmful? This question, central to Aristotle's account of akrasia in the Nicomachean Ethics (Aristotle, EN VII), finds renewed urgency in the context of addiction. While dual-system theories in cognitive neuroscience (Kahneman, 2011; Holton, 2009) capture the conflict between impulsive and deliberative processes, they often fail to explain how agents can act against evaluative judgment while remaining aware of long-term consequences (Mele, 2012). This paper proposes a novel integration between the philosophical concept of akrasia and the neurocomputational framework known as the Arbitration Model (Lee et al., 2014; Kim et al., 2024). Unlike classic dual-process views, the Arbitration Model dynamically regulates control between model-based (goal-directed) and model-free (habitual) systems based on the relative reliability and cognitive cost of each. I argue that addiction reflects a failure of this arbitration mechanism: an impaired capacity to prioritize model-based strategies in the face of entrenched, low-cost habitual routines (Sebold et al., 2014; Friston, 2018). To refine this account, I introduce a temporal bifurcation -T1 and T2- marking the shift from early, flexible arbitration (T1) to later, rigid dominance of model-free control (T2). This distinction captures the transition from voluntary use to compulsive behavior, preserving the philosophical insight that akratic agents retain evaluative clarity even in advanced stages of addiction (Erginel, 2016; Donini, 1977). Rather than assuming a pre-existing integration of predictive processing and reinforcement learning, this approach juxtaposes them as complementary explanatory frameworks. Each illuminates different aspects of the failure of arbitration in addiction -maladaptive precision weighting in PP (Friston et al., 2012), and strategy misvaluation in RL (Sutton & Barto, 1998; Redish, 2004). Recent findings on vlPFC–putamen arbitration circuits (Lee et al., 2014) further support this multi-layered explanation. This interdisciplinary account bridges philosophical theories of action and judgment with formal models of decision-making, offering a new framework to understand the persistence of akrasia in addiction (Radoilska & Fletcher, 2016).

References

Aristotle. Nicomachean Ethics, Book VII. Donini, P. (1977). Incontinenza e sillogismo pratico nell'Etica Nicomachea. Rivista Critica di Storia della Filosofia, 32(2), 174–194.

Erginel, M. M. (2016). Akrasia and conflict in the Nicomachean Ethics. British Journal for the History of Philosophy.

Friston, K. (2018). Policies and Priors. In Active Inference (Springer).

Friston, K., et al. (2012). Active inference and agency. Philosophical Transactions of the Royal Society B, 367(1594), 484–503.

Holton, R. (2009). Willing, Wanting, Waiting. Oxford University Press.

Kahneman, D. (2011). Thinking, Fast and Slow.

Farrar, Straus and Giroux. Kim, S. H., et al. (2024). Arbitration between model-based and model-free systems in obsessive-compulsive disorder. Biological Psychiatry.

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural computations underlying arbitration between model-based and model-free learning. Neuron, 81(3), 687–699.

Mele, A. (2012). Backsliding: Understanding Weakness of Will. Oxford University Press.

Radoilska, L., & Fletcher, K. (2016). Lessons from akrasia in substance misuse: a clinicophilosophical discussion. BJPsych Advances, 22(4), 234–241.

Redish, A. D. (2004). Addiction as a computational process gone awry. Science, 306(5703), 1944–1947.

Sebold, M., et al. (2014). Model-based and model-free decisions in alcohol dependence. Neuropsychobiology, 70(2), 122–131.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement Learning: An Introduction. MIT Press.

Liberty Severs - Organisation before representation: how control structures support adaptive strategies for learning and decision-making

Ruhr-Universitat Bochum/University of Lisbon
libertysevers@gmail.com

Philosophical debates have long focused on whether (and how) representations are necessary to account for cognitive phenomena (Barandiaran & Moreno, 2006; Cao, 2022; Chemero, 2009; Millikan, 1984; Neander, 2017; Shea et al., 2018). Meanwhile, formal and empirical approaches frequently assume the presence of value functions, cognitive maps and internal models as necessary theoretical posits for modelling how agents perceive, decide, and act in dynamic environments (Collins, 2019; Daw et al., 2006; Friston, 2010; Momennejad et al., 2017; Niv, 2009; Tolman, 1948; Schuck et al., 2016). Against this backdrop, a number of studies in unicellular and non-neural organisms suggest that basic forms of learning and decision-making-like behaviour can emerge in the absence of content-bearing states or internal models. These results raise the question of whether, despite differences in evolutionary taxonomy and constraints, there are convergent functional architectures that support adaptive behaviour across species (Armus et al., 2006; Bich & Moreno, 2016; Bonato et al., 2023; Brette, 2021; Boisseau et al., 2016; Dussutour, 2021; Dussutour et al., 2010; Eckert et al., 2024; Gershman, 2021; Lyon, 2006; Lyon et al., 2021; Nakagaki et al., 2000; Reid et al., 2012, 2016; Saigusa et al., 2008; Sims, 2021, 2023; Wang et al., 2023). To disambiguate this question, I trace the convergence in such cases to the organisational roots of cognition: roots that predate the evolution of nervous systems, yet share key features with their underlying functional architectures. I then characterise how the organisation of control-sensitive structures can account for variation in these functional capacities. Finally, I elaborate on the role of regulatory dynamics over cognitive processes, and propose a common mechanistic framework to reconcile representational and non-representational forms of learning and decision-making, accounting for their origins in simple regulatory loops and relationship to complex model-based planning (Bechtel & Bich, 2021; Bich et al., 2016; Bich & Bechtel, 2022; Cisek, 2019, 2021, 2022; Gómez-Marin & Ghazanfar, 2019; Lyon, 2006). To illustrate these ideas, I focus on Physarum polycephalum, a unicellular slime mould that exhibits complex behaviours such as maze-solving, nutrient-based path optimisation, and dynamically encoded memory (Nakagaki et al., 2000; Reid et al., 2016). This organism demonstrates how non-representational control architectures can underwrite proto-cognitive capacities, including the formation of preferences, outcome sensitivity, and flexible action selection (Bechtel & Bich, 2021), which are thought to be dependent on decentralised oscillatory dynamics and morphological feedback (e.g., reinforcement of nutrient-rich paths by modulating cytoplasmic flow in response to chemical gradients, see Boisseau et al., 2016; Dussutour et al., 2019). However, there are also constraints over their abilities in lieu of these underlying control mechanisms and processes. In particular, attempts to establish more complex forms of learning (e.g., associative) and decision-making (e.g., planning) have been unsuccessful or otherwise contentious. I therefore suggest that

representational strategies plausibly emerge over the course of evolution and development in cases when behaviour can be 'controlled' (and regulated) in ways that extend beyond reactive mechanisms, placing strict conditions on cognitive abilities like planning, counterfactual reasoning, and delayed credit assignment (Cisek, 2022). On this view, representation is not the foundation of cognition, but a developmental achievement: a means of extending and elaborating an already-operant control structure. Overall, this perspective supports a layered model of cognitive explanation, reviving the concept of biological autonomy and minimal models of cognition and agency (Moreno & Mossio, 2015; Godfrey-Smith, 2016; Barandiaran & Moreno, 2006).

References

Armus, H. L., Montgomery, A. R., & Gurney, R. L. (2006). Discrimination learning and extinction in paramecia (P. caudatum). Psychological reports, 98(3), 705–711. https://doi.org/10.2466/pr0.98.3.705-711

Barandiaran, X., & Moreno, A. (2006). On what makes certain dynamical systems cognitive: A minimally cognitive organization program. Adaptive Behavior, 14(2), 171-185.

Bechtel, W., & Bich, L. (2021). Grounding cognition: heterarchical control mechanisms in biology. Philosophical Transactions of the Royal Society B, 376(1820), 20190751.

Bich, L., Mossio, M., Ruiz-Mirazo, K., & Moreno, A. (2016). Biological regulation: controlling the system from within. Biology & Philosophy, 31, 237-265.

Bich, L., & Bechtel, W. (2022). Control mechanisms: Explaining the integration and versatility of biological organisms. Adaptive Behavior, 30(5), 389-407.

Boisseau, R. P., Vogel, D., & Dussutour, A. (2016). Habituation in non-neural organisms: evidence from slime moulds. Proceedings. Biological sciences, 283(1829), 20160446.

Bonato, B., Simonetti, V., Bulgheroni, M., Wang, Q., Guerra, S., Quaggiotti, S., et al. (2023). Evidence of motor intentions in plants: A kinematical study. Journal of Comparative Psychology, 137(4), 228.

Brette R. (2021). Integrative Neuroscience of Paramecium, a "Swimming Neuron". eNeuro, 8(3), ENEURO.0018-21.2021.

Cao, R. (2022). Putting representations to use. Synthese, 200(2), 151.

Chemero, A. (2013). Radical embodied cognitive science. Review of General Psychology, 17(2), 145-150.

Cisek, P., & Hayden, B. Y. (2022). Neuroscience needs evolution. Philosophical Transactions of the Royal Society B, 377(1844), 20200518.

Cisek, P. (2022). Evolution of behavioural control from chordates to primates. Philosophical Transactions of the Royal Society B, 377(1844), 20200522.

Cisek, P. (2019). Resynthesizing behavior through phylogenetic refinement. Attention, perception, & psychophysics, 81, 2265-2287.

Collins, A. (2019). Reinforcement learning: bringing together computation and cognition. Current Opinion in Behavioral Sciences, 29, 63-68.

Daw, N., Courville, A., Touretzky, D. (2006). Representation and timing in theories of the dopamine system. Neural Comput 18:1637–77

Eckert, L., Vidal-Saez, M. S., Zhao, Z., Garcia-Ojalvo, J., Martinez-Corral, R., & Gunawardena, J. (2024). Biochemically plausible models of habituation for single-cell learning. Current Biology, 34(24), 5646-5658.

Figdor, C. (2023). What are we talking about when we talk about cognition? Human, Cybernetic, and Phylogenetic Conceptual Schemes, JoLMA, 4(2), 149–162.

Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature reviews neuroscience, 11(2), 127-138.

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: carving nature at its joints. Current opinion in neurobiology, 20(2), 251-256.

Gershman, S. J., Balbi, P. E., Gallistel, C. R., & Gunawardena, J. (2021). Reconsidering the evidence for learning in single cells. Elife, 10, e61907.

Godfrey-Smith, P. (2016). Individuality, subjectivity, and minimal cognition. Biology & Philosophy, 31(6), 775-796. Gómez-Marin, A., & Ghazanfar, A. A. (2019). The life of behavior. Neuron, 104(1), 25–36.

Lyon, P. (2006). The biogenic approach to cognition. Cognitive Processing, 7(1), 11-29.

Lyon, P., Keijzer, F., Arendt, D., & Levin, M. (2021). Reframing cognition: getting down to

biological basics. Philosophical Transactions of the Royal Society B, 376(1820), 20190750.

Momennejad, I., Russek, E.M., Cheong, J.H. et al. The successor representation in human reinforcement learning. Nat Hum Behav 1, 680–692 (2017).

Moreno, A., & Mossio, M. (2015). Biological autonomy: A philosophical and theoretical enquiry. Springer.

Nakagaki, T., Yamada, H., & Tóth, Á. (2000). Maze-solving by an amoeboid organism. Nature, 407(6803), 470-470.

Neander, K. (2017). A mark of the mental: In defense of informational teleosemantics. MIT press.

Niv, Y. (2009). Reinforcement learning in the brain. Journal of Mathematical Psychology, 53(3), 139-154.

Pezzulo, G., & Cisek, P. (2016). Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. Trends in Cognitive Sciences, 20(6), 414–424.

