Models of Consciousness 6


上海科技大学
ShanghaiTech University




















Organising and Supporting Institutions:

Association for Mathematical Consciousness Science

上海科技大学
ShanghaiTech University

CHAIN
Center for Human Nature, Artificial Intelligence, and Neuroscience

BAMΞ

University of Bamberg

UNIVERSITY OF OXFORD

SAINT ANSELM COLLEGE
1889

RUB

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

Barbara Wengeler Stiftung

OMCAN
Oxford Mathematics of Consciousness and Applications Network

Programme

Table of Contents

Welcome ____ 4

Everything you need to know ____ 5

The conference team ____ 6

Excursion ____ 8

Conference dinner ____ 10

Discussion sessions ____ 12

Conference venue ____ 13

Floor maps ____ 14

Area maps ____ 15

Programme schedule ____ 16

Talk titles and abstracts ____ 24

Welcome!

It is a great pleasure to welcome you all to the sixth instalment of the *Models of Consciousness* conference series (MoC6). Following the conferences at *Oxford University* (2019, MoC1), online (2021, MoC2), *Stanford University* (2022, MoC3), again *Oxford University* (2023, MoC4), and Bamberg University (2024, MoC5), it is a great pleasure to host this year's conference at *Hokkaido University*, by kind invitation of Hokkaido University's *Center for Human Nature, Artificial Intelligence, and Neuroscience* (CHAIN).

Many approaches in consciousness science focus on the philosophical underpinnings of models of consciousness and the application of mathematics to make phenomenology more tractable. Apart from this, there is an ongoing interest in quantum approaches to consciousness, which differ in central respects from classical approaches. As advancements in Artificial Intelligence (AI) continue to accelerate, there has also been a growing body of work dedicated to investigating the possibility of artificial consciousness.

This conference aims to bring these developments together, in hopes of mutual inspiration and constructive interference. Correspondingly, this year's core themes have been chosen as:

- Philosophy of Consciousness and Mathematized Phenomenology
- Quantum Approaches to Consciousness and their Assumptions
- Research on Artificial Consciousness

We hope to create an open and inspiring atmosphere that enables us all to make further progress on these and other questions together. Thanks for contributing your talks and discussions!

Welcome to Sapporo! — *The organisers*

Everything you need to know...

... initially ;-)

Conference Venue

Clark Memorial Student Centre, Hokkaido University
Address:
7 Chome Kita 8 Jonishi, Kita Ward,
Sapporo, Hokkaido 060-0808, Japan

Google Maps Link - OpenStreetMap Link

Wi-Fi Access:

You can either use **eduroam** or
receive a Wi-Fi guest account on site.

Conference Team

Organisers

Don‘t hesitate to let us know in case you have any questions or are in any need of assistance!



Robert Prentner
Institute of Humanities, ShanghaiTech University



Keisuke Suzuki
Center for Human Nature, Artificial Intelligence, and Neuroscience (CHAIN)
Hokkaido University, Japan





Johannes Kleiner
Institute for Psychology, University of Bamberg
Munich Center for Mathematical Philosophy, Ludwig Maximilian University of Munich

Joanna Szczotka
Wisconsin Institute for Sleep and Consciousness, University of Wisconsin-Madison





Wanja Wiese
Institute for Philosophy II, Ruhr University Bochum

Camilo Miguel Signorelli
Department of Communication,
University of Copenhagen





Jonathan Mason
Oxford Mathematics of Consciousness and Applications Network

Robin Lorenz
Quantinuum

Advisory Board



Ian Durham

Department of Physics, Saint Anselm College



Lenore Blum

Distinguished Career Professor of Computer Science Emerita
Carnegie Mellon University



Yakov Kremnitzer

Oxford Mathematics of Consciousness and Applications Network
Mathematical Institute, University of Oxford

Excursion

In tradition of previous conferences, the first day of the conference is devoted to an excursion designed for participants to connect and discuss their ideas, research and interests on a one-on-one basis.