Reid, C. R., Latty, T., Dussutour, A., & Beekman, M. (2012). Slime mold uses an externalized spatial "memory" to navigate in complex environments. Proceedings of the National Academy of Sciences of the United States of America, 109(43), 17490–17494.

Reid, C. R., Garnier, S., Beekman, M., & Latty, T. (2015). Information integration and multiattribute decision making in non-neuronal organisms. Animal Behaviour, 100, 44-50.

Reid, C. R., MacDonald, H., Mann, R. P., Marshall, J. A., Latty, T., & Garnier, S. (2016). Decision-making without a brain: how an amoeboid organism solves the two-armed bandit. Journal of The Royal Society Interface, 13(119), 20160030.

Saigusa, T., Tero, A., Nakagaki, T., & Kuramoto, Y. (2008). Amoebae anticipate periodic events. Physical Review Letters, 100(1), 018101

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human orbitofrontal cortex represents a cognitive map of state space. Neuron, 91(6), 1402-1412.

Shea, N., Godfrey-Smith, P., & Cao, R. (2018). Content in simple signalling systems. The British Journal for the Philosophy of Science.

Sims, M. (2021). A continuum of intentionality: linking the biogenic and anthropogenic approaches to cognition. Biology & Philosophy, 36(6), 51.

Sims, M. (2023). Many paths to anticipatory behavior: Anticipatory model acquisition across phylogenetic and ontogenetic timescales. Biological Theory, 18(2), 114-133.

Wang, Q., Guerra, S., Bonato, B., Simonetti, V., Bulgheroni, M., & Castiello, U. (2023b). Decision-making underlying support-searching in pea plants. Plants, 12(8), 1597.

Ayoob Shahmoradi - Does thinking require sensory grounding?

Ruhr University Bochum
ayoobsh@gmail.com

The question of whether thinking requires sensory grounding is interesting for several reasons. One important reason concerns the possibility of AI systems—such as large language models (LLMs)—possessing or developing the capacity for genuine thought without relying on any sensory systems. Recently, philosophers such as David Chalmers have argued that thinking does not require sensory grounding Chalmers (2024). According to this view, systems like LLMs could, in principle, think purely on the basis of internal computational processes, without any need for sensory interaction with the world. In contrast, I argue that thinking does require sensory grounding. More specifically, I argue that the notion of "content" these philosophers rely on is not sufficient to support the kind of robust thought that is necessary for genuine thinking. Thinking a thought involves more than just manipulating symbols or internal representations; it requires representing things as being some way or another. However, without some kind of relation to a relevant object, it is unclear how the capacity for genuine representation could be acquired in the first place. Sensory mechanisms help explain how mental states including thoughts could acquire their representational capacities. That is, sensory mechanisms provide the connection between mental states and the external world, thereby explaining how mental states can possess representational properties.

Lu Teng - Metacognition in Aphantasia: Taking the “Conscious” View Seriously

The Australian National University
lu.teng@anu.edu.au

Aphantasia, where individuals report being unable to voluntarily generate conscious visual imagery, has increasingly captured attention in empirical and philosophical literature (Blomkvist, 2023; Michel et al., 2025; Zeman, 2024). One fascinating aspect of this condition is that both behavioral and neuroimaging data suggest that aphantasics do employ visual strategies to perform tasks. For example, in mental rotation tasks, aphantasics align with controls in exhibiting a steady increase in response times as greater degrees of rotation are required (Pounder et al., 2022). In tasks concerning object appearance, aphantasics also demonstrate competence in determining the shapes and colors of non-present objects (Liu & Bartolomeo, 2023). Neuroimaging studies observed overlapping brain activation patterns in aphantasics and controls (Liu et al., 2025). The discrepancy between subjective reports and objective findings motivate some researchers to adopt the “unconscious” view, according to which aphantasics rely on unconscious visual imagery to perform the relevant tasks (Michel et al., 2025; Nanay, 2021). In this talk, I will argue for the alternative, “conscious” view through the lens of metacognition in aphantasia. When it comes to eliciting reliable introspective reports, consciousness research highlights a distinction between “visibility” measures and “confidence” measures. The former instruct participants to rate the visibility of a stimulus, and confidence measures require participants to make a metacognitive judgment—usually in the form of a confidence rating—about the correctness of their perceptual decision. While both kinds of measures are widely employed, some researchers believe that confidence measures are superior to visibility measures since the former explicitly direct attention to features relevant to perceptual-task performance (Michel, 2023; Morales & Lau, 2022). I note that all empirical studies on aphantasia uses the Vividness of Visual Imagery Questionnaire (VVIQ) to recruit participants, where individuals who score below a certain threshold are categorized as having aphantasia. However, as the VVIQ asks participants to only report the “vivacity” of their mental images, it closely resembles visibility measures. I contend that in determining whether aphantasics indeed rely on unconscious visual imagery, we should further investigate their confidence levels with respect to imagery-task performance. If aphantasics are highly confident in the 2 correctness of their decisions, then this actually provides supporting evidence that they engage in conscious visual imagery. To the very best of my knowledge, very few studies on aphantasia have included confidence ratings on imagery decisions as part of their investigations. Intriguingly, like controls, aphantasics typically seem quite confident in their imagery decisions. For example, in Jacob et al. (2018), participants were instructed to imagine a shape and determine whether a given dot fell within the border. The aphantasic individual performed this task with high accuracy and confidence levels like controls (also see Reeder et al., 2024). I highlight that this marks a key dicerence between aphantasia and blindsight, where blindsighted individuals often express great uncertainty about the

correctness of their perceptual decisions. The examination of metacognition in aphantasia urges us to take the "conscious" view seriously!

References

Blomkvist, A. (2023). Aphantasia: In search of a theory. Mind & Language, 38(3), 866–888. https://doi.org/10.1111/mila.12432

Jacobs, C., Schwarzkopf, D. S., & Silvanto, J. (2018). Visual working memory performance in aphantasia. Cortex, 105, 61–73. https://doi.org/10.1016/j.cortex.2017.10.014

Liu, J., & Bartolomeo, P. (2023). Probing the unimaginable: The impact of aphantasia on distinct domains of visual mental imagery and visual perception. Cortex, 166, 338–347. https://doi.org/10.1016/j.cortex.2023.06.003

Liu, J., Zhan, M., Hajhajate, D., Spagna, A., Dehaene, S., Cohen, L., & Bartolomeo, P. (2025). Visual mental imagery in typical imagers and in aphantasia: A millimeter-scale 7-T fMRI study. Cortex, 185, 113–132. https://doi.org/10.1016/j.cortex.2025.01.013

Michel, M. (2023). How (not) to underestimate unconscious perception. Mind & Language, 38(2), 413–430. https://doi.org/10.1111/mila.12406

Michel, M., Morales, J., Block, N., & Lau, H. (2025). Aphantasia as imagery blindsight. Trends in Cognitive Sciences, 29(1), 8–9. https://doi.org/10.1016/j.tics.2024.11.002

Morales, J., & Lau, H. (2022). Confidence tracks consciousness. In J. Weisberg (Ed.), Qualitative Consciousness (1st ed., pp. 91–106). Cambridge University Press. https://doi.org/10.1017/9781108768085.007

Nanay, B. (2021). Unconscious mental imagery. Philosophical Transactions of the Royal Society B: Biological Sciences, 376(1817), 20190689. https://doi.org/10.1098/rstb.2019.0689

Pounder, Z., Jacob, J., Evans, S., Loveday, C., Eardley, A. F., & Silvanto, J. (2022). Only minimal dicerences between individuals with congenital aphantasia and those with typical imagery on neuropsychological tasks that involve imagery. Cortex, 148, 180–192. https://doi.org/10.1016/j.cortex.2021.12.010

Reeder, R. R., Pounder, Z., Figueroa, A., Jüllig, A., & Azañón, E. (2024). Non-visual spatial strategies are ecective for maintaining precise information in visual working memory. Cognition, 251, 105907. https://doi.org/10.1016/j.cognition.2024.105907

Siena, M. J., & Simons, J. S. (2024). Metacognitive awareness and the subjective experience of remembering in aphantasia. Journal of Cognitive Neuroscience, 36(8), 1578–1598. https://doi.org/10.1162/jocn_a_02120

Zeman, A. (2024). Aphantasia and hyperphantasia: Exploring imagery vividness extremes. Trends in Cognitive Sciences, 28(5), 467–480. https://doi.org/10.1016/j.tics.2024.02.007

Rémi Tison - Can ChatGPT read my mind? AI social cognition and mindshaping

George Washington University
remitison94@gmail.com

Recent advances in Large Language Models (LLMs) have sparked intense philosophical and scientific debates concerning their cognitive abilities. Despite their impressive achievements, LLMs still fall short of human performance in the domain of social cognition, particularly in the ability to attribute mental states such as beliefs, desires and intentions to others. In this presentation, I argue that many LLM socio- cognitive limitations have a root which is often neglected in the literature. Human social cognition crucially involves normative cognition. Roughly, normative cognition consists in our ability to adopt normative attitudes; that is, to evaluate thoughts and behaviors as correct or incorrect with respect to a normative standard, and to regulate others and ourselves in light of such evaluations. From the perspective presented here, attributing beliefs or intentions to others crucially relies on normative cognition because it not only has the function of predicting and explaining behavior, but also of justifying and rationalizing it with respect to social and rational norms. The idea that mental state attribution has such a normative or regulative function receives theoretical support from the normativist tradition in analytical philosophy of mind as well as empirical support from numerous recent studies in social psychology on human practices of mental state attribution. If normative cognition is indeed constitutively involved in human mental state attribution, then implementing human social cognition in LLMs will also require implementing normative attitudes. However, the normative attitudes which are necessary for uniquely human socio-cognitive abilities have not yet been implemented in LLMs, which explains some of their current difficulties. I suggest that this is not due to some inherent technological challenge. Rather, proper normative attitudes have not been implemented in artificial agents in general because of ethical and political concerns: this would entail conferring on them substantial ethical and political status in our communities, notably, allowing them to regulate human behavior in ways which do not currently seem acceptable. Making artificial agents more human-like in their social cognition may require providing them with capacities entitling them to something akin to the normative status of persons, which is in tension with the goal of using them as tools.

Uku Tooming - What is exactly imaginative about imagination's rational powers?

University of Tartu
uku.tooming@gmail.com

There has been active debate over whether and how imagination contributes to justifying empirical beliefs, with examples such as a hunter imagining whether he can jump across a stream (Williamson 2016) or someone imagining moving furniture through a doorway. In these cases, imaginative processes appear to justify beliefs that are formed on their basis. A popular assumption in the literature is that for imagination to justify beliefs, it must be properly constrained. Only under appropriate constraints can imagination reliably provide justification In this paper, I investigate the justificatory role of imagination and the relevant constraints that are supposed to make it possible. While it is widely agreed that imagination can justify beliefs, whether its justificatory role is reducible to other cognitive processes is much more debatable. The main question in this paper is whether the constraints that are relevant for the justificatory role of imagination are intrinsic to imagination itself or whether they are reducible to the constraints that are intrinsic to non-imaginative processes. If they are reducible, it suggests that the epistemic role of imagination is reducible to that of inference or perception. The existing accounts of imagination's justificatory role are vulnerable to such a line of reasoning. For example, in the case of the hunter, it can be argued that his imagination is constrained by prior beliefs about his abilities and the situation. However, if that is the case, then his imaginative process may derive its epistemic force from these beliefs and inferential processes and their characteristic constraints, suggesting that imagination's role is reducible to inference (Kinberg and Levy 2023). Alternatively, it can be argued that the hunter's imagination plays a justificatory role in virtue of constraints that do not derive from inference, such as principles of intuitive physics or core object-system principles (see Miyazono & Tooming 2024). However, these constraints, although not accessible to inferential processes, seem to be accessible to perceptual processes, raising the possibility that imagination's epistemic contribution in such contexts can be reduced to the contribution of inference or perception. In light of the aforementioned argument, the positive part of the paper consists in exploring the possible ways in which we can make sense of an irreducible rational role of imagination. Two options will be considered. First, imagination can be seen as a sui generis process that operates with unique contents and is governed by unique constraints, distinct from those of inference or perception. Second, imagination can be seen as a process that accesses both imagistic and propositional contents, thereby integrating perceptual and inferential processes (cf. Shea 2024). Since it is an exploratory paper, I will not take a stand on which of the options is preferable. Both options illustrate how examining the nature of constraints that are intrinsic to imagination, we can better understand the unique epistemic contributions of imagination and whether these contributions extend beyond the contributions of perception and inference.

Daniel Weger - Defending phenomenal structuralism: An error-theoretic account of phenomenal intrinsicalism

Goethe University Frankfurt
d.weger5@googlemail.com

Phenomenal structuralism claims that each phenomenal character is fully individuated by its relations to other phenomenal characters. For example, a red experience is what it is in virtue of being more similar to an orange experience than to a yellow experience, more similar to a purple experience than to a blue experience, and most dissimilar to a green experience, and so on. This relationalist view conflicts with the widely held view of phenomenal intrinsicalism, which maintains that what it is like to undergo a particular experience is purely a matter of that experience's intrinsic properties. While intrinsicalists acknowledge similarity relations between experiences, they deny that phenomenal character can be individuated solely on the basis of these relations. Rather than directly defending phenomenal structuralism, this talk develops an error theory about intrinsicalism. The core idea is that intrinsicalist intuitions are fundamentally mistaken. The key task is thus to explain how we come to endorse these intuitions, why we hold onto them, and why they are nevertheless mistaken. To this end, I will appeal to the following con- siderations: First, introspection and memory suggest that experiences are independent and self-contained, fostering the impression that what it is like to undergo a particular experience is wholly inde- pendent of other experiences. However, it is not at all clear whether the idea of having just a single experience is intelligible and whether we could tell what it is like to have such an expe- rience in isolation. Moreover, there is no guarantee that either introspection or memory or both reveal the deep metaphysical nature of experience and its phenomenal character to us. Second, many properties that strike us as intrinsic, such as mass, are in fact relational. Although we intuitively treat mass as intrinsic, modern physics shows it is relational, dependent on grav- itational interactions. While treating mass as intrinsic is cognitively more efficient and helps us navigate the world effectively, this does not mean that it is in fact non-relational. Likewise, conceiving phenomenal properties as intrinsic may be cognitively efficient without being metaphysically accurate. Third, linguistic behavior further reinforces the impression that phenomenal character is intrin- sic. While adjectives like "big", "expensive", or "delicious" seem to denote intrinsic properties, something is only big, expensive, or delicious relative to a specific reference class of objects. In much the same way, our linguistic habits might lead us astray when describing experienced objects as "red", "squeaky", and "fruity". While this is not to prove that such terms denote relational properties, it casts doubt on the assumption that they denote intrinsic ones. Taken together, these considerations provide an error-theoretic account of phenomenal intrinsicalism. They explain its intuitive appeal, while also pointing out where it goes wrong. The up- shot is that phenomenal intrinsicalism is much less plausible than it is commonly taken to be. This clears the way for a less presuppositional approach to theorizing about the nature of phe- nomenal character in general, and phenomenal structuralism in particular.

Nicola Weiss - Sensorimotor Enactivism: Three Responses to Dreaming

University of Edinburgh
n.weiss@sms.ed.ac.uk

Sensorimotor enactivism (SE) is a theory of embodied cognition whereby the content of perceptual experience is revealed through dynamic interaction with the environment. This sensory input and motor output results in a cognitive state constitutive of these phenomenological sensory experiences. The sensorimotor enactvist therefore contends that instead of cognition involving latent representations of the world, it is the content of our perceptual experience (and its supervenient modality) that determines implicit knowledge maps of the external world. If we take this account of embodied cognition to be true, how then can it be the case that dreams involve perceptual, phenomenological experience whilst our senses are impaired in a state of sleep? I will refer to this idea of dreams as perceptual experiences as the 'Percept Objection'. This paper aims to explain why the percept objection does not pose an existential threat to the SE account of cognition. In addressing this objection, I will give three different methods to which a sensorimotor enactivist may respond. Each response will account for a different variation of a popular dream-type experience. Each response will also include a counter-objection from the opposing view on the cognition spectrum: the representationalist/internalist. First, I will provide a response from Noe (2004) that denies the truth of the percept objection, asserting that dreams are not accurate portrayals of perceptual experience and are instead detail-sparse and unstable, posing no threat to the sensorimotor account. Second, I address vivid dreams in which dreams appear to be accurate and realistic depictions of phenomenological experience. However, although vivid dreams may appear to be indistinguishable from reality, as Barkasi contends, they are incomplete. This is because wakeful, interactive perception of the environment can be the only form of complete phenomenological experience. The third response will concern contentless dreams, constituted by a perceived sense of 'nothingness'. I contend that it is these forms of dreams that provide the strongest case for SE as if there are no distal external stimuli to be perceived, it seems intuitive that sensory facilities should turn inwards to the perception of awareness itself. Hence, this paper will answer the question: if SE is true, then how can it be the case that we have perceptual experiences whilst our sensory faculties are disabled in a state of sleep?

Ben White - Techno-Wantons: Adaptive Technology and the Will of Tomorrow

University of Sussex
b.white@sussex.ac.uk

Recent work within the tradition of 4E cognitive science and philosophy of mind has drawn attention to the ways that our technological, material, and social environments can act as hostile, oppressive, and harmful cognitive scaffolding (Slaby, 2016; Timms & Spurrett 2023; Spurrett, 2024; Figà-Talamanca, 2024). These accounts of "bad scaffolding" push back against a perceived optimistic bias in the wider 4E literature, whereby, according to these critics, our engagements with technology are presented as always taking place on our terms, always to our benefit, in line with out genuine preferences, and in ways uncomplicated by underlying political and economic realities. Critics (e.g., Selinger and Engström 2007; Slaby, 2016), often point to what they call "Andy Clark style extended mind theory" as exemplary of this optimistic bias. As critics have highlighted this alleged bias, part of the work on hostile scaffolding has been to identify, taxonomize, and understand the various forms of harm that cognitive scaffolding can inflict, including epistemic harm, e.g., "echo chambers" (Osler & Krueger, 2021), cognitive harm, e.g., "deskilling" (Aagaard, 2021), and affective harm, e.g., "mind invasion" (Slaby, 2016). This paper enters into that literature and aims to do two things: first, I introduce, describe and motivate the adoption of a novel harm concept centered on agency. I draw on classic work by Harry Frankfurt, who famously introduced the notion of the "wanton" to describe an agent acting solely on the basis of what he called "first order preferences", which are instinctive preferences not subject to metacognitive reflection or control. I argue that emerging forms of adaptive technology threaten to degrade the depth and reflexivity of our temporally-nested hierarchy of preferences. Borrowing Frankfurt's concept of the "wanton", I highlight a specific form of threat from emerging technology that I call the "techno-wanton." In more colloquial terms, I argue that some technological scaffolds threaten to diminish our metacognitive capacity for self control. I thus draw a strong conclusion, based on Frankfurt's arguments, that individuals in a state of technologically scaffolded "wantonness" are, temporarily, not persons in the technical sense of the term. The second part of my argument is that in the literature on hostile scaffolding, the concept of "mind invasion" (introduced in Slaby (2016)) has been overused, and applied too broadly to describe cases better captured by the "techno-wanton" concept. "Mind invasion" refers to cases in which material and technological scaffolding imposes unwanted preferences on a user. I argue that the notion of "techno-wantonness" should replace "mind invasion" in regard to a specific class of case. I address recent work by Spurrett (2024) and Figà-Talamanca (2024), which has applied "mind invasion" to casinos and recommender algorithms respectively. I argue that in both cases, "techno-wantonness" provides a more useful description, whereby the new concept allows for a more subtle distinction between scaffolding impinging on an unwilling mind and technology that panders to the preferences of initially often very willing users.

References

Aagaard J (2021) 4E cognition and the dogma of harmony. Philosophical Psychol 34(2):165–181. https://doi.org/10.1080/09515089.2020.1845640

Osler, L., & Krueger, J. (2022). ProAna Worlds: Affectivity and Echo Chambers Online. Topoi : an international review of philosophy, 41(5), 883–893. https://doi.org/10.1007/s11245-021-09785-8

Figà-Talamanca G (2024) Digitally scaffolded vulnerability: Facebook's recommender system as an Affective Scaffold and a Tool for Mind Invasion. Topoi. https://doi.org/10.1007/s11245-024-10051-w

Frankfurt HG (1971) Freedom of the will and the concept of a person. J Philos 68(1):5–20. https://doi.org/10.2307/2024717

Selinger E, Engström T (2007) On naturally embodied cyborgs. Janus Head 9(2):553–584

Slaby J (2016) Mind invasion: situated affectivity and the corporate life hack. Frontiers Psychology 7

Spurrett D (2024) On hostile and oppressive affective technologies. Topoi. https://doi.org/10.1007/s11245-023-09962-x

Timms R, Spurrett D (2023) Hostile Scaffolding Philosophical Papers 52(1):1–30

Adrian Wieczorek - Don't Stand So Close to Me. A Critique of Radical Enactive Vision and a Plea for Representationalism

Technische Universität Berlin
a.wieczorek@tu-berlin.de

Radical Enactive Accounts of Vision [REV] (Hutto & Myin 2013; Hutto & Myin 2017; Myin & Zahnoun 2024) defend an interactive relationalism of perception, purged from sensory content and accuracy conditions. It has two closely related tenets: (1) Vision is direct, that is not individuated by content. (2) Vision is only practically normative, that is not in/accurate. In opposition to mainstream representationalism (expressed in Tye 2009, 88), REV accepts both (a) that an object o looks in an F-way to perceivers and (b) that o looks in an F-way only if it is experienced as being F, but denies (c) that this requires a priori that o is represented as being F (Hutto & Myin 2013, 121). However, representations traditionally explain the normativity of perception. Thus, to back up relationism with a normative component, REV sees vision as an interaction with environmental conditions of fitness, forged through evolutionary selection (Hutto & Myin 2017) and purely practical attunement to sensorimotor contingencies (Myin & Zahnoun 2024). Hence, in/appropriate looks are sufficiently grounded by in/appriopriate movement and not by modelling (cf. Hutto & Myin 2013, 126; Hutto & Myin 2017, 172). This talk has two goals. First, it is shown that both tenets are inconsistent and produce a dilemma: Either vision is both direct and purely practically normative but not vision. Or direct vision is secured but not only as practically normative. Concerning the first horn, it will be demonstrated how practical normativity intentionally directs perceivers to ethological properties of o (e.g. a banana's being graspable and nutritious) instead of its surface properties (e.g. a banana's shape and yellow color). However, this undermines the specifically visual and phenomenological character of seeing and the more immediate causal power of surface properties. With respect to the second horn, vision is secured, but since surface properties looking a certain way (looking yellow and curved) is not, in itself, reducible to affordance for interaction (being graspable) or fitness value (being nutritious), the way perceptions come about cannot be explained by practical standards (selectionist, sensorimotor) alone. Since o looks as being F, and F being a surface property of o (the object looks yellow, not the experience), visual experience must also be produced and normed with respect to the way how things are – and not only in terms of what the perceiver can do with them or their beneficial effects. But if such "visual distance" is at odds with tightly coupled entanglement in interaction and the rejection of sensory accuracy, this objective character remains unclear in REV. Second, to escape the dilemma and secure vision as normative, representationalism is reconsidered. Contents as looks are posited as explanantia sui generis to mediate the direct but objectively distanced character and provide the normative standard with respect to how things are, accuracy conditions. Since contents are visual looks (e.g. being yellow), they are not linguistic, high-level concepts (e.g. being mind-independent) which penetrate vision.