Option 1:

We've chosen a suitable hiking trail on Mt. Moiwa with a beautiful view.

Departure at **10:30** at **Hokkaido University South Gate**. Bus Departure for return: **14:30**.

One bottle of water per person and a lunch box will be provided at departure.

Hokkaido University South Gate (Map)



© Adam Jones CC Attribution-Share Alike 2.0 Generic

Option 2:

Due to temporary closure of the Hokkaido Museum of Modern Art, we recommend participants to visit the **Botanic Garden**, Hokkaido University, in close proximity to the conference venue (self-organized), **Address:** North3, West8, Chuo-ku, Sapporo, 060-0003.



© 663highland CC Attribution-Share Alike 3.0

Hiking Trail Schedule

Time	Activity	Location	Duration
10:00	Meet & Check-in	Hokkaido University South Gate	-
10:30	Bus Departure	Same as above	30 min
11:00	Begin Hike	Mount Moiwa Hiking Walking trail - Start	80 min
12:20	Lunch at Summit & Free Time	Mount Moiwa Peak	2h 10m
14:30	Bus Departure for Return	Mount Moiwa Hiking Walking trail - Start	30 min
15:00	Arrive Back	Hokkaido University South Gate	-

Hiking Route

Mount Moiwa: Jikeikai Byoin Mae Course

Length: 5.4 km
Elevation gain: 543 m
Estimated time: 2.5 - 3 hours



You Miss the 14:30 Return Bus:

You must use public transport to return to Sapporo Station from the nearby **Keimei Bus Terminal (啓明バスターミナル)**.

- **Option A (Direct):** Bus **Route 51** → Sapporo Station (~30 min).
- **Option B (Transfer):** Bus **Circle Route 66** → Horohirabashi Stn. (~6 min), then Subway **Namboku Line** → Sapporo Station (~6 min).

More details about the excursion will be distributed by email to all those who have signed up.

Conference Dinner

You are invited to join the conference dinner at a charcoal grill and sushi restaurant, serving sushi, seafood, hand-made Tofu, chicken BBQ, and more!
(By registration only.)







Restaurant: **Maido! Sapporo Eki Minamiguchi Ten**

「炭火焼と海、手づくり豆富 まいど！ 札幌駅南口店」

Address: 〒060-0004 Hokkaido, Sapporo, Chuo Ward, Kita 4 Jonishi, 2 Chome, Kitako Trade BLD, 2F.

Map

Day: **Thursday, 2 October**

Time: Starting **7:00 p.m.**

Payment: *Please note that the conference dinner is **only** included if you registered and paid for the dinner when paying the conference fee.*

Maido! Sapporo Eki Minamiguchi Ten

炭火焼と海、手づくり豆富 まいど！ 札幌駅南口店



Discussion Sessions

To create an open and friendly atmosphere in which thoughts and ideas can flow and, hopefully, the spark of inspiration strike, this conference features discussion sessions in small groups. They give participants a chance to discuss open questions in the field and are intended to help develop a shared sense of purpose and long-term goals among researchers who work on new topics in consciousness science.

Each discussion session focuses on a specific question related to consciousness science. In order to let those questions emerge that receive the most interest by all participants, we are using a web interface that allows attendees to propose questions, and to indicate their interest in questions that others have proposed.

In order to access this web interface, simply access the following website with any internet device (laptop, cell phone, tablet)

www.slido.com

and enter the event code

#MoC6

for “Models of Consciousness 6”.



We hope discussion sessions at the conference are enjoyable and fruitful! Ultimately, this requires an atmosphere of trust and tolerance supported by all attendees of a session, much like described in the two quotes by Isaac Asimov and a collaborator of Werner Heisenberg on the back of this booklet.