References

Hutto, D.D.,Myin, E. (2013). Radicalizing Enactivism: basic minds without content. MIT Press https://doi.org/10.7551/mitpress/9780262018548.001.0001

Hutto, D.D.,Myin, E. (2017). Evolving Enactivism. Basic Mind meet Content. MIT Press https://doi.org/10.7551/mitpress/10366.001.0001

Myin, E., Zahnoun, F. (2024). Enactive Vision. Shapiro, L., Spaulding, S. (Eds.). The Routledge Handbook of Embodied Cognition (2nd ed.). Routledge, https://doi.org/10.4324/9781003322511

Tye, M. (2009), Consciousness Revised: Materialism without Phenomenal Concepts. MIT Press. https://doi.org/10.7551/mitpress/9780262012737.001.0001

Iwan Williams - Intention-like representations in language models? The case of function vectors

University of Copenhagen
iwan.r.williams@gmail.com

Despite some scepticism, a growing chorus of AI researchers and philosophers have posited internal representations in LLMs. But how do these representations relate to the kind of mental states we routinely ascribe to our fellow humans? In particular, do LLMs have representations that function like intentions? A minimal condition for intention-like representations is that they have directive content, as opposed to descriptive content. Directive representations function to represent a goal state and direct the system to bring it about. Most representations posited in language models, by contrast, appear to be descriptive – tracking or encoding background facts or features of the input – and some theoretical assumptions about language models seem to imply that they only traffic in descriptive contents. Recent research positing function vectors (Todd et al. 2024; Hendel et al. 2024) challenges this picture – at first glance, these appear to be directive representations, and thus intention-like in this minimal sense. However, I argue that other interpretations are available, and that determining whether these representations qualify as directives requires careful conceptual and empirical work.

Chen-Wei Wu - Conspiracy Theories as Addictive Goods

Rice University
cw205@rice.edu

Conspiracy theories seem worrisome. Beliefs in such theories can lead to poor political decisions, refusal to receive vaccinations, or even acts of terrorism (Brotherton, 2015). Nonetheless, not all conspiracy theories bring about these harms. For example, believing that Elvis is alive doesn't obviously encourage any of the above actions. Is there a reason why we should be suspicious of conspiracy theories in general? In this essay, I propose and examine a hypothesis that conspiracy theories are addictive goods. The hypothesis implies that we should remain vigilant and treat conspiracy theories cautiously as how we treat addictive goods. After reviewing the relevant literature, I suggest that while conclusive evidence for the hypothesis is lacking, there are similarities between the two phenomena such that our normative reactions to addictive goods can be justifiably applied to conspiracy theories. I will first present an overview of the psychology behind conspiracy theories. Particularly, I will focus on how people become "conspiracy-minded," that is, their tendency to keep endorsing conspiracy theories (Brotherton, 2015; Goertzel, 1994). People start to believe conspiracy theories because of various motives, such as helping them make sense of previously unexplainable events (Douglas et al., 2019; Nguyen, 2021). However, these benefits are likely short-term (Douglas et al., 2017; Van Prooijen, 2022). Instead, what sustains their tendency to believe conspiracy theories might be their distrust toward epistemic authority and the fact that they are immersed in a bad epistemic environment (Franks et al., 2017; Harris, 2023; Nguyen, 2020). It can be challenging for conspiracy theorists to stop believing the theories since it is difficult to step out of their echo chambers and rebuild trust in standard epistemic authority. With the above background, I will articulate and evaluate the hypothesis. Specifically, the hypothesis identifies the state of "conspiracy-mindedness" as a kind of addiction, and conspiracy theories are the "substances" leading to the state of conspiracy-mindedness. Conspiracy theorists can "consume" these theories by constructing, defending, mentally rehearsing, or other processes that strengthen their beliefs in these theories. Despite some similarities, conspiracy theories do not exhibit all the features of substance addiction. To start, conspiracy theorists do not experience typical withdrawal effects. They are not chemically dependent upon the theories, and their dependency on conspiracy theories might not share the same mechanism that underlies addiction. Relatedly, it is also unclear if conspiracy theorists experience a feeling of craving if they stop consuming conspiracy theories. More critically, there is no conclusive evidence on the degree to which conspiracy theorists lose their ability to regulate their beliefs. On the contrary, they seem sensitive to the epistemic values and the practical costs of believing conspiracy theories (Dentith, 2019; Wood & Douglas, 2013). Despite these shortcomings, I think the hypothesis rightly captures the idea that conspiracy theories can create a mind trap similar to addictive substances. While a full-fledged addiction theory of conspiracy theories could be far-fetched, an analogy might still be defensible and could ground our normative judgments about conspiracy theories.

References

Brotherton, R. (2015). Suspicious Minds: Why We Believe Conspiracy Theories. Bloomsbury Publishing.

Dentith, M. R. (2019). Conspiracy theories on the basis of the evidence. Synthese, 196, 2243-2261.

Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The psychology of conspiracy 2 theories. Current directions in psychological science, 26(6), 538-542.

Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., & Deravi, F. (2019). Understanding conspiracy theories. Political psychology, 40, 3-35.

Franks, B., Bangerter, A., Bauer, M. W., Hall, M., & Noort, M. C. (2017). Beyond "Monologicality"? Exploring Conspiracist Worldviews. Frontiers in Psychology, 8, 861.

Goertzel, T. (1994). Belief in conspiracy theories. Political psychology, 731-742.

Harris, K. R. (2023). Conspiracy theories, populism, and epistemic autonomy. Journal of the American Philosophical Association, 9(1), 21-36.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. Episteme, 17(2), 141-161.

Nguyen, C. T. (2021). The seductions of clarity. Royal Institute of Philosophy Supplements, 89, 227-255.

Van Prooijen, J. W. (2022). Psychological benefits of believing conspiracy theories. Current Opinion in Psychology, 47, 101352.

Wood, M. J., Douglas, K. M., & Sutton, R. M. (2012). Dead and alive: Beliefs in contradictory conspiracy theories. Social psychological and personality science, 3(6), 767-773.

Han Zhiheng & Mizumoto Masaharu - Moral Cognition Across Cultures: The Influence of Intent and Perspective in the Trolley Problem

Japan Advanced Institute of Science and Technology
hzh951030@gmail.com

This preregistered study investigates how intention and perspective influence moral judgments and moral choice, particularly in the context of the trolley problem. Drawing on prior findings in moral psychology and cross-cultural research, we focus on the role of intentionality attribution, moral judgment, and moral choice across cultures, with special attention to Eastern and Western populations. While Western participants are known to distinguish between intentional and unintentional consequences(Greene 2009, Bago et al. 2022), this distinction has not been consistently observed in Eastern samples(Bago et al. 2022). While previous research, such as Greene et al. (2009) and Bago et al. (2022), distinguishes between intentional and unintentional actions in their experimental scenarios, they do not clarify how participants interpret intentionality (it was simply stipulated). Additionally, the actor-observer bias, a well-known cognitive tendency where people judge others' actions differently than their own, is explored in the moral domain among Asian participants. To examine these questions, we employ a 2 (perspective: first-person vs. third-person) × 2 (personal force: switch vs. push) × 2 (intent structure: means vs. side effect) between- subjects design, resulting in eight experimental conditions. Participants in three countries(N = 1200) will be randomly assigned to one of these conditions. Three key dependent measures will be collected: moral choice (Q1), moral judgment (Q2), and intentionality attribution (Q3). Hypothesis 1 tests the interaction of intent, moral judgment, and culture. Hypothesis 2 addresses the discrepancy between moral judgment and choice across cultures, particularly comparing American, Japanese, and Chinese participants. Hypothesis 3 examines whether the actor-observer bias affects moral judgments in Asian populations. We will analyze the data using logistic regression for the interaction effects (H1 and H3), and Fisher's exact test (one-tailed) for the consistency between moral choice and judgment (H2). Additionally, demographic variables other than country, such as age and gender will be explored as potential moderators. This study aims to contribute to the understanding of cross-cultural moral cognition and clarify whether differences in intentionality processing and perspective underlie divergent moral behaviors in different cultural contexts.

References

Bago, B., Kovacs, M., Protzko, J. et al. Situational factors shape moral judgements in the trolley dilemma in Eastern, Southern and Western countries in a culturally diverse sample. Nat Hum Behav 6, 880–895 (2022). https://doi.org/10.1038/s41562-022-01319-5

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. Cognition, 111(3), 364-371.

Silvano Zipoli Caiani & Claudio Fabbroni - Teleosemantics, Normativity and the Fear of Infinite Regress

Università degli Studi di Firenze
Università degli Studi di Napoli Federico II
claudio.fabbroni@unina.it

A common view in cognitive science holds that mental representations are biological traits that carry content about targets and are therefore normative: they can succeed or fail. This raises a classic problem for naturalistic accounts: how can normativity arise from biology without presupposing prior normative elements? Teleosemantics (Millikan, 1984; Neander, 2017; Papineau, 2016) addresses this by grounding representational content in evolutionary function: a trait represents x because representing x enhanced fitness in the past. Thus, normativity derives from proper biological function. However, this account faces a well-known challenge: if attributing function depends on normative contexts, like the explanatory aims of a research community, then appealing to function to explain normativity risks circularity or regress (Ramsey, 2007; Hutto & Myin, 2012). If every normative trait depends on another, no foundational explanation is possible. This paper questions whether infinite regress really undermines representational theories, particularly teleosemantics. We introduce a distinction between global explanations, which aim to account for the existence of normative traits in general, and local explanations, which explain why a specific trait is normative in a given context. While global explanations may encounter regress, local ones can remain viable even if they rely on normative assumptions. We support this view through the case of hippocampal place cells, whose activity correlates with spatial navigation (O'Keefe & Nadel, 1978; Bechtel, 2016). Scientists treat these neurons as representational because it serves specific explanatory goals, such as understanding behaviour. Attributing function here involves selecting one effect (e.g., spatial mapping) among many, guided by contextual research aims. Our core claim is that teleosemantics functions effectively as a local explanatory framework. It can clarify why a given trait is seen as normative within a specific scientific practice, even if it does not explain normativity globally. In this light, the regress is not vicious: each step in the explanatory chain offers a coherent local account, analogous to defensible infinite chains in epistemology (Klein, 2003). This local perspective allows teleosemantics to preserve both its naturalistic grounding, since it deals only with traits and functions, and its scientific relevance, without needing to resolve normativity once and for all. As such, it aligns with pragmatic and instrumentalist approaches to representation (Egan, 2020; Coelho Mollo, 2022; Taylor, 2022), which see representational claims as tied to explanatory goals rather than metaphysical facts. By scaling down its explanatory ambition from global to local, teleosemantics avoids the threat of regress and remains a viable theory for understanding representational normativity in cognitive science. This shift points toward a more flexible, context-sensitive naturalism, consistent with both scientific practice and philosophical analysis.

References

Bechtel, W. (2016). Investigating neural representations: The tale of place cells. Synthese, 193(5), 1287–1321.

Coelho Mollo, D. (2022). Deflationary realism: Representation and idealisation in cognitive science. Mind & Language, 37(5), 1048–1066.

Egan, F. (2020). A deflationary account of mental representation. In Smortchkova, J., Dolega, K., & Schlicht, T. (Eds.), What Are Mental Representations? Oxford University Press.