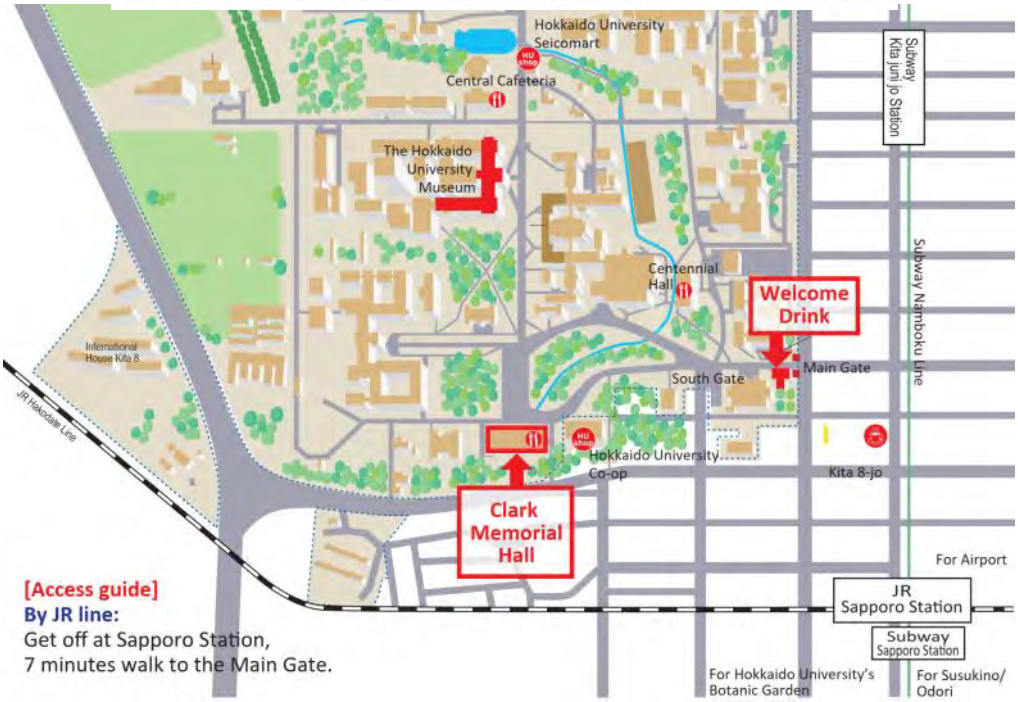
How to use the web interface?

→ In order to **indicate that you find a question interesting**, click the small “thumbs up” on the right hand side of a question.

→ In order to **add a new discussion topic**, simply enter the question into the field at the top and press “send”.

Conference Venue





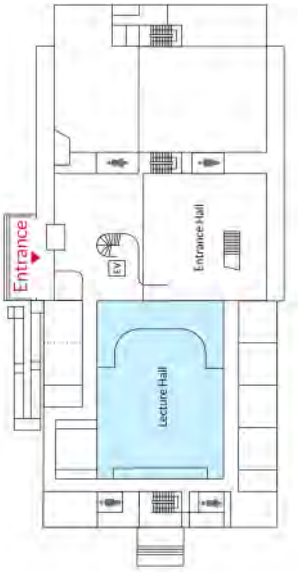
Clark Memorial Hall

***Address*:**
Hokkaido University, 7 Chome Kita 8 Jonishi,
Kita Ward, Sapporo, Hokkaido 060-0808, Japan

Floor Map



Floor 1



Floor 2



Floor 3

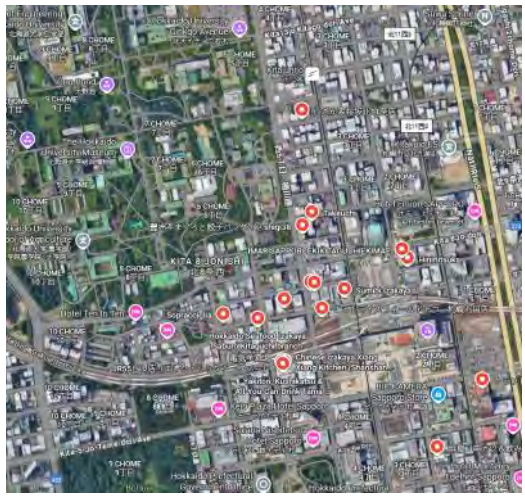
Clark Memorial Hall

***Address*:**
Hokkaido University, 7 Chome Kita 8 Jonishi, Kita Ward, Sapporo, Hokkaido 060-0808, Japan

Lunch around Hokkaido University



Bars & Restaurants



Vegan & Vege-tarian Options



Wednesday, 1 October

Time	Lecture Hall
8:30 - 9:00	Welcome & Registration
9:00 - 9:15	Opening Remarks
09:15 - 10:15	**Invited Talk: Ryota Kanai**
10:15 - 11:15	Discussion
11:15 - 11:45	Coffee Break
11:45 - 12:35	**Parallel Session: Conscious Machines** – Joscha Bach – Michael Timothy Bennett
12:35 - 14:00	Lunch
14:00 - 15:00	**Invited Talk: Fernando Rosas**
15:00 - 15:45	Discussion
15:45 - 16:15	Coffee Break
16:15 - 17:30	**Parallel Session: Beyond the Substrate** – Hikari Sorensen – Johannes Kleiner – Wanja Wiese
18:00 -	Drinks Reception

Time	Meeting Room 1	Meeting Room 2
Welcome & Registration		
Coffee Break		
11:45 - 12:35	**Parallel Session:** **Computational Phenomenology** – Will Yun-Farmbrough – Joanna Szczotka	**Parallel Session:** **Embodied Dynamics** – Liya Zou – Alexander Hölken
Lunch		
Coffee Break		
16:15 - 17:30	**Parallel Session:** **Formal Approaches** – Ryuzo Hirota – Sophie Taylor – Francesco Giorlando	**Parallel Session:** **Rethinking the Intentional** – Adel Chaibi – Patrick Gruneberg – Adam Rostowski
Drinks Reception		

SCHEDULE

Thursday, 2 October

Time	Lecture Hall
09:00 - 10:00	**Invited Talk: Takashi Ikegami**
10:00 - 11:00	Discussion
11:00 - 11:30	Coffee Break
11:30 - 12:40	**Parallel Session: Relations and Metaphysics** – Robert Prentner – Zoe Lee-Youngzie – Etienne Jacques – Gustaf Malmberg
12:40 - 14:00	Lunch
14:00 - 15:00	**Invited Talk: Tom Froese**
15:00 - 15:45	Discussion
15:45 - 16:15	Coffee Break
16:15 - 16:45	**Showcase Talk: Sid Kouider**
16:45 - 17:35	**Parallel Session: Quantum and Brain** – Chris Rourk – Joachim Keppler
19:00 -	Conference Dinner

Time	Meeting Room 1	Meeting Room 2
Coffee Break		
11:30 - 12:40	**Parallel Session:** **Physics and Phenomenology** – Ivan Serwano – Kohtaro Tadaki – Ian Durham – Paul Skokowski	**Parallel Session:** **Architectures for AI Consciousness** – Erik Nemeth – F. Rodrigues-Vergara – Igor Balaz – Ben Gaskin
Lunch		
Coffee Break		
16:45 - 17:35	**Parallel Session:** **Attribution and Representation** – Anna Eiserbeck – Francesco Lassig	**Parallel Session:** **Functionalist Interpretations** – Anna Eiserbeck – Nathaniel Virgo
Conference Dinner		

SCHEDULE

Friday, 3 October

Time	Lecture Hall
09:00 - 10:00	**Invited Talk: Susan Schneider**
10:00 - 11:00	Discussion
11:00 - 11:30	Coffee Break
11:30 - 12:45	**Special Session: Quantum Cognition Mini Session** – Peter Bruza – Michael Schnabel – Yuko Ishihara
12:40 - 14:00	Lunch
14:00 - 15:00	**Invited Talk: Shigeru Taguchi**
15:00 - 15:45	Discussion
15:45 - 16:15	Coffee Break
16:15 - 17:15	**Invited Talk: Miguel Aguilera**
17:15 - 18:10	**Parallel Session: Quantum Participation** – Valtteri Arstila – Chetan Prakash – Renaud Gauthier