Hutto, D. D., & Myin, E. (2012). Radicalizing Enactivism: Basic Minds Without Content. MIT Press.

Klein, P. (2003). When infinite regresses are not vicious. Philosophy and Phenomenological Research, 66(3), 718–729.

Millikan, R. G. (1984). Language, Thought, and Other Biological Categories. MIT Press.

Neander, K. (2017). A Mark of the Mental: In Defense of Informational Teleosemantics. MIT Press.

Papineau, D. (2016). Teleosemantics. In D. L. Smith (Ed.), How Biology Shapes Philosophy. Cambridge University Press.

O'Keefe, J., & Nadel, L. (1978). The Hippocampus as a Cognitive Map. Oxford: Clarendon Press.

Taylor, S. D. (2022). Cognitive instrumentalism about mental representations. Pacific Philosophical Quarterly, 103(3), 518–550.

Wright, C., & Bechtel, W. (2007). Mechanisms and psychological explanation. In Thagard, P. (Ed.), Philosophy of Psychology and Cognitive Science. Elsevier.

Keynote symposium

Mapping non-neural cognition: a synthesis of experimental and theoretical perspective

Margherita Bianchi - Multisensory, multimodal, and temporal aspects in plant signaling and behavior

University of Rome "La Sapienza"

The assumption is that ongoing advancements in comparative studies on ecological signaling and interkingdom communication, including from a biosemiotic perspective, may offer valuable insights into certain aspects of non-neural cognition. Based on current research into the multisensory perception of plants, the theoretical hypothesis is that plants may possess multimodal communication abilities. These abilities could contribute to the flexible regulation of behavior in response to diverse environmental stimuli, especially in variable or noisy contexts where the risk of signal loss increases. The question is whether this aspect might be linked to the capacity to shift from one mode of communication to another in the pursuit of survival goals. One objective will be to address contextual variation, whether at the individual or interactional level, considering the strategies of other organisms, as well as the timing and modalities of communication in plant behavior. Research required to pursue these investigations involves a deeper exploration of the temporal dimension (including perception, duration, regularity and integration) of plant signaling and behavioral processes. Through further inquiry, it may be possible to better understand what is specific and unique to plants and what is shared with other living systems. These interdisciplinary analyses will also foster reflection on the limits, possibilities, and constraints of our observation and interpretation of the processes under consideration.

Leonardo Bich - Minimal cognition is grounded in biological regulation

University of the Basque Country

This talk discusses minimal cognitive capabilities in the context of the organization of biological systems. Based on the organizational account, it analyzes a fundamental requirement for the realization of minimal cognition: the capacity to make decisions. It provides a theoretical model describing how an organism, in order to maintain itself, can assess its internal states and the state of its environment, and make decisions based on the evaluation of these assessments. The talk argues that: (1) this capability for making decisions is grounded in the molecular domain; (2) it originates from mechanisms of biological regulation; (3) it is shared by all present-day living beings; and (4) does not necessarily depend on the presence of a nervous system. Basic decision-making will be discussed by analyzing examples from bacteria

and other organisms in which decisions are made on the basis of measurements performed by regulatory mechanisms. A comparison will be made with abiotic systems such as self-propelling oil droplets, which have been proposed as possible candidates for minimal cognition, but which lack the organization necessary to support decision-making.

Qiuran Wang - Mind the mould: From empirical study to theoretical insights into non-neural learning and decision-making

Ruhr University Bochum

This talk examines recent developments in the study of cognitive behaviours beyond neural systems, drawing on the interdisciplinary work of the non-neural cognition research (NNCR) Lab. Focusing on decentralised and distributed systems, I synthesise experimental findings and theoretical insights that challenge the assumption that learning and decision-making are confined to neural organisms. Case studies on slime mould from our lab include investigations into whether Physarum polycephalum is capable of associative learning and, if so, how such behavioural capacities might be probed in non-neural life forms. I also consider how bio-inspired computational modelling, particularly in the context of decision-making in foraging and migration behaviours in P. polycephalum, may inform novel approaches to reinforcement learning. These inquiries invite broader reflection on how cognitive processes in unconventional organisms can reshape our understanding of cognition and prompt fresh dialogue between philosophy of mind and the empirical sciences.

Contributed symposia

Advances in Neural Representation

The concept of representation has been a historical point of contention among philosophers of the mind sciences. Views on the status of neural representations range from full-blooded antirealism (Port and van Gelder, 1995; Hutto and Myin, 2012; Chemero, 2009; Favela and Machery, 2023), to positions treating them as useful or pragmatic theoretical posits (Chirimuuta, 2024), to accounts holding that the vehicles are real while their cognitive contents are merely theorist-imposed glosses (Egan, 2025), and finally to views committed to their reality in one form or another (Thomson and Piccinini, 2018; O'Brien and Opie, 2004; Gładziejewski and Miłkowski, 2017; Lee and Calder, 2023). This symposium will showcase recent work from three junior scholars that advances our understanding of neural representations. Our first speaker, Bryce Counts, proposes a multilevel integrationist perspective on neural representation. In contrast to claims that there is a privileged level of neural representation (Saxena and Cunningham, 2019; Yuste 2015; Eichenbaum 2018), or that neuroscientists are confused about the level at which neural representation occurs (Favela and Machery 2023), Counts argues that representational operations span many levels, both above and below the single neuron, involving (although no limited to) population coding, single-neuron feature detection, and sub-neuronal dendritic computation. Our second speaker, Johan Heemskerk, identifies an implicit theory of content at work in neuroscientific practice, which he labels maxMI: the idea that a representation's content is the external item with which it shares maximal mutual information. He traces how this assumption underlies methodologies, then addresses theoretical concerns such as the reference class problem. In doing so, he aims to offer indirect support for a realist stance on neural representation. Our last speaker, Matthew Nestor, develops an emerging approach to neural representation called structural representation, according to which informational content is determined by a relation of homomorphism between a representing vehicle and its content. Drawing on a branch of mathematics called measurement theory, Nestor explores the question of how the content-carrying structures of neural populations can be causally efficacious of behavior.

References

Calder, D., & Lee, J. The many problems with S-representation (and how to solve them). (2023). Philosophy and the Mind Sciences, 4. https://doi.org/10.33735/phimisci.2023.9758

Chemero, A. (2009). Radical embodied cognitive science. MIT Press.

Chirimuuta, M. (2024). The brain abstracted: Simplification in the history and philosophy of neuroscience. MIT Press.

Egan, Frances (2025). Deflating Mental Representation (The Jean Nicod Lectures). MIT Press (openaccess).

Eichenbaum, H. (2018). Barlow versus Hebb: When is it time to abandon the notion of feature detector and adopt the cell assembly as the unit of cognition? Neuroscience Letters, 680, 88-93. doi:

10.1016/j.neulet.2017.04.006

Favela LH and Machery E (2023) Investigating the concept of representation in the neural and psychological sciences. Front. Psychol. 14:1165622. doi: 10.3389/fpsyg.2023.1165622 Gładziejewski P, Miłkowski M. Structural representations: causally relevant and different from detectors. Biol Philos. 2017;32(3):337-355. doi: 10.1007/s10539-017-9562-6. Epub 2017 Feb 2. PMID: 28713187; PMCID: PMC5491603.

Nestor, M. (2017). "Narrow Structuralism: Paving a Middle Path Between Cummins and Millikan". The Australasian Journal of Philosophy 95 (1), pp. 109–123.

O'Brien, G., & Opie, J. (2004). Notes Toward a Structuralist Theory of Mental Representation. In H. Clapin, P. Staines, & P. Slezak (Eds.), Representation in mind: New approaches to mental representation (pp. 1–20). Elsevier Science Ltd.

Port, R. F., & van Gelder, T. (Eds.). (1995). Mind as motion: Explorations in the dynamics of cognition. The MIT Press.

Ramsey, W. (2007), Representation Reconsidered, CUP, New York.

Saxena, S., & Cunningham, J. P. (2019). Towards the neural population doctrine. Current Opinion in Neurobiology, 55, 103-111. https://doi.org/10.1016/j.conb.2019.02.002

Swoyer, C. (1991), "Structural representation and surrogative reasoning", Synthese, 87(3), 449-508.

Thomson, E., Piccinini, G. Neural Representations Observed. Minds & Machines 28, 191–235 (2018). https://doi.org/10.1007/s11023-018-9459-4

Williams, D. and Colling, L. (2018). "From symbols to icons: the return of resemblance in the cognitive neuroscience revolution". Synthese 195, pp. 1941–1967.

Yuste, R. (2015). From the neuron doctrine to neural networks. Nature Reviews Neuroscience, 16, 487-497.

Bryce Counts

University of Missouri

A central issue in contemporary neuroscience is how the brain computes over internal representations in order to guide behavior. Recently, a historical debate has resurfaced about whether this ability is the result of processes involving individual neurons and their connections, or instead, entire populations. If novel information processing abilities reside in neural populations only, and not at the level of single neurons and their connections, then an

explanation and analysis of that function can proceed autonomously without modeling or taking into account the nature of information processing at the resolution of single neurons. Advocates for population autonomy invoke an appropriate computational level of analysis that is sometimes called the "unit of cognition." In this paper, I argue against the idea of an individual unit of computation, representation, or cognition. Instead, I present evidence that computational and representational mechanisms span many levels and argue that higher-level computations and representations are realized and constrained by the properties of their lower-level components.

Johan Heemskerk

University of Warwick

In cognitive neuroscience, representational explanations typically involve the attribution of 'contents' to the posited representations. Contents are items external to the representation itself. For example, face recognition is thought to require the representation of a specific shape profile found on human faces. However, neuroscientists have not articulated a general 'theory of content'; a theory which specifies, in theoretical terms, what it is that makes X the content of any given representation, R. Arguably, it is this lack of theoretical footing which has encouraged understandable anti-realist sentiment among philosophers. In this paper, I argue that while neuroscientists have no explicit theory of content, various methodologies used assume an implicit theory of content – which I call maxMI. According to maxMI, very roughly, the content of a representation is the external item with which the representation shares maximal mutual information. I go through the central argument, a run-down of various methodologies detailing how they implicitly assume maxMI. I then consider theoretical issues such as the 'reference class problem'. I then provide a justification for why neuroscientists implicitly assume maxMI: given some constraints, which I lay out, it provides a measure of the information available to the system itself, isolating contents a change in which results in a change to the system. Since differences which make no difference appear to underlie certain anti-realist positions, I hope to provide some indirect support for realism.

Matthew Nestor

University of Adelaide

Structuralism is the view that neural representations are structural representations (O'Brien & Opie, 2004; Williams & Colling 2018). A structural representation is one in which the informational content of the representing vehicle is fixed, at least in part, via a structure-preserving mapping, or homomorphism, between the vehicle and what it represents (Swoyer, 1991). A topographic map is arguably a (public) structural representation in this sense: the relative spacings among contour lines in the map carry information by mirroring the pattern of inclines among evenly spaced horizontal cross-sections of the terrain. Now, it is widely believed that any adequate theory of neural representation must explain how neural vehicles

causally shape downstream activity (and ultimately behavior) in a way that tracks the informational content they carry (Ramsey 2007). For structuralism, this amounts to showing that, minimally, downstream processes are causally sensitive to the content-carrying structure of upstream neural populations (O'Brien & Opie, 2004; Nestor 2017; Shea 2014, 2018; Gładziejewski and Miłkowski, 2017). But what is it for a process to be causally sensitive to structure? In this talk, I sketch an answer to this question by drawing from a branch of mathematics called measurement theory. I illustrate the account with examples from neuroscience and electronics.

Current discussions in the Philosophy of Cognitive Sciences

In this symposium, as members of the Buenos Aires Philosophy of Cognitive Science Research Group (GIFICC) we selected some debates regarding different topics about explanation in philosophy of cognitive science. In particular, the symposium will address: (i) antireductive mechanistic explanation and its potential for commonsense revisionism, (ii) the explanatory power of endorsing a plurality of conceptual formats, (iii) the application of dual-processing explanations to both human and non-human animals and (iv) unification power of predictive processing explanations. In each case, we will critically examine arguments in favor and against and offer our own conclusions regarding the current state of the topic.

Nicolás Alejandro Serrano - Mechanistic explanation: antireductive revisionism

Universidad de Buenos Aires

In her renowned book, Neurophilosophy, P.S. Churchland (1986) holds that the rebuttal of the analytic-synthetic distinction (Quine 1951) implies that pre-theoretic intuitions can't be valid reasons to defend the irreducibility of the mind to the brain. Therefore, she considers that a progressively reductive co-evolution of psychological and neurophysiological theories should be both looked for and expected, eventually leading to the reduction of folk psychology to neuroscience. However, she also notices that in philosophy of cognitive science (and philosophy of mind in particular) reductionism is often frowned upon, and rejected on the basis of what I'll be calling "conservative conceptual arguments" (CCAs). These arguments threaten to stifle scientific research and conceptual engineering in favor of common sense conceptions of the mind. In this talk, I'll hold that the mechanistic view of explanation (Craver 2007, Francken, Slors y Craver 2022) offers a different type of anti- reductionist argument, that neither falls into the typical types of CCAs, nor threatens to stifle scientific research. In order to do so, I'll compare the mechanistic anti-reductionist position with the typical types of CCAs considered by Churchland (1986, 1996) to show that, while the mechanistic view does offer a "conceptual/transcendental" argument against reduction, it is neither based on pre-theoretic intuitions nor rejects engineering and revision of our common sense concepts about the mind. In particular, according to the mechanistic view, explaining a given phenomenon amounts to finding and specifying the mechanism responsible for its production, i.e. a complex system that produces the phenomenon by the interaction of a number of parts, where the interaction between parts can be characterized by direct, invariant, change-relating generalizations (Glennan 2002). Yet, specifying said mechanism presupposes a characterization of the phenomenon of interest in order to identify the mechanism that produces it (Craver 2007,

Francken, Slors y Craver 2022). This implies that the high-level characterization of the phenomena is a necessary part of the explanation that cannot be reduced to its mechanistic base. However, this also allows for a revision of said characterization in light of the discovered mechanism, its parts and organization, leading to a (non-reductive) co-evolution of cognitive psychology and neuroscience.

References

Churchland, P. S. (1986). Neurophilosophy. The MIT Press.

Churchland, P. S., (1996) Toward a Neurobiology of the Mind. The Mind-Brain Continuum – ed. Llinas, Churchland. Cambridge, MA: MIT Press. 281-303.

Craver, C.F. (2007). Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience, Oxford: Clarendon Press

Francken, J.C., Slors, M. & Craver, C.F. (2022). "Cognitive ontology and the search for neural mechanisms: three foundational problems". Synthese 200, 378.

Quine, W. V. O. (1951). "Two Dogmas of Empiricism". The Philosophical Review. 60 (1): 20–43. Glennan, S.S (2002). "Rethinking Mechanistic Explanation", Philosophy of Science, 69(S3): S342–S353.

Sabrina Haimovici - The plurality of conceptual formats and the role of language-like representations

Universidad de Buenos Aires

According to evidence from neuroscience, it does not seem likely that the brain implements and processes discrete language-like representations like the ones posited by the classical language of thought program (Piccinini 2020). The evidence though seems compatible with a weaker version of the language of thought, in the sense that there seem to be language-like representations at least for language processing. In this paper I will argue that this weaker version still has significant explanatory power in accounting for phenomena addressed by theories of concepts, in particular the possibility of abstract representation, productivity and systematicity and some forms of logical reasoning. Moreover, there are current pluralist approaches to the format of concepts that, in spite of differences in the specific representations posited, agree in incorporating symbolic language-like representations as part of the conceptual repertoire. For example, there are proposals that include linguistic and sensorimotor representations (Dove 2009, 2011), symbolic and sensorimotor representations (Zwaan 2016), symbolic and iconic representations (Espino & Byrne 2018, Orenes, Beltrán & Santamaría 2014), and linguistic and cartographic representations (Aguilera 2020). What these accounts have in common is that they postulate language-like representations, but do not restrict conceptual processing to the operations performed over them. There is a growing consensus in a pluralist view that combines language-like representations with other formats, which preserves some of the explanatory power of the classical language of thought hypothesis, without committing to neurally implausible approaches.

References
Aguilera, M. (2020). Mapas, lenguaje y conceptos: hacia una teoría pluralista del formato de los conceptos. Principia: An International Journal of Epistemology 24(1): 121–146.

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. Cognition, 110, 412-43.

Dove G (2011) On the need for embodied and dis-embodied cognition. Front. Psychology 1:242. doi: 10.3389/ fpsyg.2010.00242

Espino, O. & Byrne, R. (2018). Thinking About the Opposite of What Is Said: Counterfactual Conditionals and Symbolic or Alternate Simulations of Negation. Cognitive Science, 1-43.

Orenes, I., Beltrán, D. & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. Journal of Memory and Language, 74, 36-45.

Piccinini. G. (2020). Neurocognitive mechanisms: Explaining biological cognition. Oxford University Press.

Zwaan, R. (2016). Situation models, mental simulations, and abstract concepts in discourse comprehension. Psychon Bull Rev 23:1028–1034.

Fernanda Velázquez - Dual processing and animal minds

Universidad de Buenos Aires

Dual processing theorists argue that we share type 1 processing with non-human animals, but that type 2 processing is distinctively human (Evans 2010, Evans & Stanovich 2013). This claim suggests a strong discontinuity between the minds of animals and humans (Carruthers 2019). However, some argue that one can extend type 2 processing to animal cognition (e.g., Carruthers 2019), wild animal behavior (Teichroeb et al. 2023) and even to insects (Kelly & Barron 2022). If correct, these positions and their corresponding evidence would weaken the strong discontinuity approach. In this presentation, I will examine the interpretation of the evidence about insects as supporting the existence of dual processing in invertebrates. The purpose of my analysis will be to clarify which type of argument is more appropriate to state that non-human animals exhibit type 2 processing. To this end, I will also briefly characterize dual processing in humans and its main theories and controversies (e.g., De Neys, 2023). Regarding dual systems in arthropods, three types of arguments have been offered (Kelly & Barron, 2022). One distinguishes two processing pathways from features of the insect nervous system. This argument relies on a strategy that matches cerebral structure and processing type one-to-one, so that exactly two brain systems are underlying two forms of processing. However, this claim is stronger than what dual processing theorists would want to assert (Evans & Stanovich 2013). The second type of argument draws on similarities between inhibitory processes in arthropod nervous systems and dual systems architectures. However, this path wouldn't be advisable either, as some dual processing theories defend a competitive architecture instead of a default interventionist model. Finally, the third type of argument aims at equating processing features

in insects and humans, yet only considers concurrent (but not defining) features of the processing types (Evans & Stanovich 2013). Therefore, I'll hold that all three types of argument are problematic and that the most promising argumentative strategy for supporting the continuity thesis is to look for the defining features of processing types in non-human animals.

References

Carruthers, P. (2019). "Animal Minds", in Human and Animal Minds (pp.29-51). Oxford: Oxford University Press.

De Neys, W. (2023). Advancing theorizing about fast-and-slow thinking. Behavioral and Brain Sciences, 46

Evans J. St. B. T. (2010). Thinking twice: Two minds in one brain. Oxford, England: Oxford University Press.

Evans J. St. B. T. & Stanovich, K. (2013). Dual-Process theories of higher cognition: Advancing the debate. Perspectives on Psychological Science, 8 (3), 223-241.

Kelly, M. & Barron, A. B. (2022). The best of two worlds: Dual systems of reasoning in animals and AI. Cognition, 225, 105118.

Julie A. Teichroeb, J.A, Smeltzer, E.A., Mathur, V., Anderson, K.A., Fowler, E.J., Adams, F.V., Vasey, E.N., Kumpan, L.T., Stead, S.M., Arseneau-Robar, T.J. (2023). How can we apply decision-making theories to wild animal behavior? Predictions arising from dual process theory and Bayesian decision theory. American Journal of Primatology, 2025, 87:e23565.

Pedro Bianchi - A Possible Unifying Explanation of Cognitive Faculties

Universidad de Buenos Aires

One of the greatest difficulties within cognitive sciences is finding a model capable of unifying the diversity of mental phenomena that occur in the daily life of biological beings, primarily humans. For example, there is a long-standing and well known debate regarding the scope and limits of classical-deductive Turing-like architectures in comparison with connectionist-statistical models such as neural networks, particularly when it comes to explaining thought as a recursive and inferential phenomenon from a general point of view (Fodor & Pylyshyn 1988, Smolensky 1987). An upshot of this kind of debates often is that different cognitive faculties are better explained by different types of models, which in turn implies that different faculties (like perception, thought, and action) have different natures and objectives. However, a theory has emerged recently that could offer a conciliatory and unifying explanation across all cognitive faculties, according to which their only objective would be to predict and process patterns in the environment (Clark 2015). Within this predictive processing framework, perception, thought and action of any living being are considered as faculties that should be able to actively infer environmental regularities from that organism's own generative models, in order to reduce the surprise it experiences (in a computational sense of the term,

see Friston et al., 2006, Friston et al., 2022). In this talk, I argue that this framework has the potential to propose a unifying explanation for various cognitive faculties, as long as it also holds as a principle that the goal of any cognitive function is to algorithmically compress or reduce the information it is actively processing. This idea makes it possible to relate the classical computational characteristics of a symbolic Turing-like deductive architecture with those of a connectionist or statistically- based neural network, without needing to dwell on the specific constitution of each. What is relevant is not how these architectures are constructed but rather how they perform the task of processing the information per se. This can be observed when comparing two different theories of cognitive faculties such as those of Stanislas Dehaene (2022) and Andy Clark (2015), the former being of a classical computational type and the latter of a context- dependent type. The common point between both is the notion of simplicity (Chater & Vitanyi 2003, Chaitin 2003) in each cognitive act, i.e. the reduction of informational complexity performed by each cognition or faculty proposed by both models.