Time	Meeting Room 1	Meeting Room 2
Coffee Break		
Lunch		
Coffee Break		
17:15 - 18:10	**Parallel Session:** **Structures of the Self** – Yuria Shimizu – Anya Daly – Elay Shech	**Parallel Session:** **Algorithmic Consciousness** – Jad Tarifi – Giulio Ruffini – Francesca Castaldo

SCHEDULE

Saturday, 4 October

Time	Meeting Room 1
09:00 - 10:00	**Invited Talk: Yuko Ishihara**
10:00 - 11:00	Discussion
11:00 - 11:30	Coffee Break
11:30 - 12:00	**Showcase Talk: Moritz Kriegleder**
12:00 - 12:50	**Parallel Session:** **Entropy and Consciousness** – Yifeng Chen – Jeff Sanders
12:50 - 14:00	Lunch
14:00 - 14:30	**Showcase Talk: Manuel Baltieri**
14:30 - 15:20	**Parallel Session:** **Philosophical Topics** – Nicolas Loerbroks – N. M. K. Tütüncüoğlu
15:20 - 15:50	Coffee Break
15:50 - 16:20	**Showcase Talk: Kobi Kremnitzer**
16:20 - 17:00	Closing Remarks

Time	Meeting Room 2
Coffee Break	
12:00 - 12:50	**Parallel Session:** **Quantum Cognition** – Masanao Ozawa – Sean Tull
Lunch	
14:30 - 15:20	**Parallel Session:** **Geometry of Hallucination** – Paweł Motyka – Trevor Hewitt
Coffee Break	

SCHEDULE

Titles and Abstracts

1 Ryota Kanai (invited talk)

Araya Inc., Japan

Qualia and Symmetry

The subjective qualities of conscious experience, known as qualia, pose a fundamental challenge to scientific explanation. In this talk, I will present a novel mathematical framework, developed with my collaborators Masafumi Oizumi and Chanseok Lim, that characterizes the structure of qualia through the lens of symmetry and principal bundle geometry. Our central proposal is that for a G-equivariant neural network, its state space is organized by a principal bundle structure, giving rise to a three-level hierarchy of qualia. First, the algebraic structure of the symmetry group G itself defines the fundamental nature of a qualia modality (e.g., vision versus audition). Second, a "qualia signature," representing the invariant identity of a percept (e.g., "cat"), corresponds to a point in the bundle's quotient space. Third, a "qualia attribute," representing continuous variations of that percept (e.g., the cat's location), is a point on a specific orbit induced by the group's action. This framework reveals a fundamental duality: the geometry of qualia attributes is topologically rigid, inheriting its structure directly from the symmetry group G, which explains their universal and stable nature across individuals. In contrast, the geometry of the quotient space, which encodes qualia signatures, is plastic and can be shaped by an individual's learning and experience. We further speculate that the compositionality of experience can be modeled through hierarchical or product structures of these bundles. This geometric approach provides a rigorous language for describing subjective experience, generates empirically testable predictions for comparing qualia structures, and offers a systematic research program for a scientific understanding of consciousness.

2 Joscha Bach

California Institute for Machine Consciousness, USA

Request Confirmation Networks as a Model of Consciousness

Consciousness may be characterized by its phenomenology and its functionality (how do mental states change in the presence of consciousness?). We explore the hypothesis that consciousness is functionally a causal pattern that establishes and increases coherence in mental representations within working memory, similarly to how a conductor establishes coherence in the music played by the different instruments of an orchestra. We understand coherence as the elimination of constraint violations across currently active perceptual representations, and propose a computational model of bottom-up/top-down perception that interacts with lateral constraint propagation, creating a working memory state that is coherent with current perceptual content and action control: the contents of a perceived present, which is reflected in a simulated observer. Request Conformation Networks (ReCoNs) are a model of how to achieve perceptual binding and working memory synchronization by periodic propagation of spreading activation in a recurrent paradigm. A ReCoN is composed of nodes and a small number of link types, where each node implements a state machine, facilitating binding, learning and execution of perception and action routines.
Our work on Request Confirmation Networks is part of a larger project to understand the mechanisms of consciousness in a computational paradigm, using testable simulation models.