References

Chaitin Gregory J. (2003) Leibniz, Information, Math and Physics, IBM Research Division, Yorktown Heights, NY

Chater Nick & Vitanyi Paul (2003) Simplicity: a unifying principle in cognitive science?, TRENDS in Cognitive Sciences, Vol7 No.1 January, 19-22

Clark Andy (2015) Radical Predictive Processing, The Southern Journal of Philosophy, Volumen 53, Spindel Supplement, 3-27

Dehaene Stanislas, et al, (2022) Symbols and mental programs: a hypothesis about human singularity, Trends in Cognitive Sciences, 761 - 766

Fodor Jerry & Pylyshin Zenon, (1988) Connectionism and cognitive architecture: A critical analysis, Cognition No28, 3-71

Friston Karl & Kilner James & Harrison Lee (2006) A free energy principle for the brain, Journal of Physiology, Paris, 70 - 87

Friston Karl & Pezzullo Giovanni & Parr Thomas (2022) Active Inference, The Free Energy Principle in Mind, Brain and Behavior, MIT Press

Smolensky Paul (1987) The Constituent Structure of Connectionist Mental State: A reply to Fodor and Pylyshin, The Southern Journal of Philosophy, n26, 137-161

Laws of Appearance: Between Contingency and Necessity

EJ Green - Empirical Explanations of the Laws of Appearance

John Hopkins University

It is widely thought that there are limits to how things can appear to us in perceptual experience. For instance, nothing can appear both square and circular, or both pure red and pure blue. Adam Pautz has dubbed such constraints “laws of appearance.” Most of the laws of appearance involve dependencies wherein the experience of one sort of property either requires or precludes the experience of other sorts of properties. But if the laws of appearance obtain, then what explains them? Here I examine the prospects for an empirical explanation of the laws of appearance. First, I challenge extant empirical explanations that appeal purely to the format of perceptual representation, and specifically its alleged iconic or depictive character. I argue that such explanations are committed to an implicit but unwarranted uniqueness assumption about the representational underpinnings of perceptual experience — i.e., that perceptual experience is uniquely underpinned by a single map or iconic array. I then develop a hybrid approach, on which the laws are explained not merely by format, but by two further factors: ecological constraints imposed by our environments, and computational constraints embodied by our perceptual systems. I explain how these factors might produce dependencies between the experiential representation of one property or dimension and the experiential representation of other properties or dimensions. While the hybrid approach implies that the laws of appearance are contingent, I argue that this implication is empirically defensible, since even some of the most intuitively compelling laws have real-world counterexamples. In particular, I adduce scientific evidence for violations of the shape exclusion law (i.e., the law that no visual experience can represent a single object as having two incompatible shape properties simultaneously).

Ned Block - Can Perception Be Inconsistent?

New York University

Many of the “Laws" of Appearance are at best ceteris paribus laws that have exceptions. For example, the “law” of appearance that says that nothing can be both red and green all over at the same level of determinacy derives from the computational fact that long wave cones excite the green-to-red opponent channel whereas medium wave cones inhibit that channel. But it is possible nonetheless to excite both ends of the channel using interventions that do not exist in normal life outside the lab. Interestingly, many of the exceptions to “laws” of appearance are cases in which it seems we have two representations of one object or property that are in a sense to be explained conflict. This talk concerns how it is possible to have conflicting simultaneous representations of the same object or property.

Alfredo Vernazzani - Necessary Laws of Appearance: Perceptual Segregation and Qualitative Discontinuities in Humans and other Creatures

Hanse-Wissenschaftskolleg

Are perceptual appearances governed by metaphysically necessary laws? I argue that they are, and advance four such laws grounded in perceptual discriminatory capacities. The first law states that all appearances are located within a spatiotemporal envelope. The second and third law govern qualitative homogeneity and phenomenal segregation. The fourth law governs apparent boundaries. Together, these laws illuminate the modal structure of experience across sensory-modalities — including those absent in humans — and offer a principled framework for understanding appearances in non-human perceivers. I conclude by situating the laws proposed here within the recent debate on whether all laws of appearance are contingent.

Philosophy of Neuromorphic AI

The news is brimming with discussions about artificial intelligence, while AI technology is evolving faster than policy makers can respond to it. Although conventional digital microchips have been at the forefront of AI for at least the last seven decades, new designs are being explored that might revolutionize the field. With Moore's Law reaching its limits and facing constraints in speed and energy efficiency, computer engineers are redirecting their focus towards systems beyond the von Neumann architecture. Chips that operate with mechanisms akin to those found in the brain have garnered significant interest. Known as neuromorphic hardware, these chips consist of interconnected physical "neurons" and "synapses," offering the promise of reducing power consumption in large AI systems by two orders of magnitude. Consequently, they hold the potential to replace conventional AI implementations within the next few decades. While some neuromorphic chips represent cutting-edge basic research, others are being developed by for-profit technology companies such as Intel or IBM. Philosophers are only now beginning to contemplate the potential implications of hardware structure that replicate aspects of neural mechanisms in the mammalian brain. The proposed symposium could serve to stimulate debate on neuromorphic technology's philosophical implications, as well as the potential risks and benefits of using neuromorphic technologies.

Symposium Description

Carver Mead (1990) first coined the term "neuromorphic" to describe computational systems of highly interconnected electronic circuits that mimic the neuro-biological architecture of the nervous system. Since then, the meaning of the term has changed and today it applies to a broad set of approaches. Three main areas of research on neuromorphic hardware design can be identified (cf. Indivieri 2021): (1) Artificial neural circuits: The original goal of neuromorphic engineering was to develop systems that directly replicate the physics of biological neural networks, and to build artificial neurons, synapses, and networks using the same organizing principles as the biological brain. This approach survived in the form of non-commercial small-scale research chips that explore different aspects of neural computation ranging from sensory systems (Kramer 2002, Wen & Boahen 2009, Liu & Delbruck 2010) to re-configurable networks with biologically plausible neural dynamics (Benjamin et al. 2014, Qiao et al. 2015), to spike-based learning and plasticity circuits (Mahvash & Parker 2013, Banjeree et al. 2015). Some systems can even be used as neural implants (Abu-Hassan et.al. 2019). (2) Digital very large-scale integration (VLSI) systems: Neuromorphic hardware also comprises mixed-signal or purely digital systems that can be used to simulate neural networks with higher speed and energy efficiency. One example is the EU-funded multi-core computer SpiNNaker developed by the University of Manchester and the Technical University Dresden (Furber et al. 2014). It is built by stacking together 600 circuit-boards, each carrying 48 processors. Similar systems have been developed for commercial purposes by IBM (Merolla et al. 2014) and Intel (Davies et al. 2018). (3) Memristors: Only recently, material and solid-state physicists started using the term "neuromorphic" to describe devices that exhibit different types of behavior akin to those of biological synapses and can be used as building blocks in large-scale AI computing systems.

These so called “memristors” can perform in-memory computation, i.e. processing and storing information in the same location. Different memristive materials are currently under development (e.g. Waser & Aono 2007, Jo et al. 2010, Saïghi et al. 2015). Philosophers are only now beginning to contemplate the potential implications of hardware structures that replicate aspects of neural mechanisms in the mammalian brain. There is a gap between the pace of technological advancement and philosophical reflection concerning neuromorphic AI. While numerous research groups, laboratories, and academic institutions worldwide are diligently exploring and enhancing neuromorphic computing, there appears to be a lack of research articles solely dedicated to its philosophical implications. Furthermore, the few papers that do discuss certain philosophical aspects of neuromorphic engineering seem to have limited interconnections and references to one another. Also, existing work seems to be focused largely on neuromorphic systems from category (2), while neural circuits (1) and memristor devices (3) might be of particular philosophical interest due to their high degree of bio-fidelity. A key component of the proposed symposium is its integration with a special issue (https://journals.ub.uni-koeln.de/index.php/phai/Calls) forthcoming with the recently founded open access journal Philosophy of AI. All confirmed presenters will have submitted a full paper related to their symposium presentation at the time of the symposium. Thus, this symposium could ignite discussions on the philosophical implications of neuromorphic technology, along with the potential risks and advantages associated with its use.

References

Abu-Hassan, K., Taylor, J. D., Morris, P. G., Donati, E., Bortolotto, Z. A., Indiveri, G., & Nogaret, A. (2019). Optimal solid state neurons. Nature Communications, 10(1), 5309. https://doi.org/10.1038/s41467-019-13232-9

Babacan, Y., Yesil, A., Tozlu, O. F., & Kacar, F. (2022). Investigation of STDP mechanisms for memristor circuits. AEU-International Journal of Electronics and Communications, 151, 154230. https://doi.org/10.1016/j.aeue.2022.154230

Banerjee, A., Kar, S., Roy, S., Bhaduri, A., & Basu, A. (2015, May). A current-mode spiking neural classifier with lumped dendritic nonlinearity. In 2015 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 714-717). IEEE. https://doi.org/10.1109/ISCAS.2015.7168745

Benjamin, B. V., Gao, P., McQuinn, E., Choudhary, S., Chandrasekaran, A. R., Bussat, J. M., & Boahen, K. (2014). Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. Proceedings of the IEEE, 102(5), 699-716. https://doi.org/10.1109/JPROC.2014.2313565

Davies, M., Srinivasa, N., Lin, T. H., Chinya, G., Cao, Y., Choday, S. H., ... & Wang, H. (2018). Loihi: A neuromorphic manycore processor with on-chip learning. IEEE Micro, 38(1), 82-99. https://doi.org/10.1109/MM.2018.112130359

Furber, S. B., Galluppi, F., Temple, S., & Plana, L. A. (2014). The spinnaker project. Proceedings of the IEEE, 102(5), 652-665. https://doi.org/10.1109/JPROC.2014.2304638

Grindrod, P., & Brennan, M. (2023). Cognition and consciousness entwined. Brain Sciences, 13(6), 872. https://doi.org/10.3390/brainsci13060872

Indiveri, G. (2021). Introducing 'neuromorphic computing and engineering'. Neuromorphic Computing and Engineering, 1(1), 010401. https://doi.org/10.1088/2634-4386/abf8f8

Jo, S. H., Chang, T., Ebong, I., Bhadviya, B. B., Mazumder, P., & Lu, W. (2010). Nanoscale memristor device as synapse in neuromorphic systems. Nano Letters, 10(4), 1297-1301. https://doi.org/10.1021/nl904092h

Kramer, J. (2002). An integrated optical transient sensor. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, 49(9), 612-628. https://doi.org/10.1109/TCSII.2002.804556

Liu, S. C., & Delbruck, T. (2010). Neuromorphic sensory systems. Current Opinion in Neurobiology, 20(3), 288-295. https://doi.org/10.1016/j.conb.2010.03.007

Mahvash, M., & Parker, A. C. (2013). Synaptic variability in a cortical neuromorphic circuit. IEEE Transactions on Neural Networks and Learning Systems, 24(3), 397-409. https://doi.org/10.1109/TNNLS.2012.2236572

Mead, C. (1990). Neuromorphic electronic systems. Proceedings of the IEEE, 78(10), 1629-1636. https://doi.org/10.1109/5.58356

Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., ... & Modha, D. S. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. Science, 345(6197), 668-673. https://doi.org/10.1126/science.1254642

Qiao, N., Mostafa, H., Corradi, F., Osswald, M., Stefanini, F., Sumislawska, D., & Indiveri, G. (2015). A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses. Frontiers in Neuroscience, 9, 141. https://doi.org/10.3389/fnins.2015.00141

Saïghi, S., Mayr, C. G., Serrano-Gotarredona, T., Schmidt, H., Lecerf, G., Tomas, J., ... & Linares-Barranco, B. (2015). Plasticity in memristive devices for spiking neural networks. Frontiers in Neuroscience, 9, 51. https://doi.org/10.3389/fnins.2015.00051

Waser, R., & Aono, M. (2007). Nanoionics-based resistive switching memories. Nature Materials, 6(11), 833-840. https://doi.org/10.1038/nmat2023 Wen, B., & Boahen, K. (2009). A silicon cochlea with active coupling. IEEE Transactions on Biomedical Circuits and Systems, 3(6), 444-455. https://doi.org/10.1109/TBCAS.2009.2029267

Peter Grindrod - What Present Neurotrophic AI Cannot Do

University of Oxford

What do human brains do that neuromorphic computation and neuromorphic AI does not do? By making full simulations of 1B neurones connected within a suitably realistic architecture (a loose network of tighter sub netwroks - modules equlivalent to neural columns in the brain), while incurring neuron-to-neuron transmission delays, we may reverse engineer the whole and

watch as it is subject to all kinds of stimulations. The system's repose is dynamic and results in competitive dynamical “modes” - patterns of activity both across the neural architecture and over time. The same is true when the inner (dense) networks are replaced by k-dimensional clocks (winding dynamics with k independent phases - easier for our ponderous binary processors to compute and simulate). To take advantage of such a neurmophic architecting (perhaps within a future neuromophic chip), neuromorphic AI should exploit the advanatges. Such AI should simulate human performance (not emulate it) with "warts and all": it will be poor at logic but great at making decisions with little or no data, based on "gut feelings"" (consistent and stable modes that precondition the cognitive system to produce more rapid yet restricted outputs). Thinking fast and slow, just like Kahnenamn.

Inês Hipolito - Neuromorphic AI and the "Computational Ether": A Call for Physical Re-foundation

Macquarie University

While neuromorphic AI promises revolutionary advances by emulating brain-like architectures, this contribution questions whether merely mimicking neural structures within prevailing computationalist frameworks can escape the foundational impasses of classical AI. I will argue that much of AI development, including aspirations for neuromorphic systems, remains implicitly tethered to the notion of 'information processing' as a fundamental explanans—a concept that functions as a 'Computational Ether,' a mediating construct necessitated by an underlying separationist ontology between system and world, or substrate and function. Drawing lessons from fundamental physics, particularly Einstein's 'Relativistic Dissipation' of superfluous entities, I will explore how this reliance on a 'Computational Ether' generates persistent conceptual puzzles regarding meaning, understanding, and genuine cognitive integration. A transformative, 'brain-like' AI, from this perspective, may require more than architectural mimicry; it may necessitate a radical physical re-foundation where cognitive dynamics are understood not as computations performed by hardware, but as intrinsic, self-referential field dynamics constituted by the physical system in its relational embedding. This shifts the challenge for neuromorphic AI from optimizing processing efficiency to realizing principles of direct physical constitution.

Derek Shiller - Neuromorphic Hardware and the Prospects of Enbrainment

Rethinking Priorities (New York)

This paper introduces the concept of enbrainment and outlines its relevance to interpretations of consciousness and personal identity. A system is enbrained to the extent that the same physical subsystems consistently perform the same computational roles over time. Contemporary AI systems typically rely on distributed, general-purpose computing

infrastructure and therefore lack the continuity required for enbrainment. In contrast, human brains exhibit enbrainment through functional specialization and anatomical persistence. Neuromorphic approaches to computing promote enbrainment by incorporating persistent, functionally specific hardware structures. Future AI paradigms may diverge along two paths: one favoring flexible, distributed, non-enbrained architectures, and another centered on dedicated, local, enbrained hardware. Given enbrainment's potential relevance to cognition and consciousness, these trajectories may carry significantly different philosophical implications.

Wanja Wiese - Does Consciousness Require Neuromorphic Processing?

Ruhr University Bochum

This talk first defines two notions of neuromorphic processing: a relatively general notion, and a slightly more specific notion. Based on the more general notion of neuromorphic processing, I introduce the “detachability criterion”: a criterion to assess whether simulated agents that are part of a computer simulation can be conscious. Although I provide an intuitive motivation for the criterion, my main aim is to argue for the following conditional claim: IF conscious experience in a computer simulation requires satisfying the detachability criterion and neuromorphic processing in the more general sense, THEN conscious experience (in computer simulations, robots, or other artificial systems) also requires neuromorphic processing in the more specific sense.

Realism in Neuroscience

Given that scientific realism is a perennial issue in the philosophy of science, it should come as no surprise that it manifests itself in the philosophy of neuroscience. This symposium explores some of these manifestations. A principal locus of realism debate in the philosophy of neuroscience concerns realism about neural/mental representations. This, in turn, is intertwined with the view that neural/mental representations are real insofar as they are hypothesized to explain behavior. Another locus of realism concerns the putative computational functions arising with computational models of cognition. In her talk, Egan proposes that, while representational vehicles are real, content is not. Content is not real, since content attributing explanations have an irreducible pragmatic component. This component is inconsistent with there being purely objective grounds for content attribution. In his talk, Piccinini defends content realism. He proposes that neural content is not hypothesized in order to explain behavior. Instead, neuroscientists observe content. This approach promises to sidestep Egan's critique of content realism. In his talk, Aizawa draws attention to cases that seem to be problematic for Piccinini's account of neural content. These are cases in which neuroscientists combine both hypothetical and "observational" empirical methods. Finally, in her talk, Williams poses a challenge for computational realism in artificial neural networks. Although ANNs are specified by the functional interactions of individual "neurons," there is generally no specification of the function a complete ANN computes. Thus, problematically, there is no epistemically accessible function for the realist to be a realist about.

Ken Aizawa - On the Evidence for Neural Representations

Rutgers University, Newark

Philosophers of psychology often note that some scientists believe in mental representations because those representations can explain behavior. Edward Tolman, for example, believed a rat's possession of a cognitive map explained some of that rat's navigation behavior. Philosophers of neuroscience also sometimes observe that neuroscientists believe that single-unit firing rates represent environmental features, in part, because of correlations between the environmental features and the single-unit firing rates. John O'Keefe and Jonathan Dostrovsky, for example, observed that the firing rates of certain hippocampal cells—later called "place cells"—correlate with a rat's orientation on a platform. Less frequently, philosophers have observed that sometimes neuroscientists envision linking the results of behavioral experiments and neurophysiological experiments. O'Keefe and Dostrovsky, for example, believed there was a connection between Tolman's work and their own. Citing Tolman's work, they proposed that specific deficits in rat behavior "could be due to the loss of the neural system which provides the animal with a cognitive, or spatial, map of its environment. Our preliminary observations on the behaviour of hippocampal units in the freely moving rat provide support for this theory of hippocampal function" (O'Keefe & Dostrovsky, 1971, p. 171). I provide two illustrations of how scientists have combined behavioral evidence with neurophysiological

evidence to support the existence of neural representations. The first will be one of the simplest I could find involving simultaneous contrast. The second will be for representations involved in retinal processing surrounding the blind spot in humans. This will bring to three the number of illustrations of the combination of behavioral and neurophysiological evidence for neural representations.

References

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. Brain Research, 34, 171-175.

Frances Egan - What Does Representational Realism Require?

Rutgers University

It is often said that we should be realist about the posits of our best scientific theories. The scientific realist's suggestion is not as straightforward as it may seem. There is always an issue about which aspects of an empirical theory should be interpreted realistically, and which aspects should be construed rather as features of the scheme used to articulate the theory. This is especially true for theories purporting to explain our cognitive capacities. Many theories in cognitive neuroscience posit neural representations that play causal roles in cognitive processes, so we should be realist about them. But what exactly does representational realism require? A commitment to representations presumes a distinction between representational vehicle and representational content. Vehicles are concreta, such things as symbols in a language of thought or characteristic patterns of activation of a neural network. They are items over which causal processes that underly cognition are defined, so we should be realist about representational vehicles. Contents, on the other hand, are abstracta. Content is the meaning "carried by" the vehicle (to employ a useful metaphor) and is outside the causal order. Though content is often said to be "causally relevant," the standard argument for scientific realism does not apply to it. Content itself is epiphenomenal. I suggest that realism about content is best understood as the idea that there are purely objective grounds for the attribution of representational content. I have argued (Egan 2025) that content realism in this sense is false; there are irreducible pragmatic determinants of content. In this talk I argue (1) that a representational realist need not endorse content realism, and (2) that we can make clear sense of the idea that content is causally relevant without being realist about it.

References

Egan, F. (2025). Deflating Mental Representation, MIT Press.

Gualtiero Piccinini - How to Observe Mental Representations, their Semantic Content, and the Causal Efficacy of their Content

University of Missouri

Mental representations, insofar as they exist, are realized in neural tissue. They are (a kind of) what neuroscientists call neural representations (cf. Piccinini 2020a, b, 2022, 2025; Piccinini and Hetherington 2025). Paradigmatically, neural representations are a kind structural representations, that is, a system of internal states with at least the following four properties: (1) they form a system structurally similar to their target, meaning that relations between internal states map onto relations between targets; (2) active states maximize mutual information with their target, usually via a causal connection (Heemskerk 2025); (3) they are decouplable from their target (meaning, they maximize mutual information with their target even in the absence of a causal connection); and (4) their structural similarity to their target allows them to guide behavior with respect to their target. The content of structural representations and their causal efficacy is a function of those four properties. Following Nestor (2024), paradigmatic structural representations have an iconic content and an indexical content. The iconic content is the structure of the representation, which is "predicated" of the target; the indexical content is the target. To Nestor's account, I add that some specialized mental/neural representations also have symbolic content, which is a secondary, arbitrary target a representation inherits from (sensorimotor) representations it activates and is activated by. Accordingly, the causal efficacy of iconic, indexical, and symbolic contents is the causal efficacy of the representational structure, primary target, and (if applicable) secondary target, respectively, on the system and its behavior. All the above items (representational structure, targets, mutual information, decouplability, behavioral guidance) are empirically observable. Therefore, mental representations, their semantic content, and the causal efficacy of their content are empirically observable (and have been observed in many neurocognitive systems).

References

Hemskeerk, J. (2025). How Informational Teleosemantics Works: Towards a Realist Theory of Content. PhD Dissertation. University of Warwick.

Nestor, M. J. (2024). From the Body to the World: Structural Representation and the Mapping Problem. PhD Dissertation. University of Adelaide.

Piccinini, G. (2020a). Neurocognitive Mechanisms: Explaining Biological Cognition. Oxford: Oxford University Press.

Piccinini, G. (2020b). "Nonnatural Mental Representation," in K. Dolega, T. Schlicht, J. Smortchkova, eds., What Are Mental Representations? Oxford: Oxford University Press, pp. 254-286.

Piccinini, G. (2022). "Situated Neural Representations: Solving the Problems of Content," Frontiers in Neurorobotics 16: 846979.

Piccinini, G. (2025). "Neural Hardware for the Language of Thought." Unpublished manuscript.

Piccinini, G., and Hetherington, S. (2025). "Knowing That as Knowing How: A Neurocognitive Practicalism." Synthese 205 (5). https://doi.org/10.1007/s11229-024-04843-1

Danielle J. Williams - An epistemic problem for the realist about mathematical functions in neuroscience

Washington University In St. Louis

Computational neuroscience aims to identify the unknown functions used to perform cognitive tasks. It has been argued that we can interpret which computation (mathematical function) the brain performs based on a computational theory (Shagrir 2022). Such a view is realist about which mathematical functions the brain performs (Williams 2025; Egan 2025). I will argue, however, that when it comes to neural network models (ANNs), the realist faces an epistemic problem: the model relies on an unknown mathematical function to solve the task. Because the computational theory only provides the mathematical functions used to define the model, and not which mathematical function the model uses to solve the problem, the realist about mathematical functions should not be a realist about the function given within the computational theory, as it is not a solution to the cognitive problem targeted by the modeler. ANNs rely heavily on the use of mathematical functions—there are functions that are used to define every part of an ANN and how it works, including its architectural details. The mathematical functions mathematically define the model and are often explicitly stated by modelers when describing their results. However, these functions are not the ones that the model uses to solve the task. Instead, to solve the task, the model relies on an unknown mathematical function that it develops through training. This is precisely what makes ANNs so powerful: so long as they are trained with enough data, they can approximate any possible mathematical function—as opposed to being "programmed" with a solution beforehand. A consequence of this flexibility is that the solution used by the ANN is epistemically opaque to the modeler (Humphreys 2009), giving rise to an epistemic problem for the realist about mathematical functions: there is no epistemically accessible function for the realist to be a realist about.

References

Egan, F. (2025). Deflating Mental Representation. MIT Press.

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. Synthese

Shagrir, O. (2022). That Nature of Physical Computation. Oxford University Press

Williams, D. (2025). How to be a realist about computational neuroscience. Synthese


ISPSM