3 Michael Timothy Bennett

How To Build Conscious Machines

I present a novel formalism for understanding existence an infinite stack of abstraction layers, where each layer embodies policies constraining possible worlds. I use this model to explore the origins of life and consciousness. I argue those systems which are alive are those that maximise weakness of constraints on function, while not maximising simplicity. This allows for complex information processing. I argue that biological systems function as bioelectric polycomputers, with consciousness emerging from complex tapestries of valence that classify objects and properties. These tapestries integrate representation and value judgement. I propose the psychophysical principle of causality, arguing qualia are an abstraction layer formed of these tapestries. I argue phenomenal consciousness begins at the first-order self, while conscious access requires second-order selves. This framework connects intelligence and consciousness, offering insights into artificial consciousness. It also suggests a surprising relationship between identity, language, cancer and the Fermi paradox. Finally, I discuss an unresolved question I call “The Temporal Gap”, which is whether consciousness must be at a point in time but extended in space, or can be “smeared across time” like in a computer. This work integrates theoretical, experimental, and philosophical perspectives to advance our understanding of consciousness and its artificial replication.

4 Will Yun-Farmbrough

Sussex Centre for Consciousness Science, School of Engineering & Infor-

Perception as Inference: from Predictive Coding to the Meta-Problem of Consciousness

This abstract addresses perceptual inference in two parts: first, computational phenomenology with predictive coding (PC) networks; second, implications for the meta-problem of consciousness.

i) PC networks are deep generative models that implement perception through local error minimisation—flexibly inverting a learned model to infer the causes of sensory input. While these iterative dynamics emulate recurrent processing evident in cortex, they fail to capture the fast, feedforward sweep that initiates perception. Hybrid PC networks address this by introducing bottom-up weights trained to approximate converged posteriors directly from data, enabling rapid, amortised inference followed by iterative refinement. This strategy mirrors hybrid approaches in machine learning and captures phenomenological aspects of gist perception, suggesting that conscious contents may arise from both bottom-up and top-down predictions.

ii) Hierarchical inference schemes imply that inferred perceptual representations may themselves become the basis for abstracted understanding in higher-level models. Intriguingly, such frameworks offer compelling accounts of distinct phenomenological aspects. However, the explanatory domain of higher-level models may exclude concepts corresponding to basic perceptual representations—understanding of a physical world built on perceptual primitives need not explain the nature of primitives themselves. With the capacity to reflect on past states, such systems may inevitably generate intuitions of qualia.

5 Joanna Szczotka

Center for Sleep and Consciousness,

The hard problem of valence – towards formalizing a missing piece in theories of consciousness

In recent decades, consciousness science has witnessed the rapid growth of theoretical frameworks, yet some argue that genuine progress has remained elusive. One reason for this perceived stagnation is that different models often target subtly different phenomena, creating a fragmented and disjointed landscape of inquiry. With AI advancing rapidly, the lack of consensus on foundational aspects is increasingly concerning. Against this background, I argue for an approach that centralizes a particular problem that so far remained largely on the back burner in consciousness science: the problem of valence – the positive or negative quality of conscious experience. Valence has often been sidelined as either too trivial or too complex at the current stage of the field. Instead, many theories have focused on more tractable questions of sensory detection, contrasting “stimulus seen” versus “stimulus unseen” brain activity. The prevailing assumption has been that we must first determine what makes a sensory stimulus conscious, and only then address supplementary problems, such as what makes a conscious stimulus evaluated as good or bad. However, this sequential approach risks obscuring how deeply intertwined mechanisms of sensory detection and evaluation are. Evidence suggests that certain implicit evaluative processes not only strongly influence the gating of sensory stimuli into consciousness but also dictate the temporal dynamics of the conscious stream, challenging the view that valence is merely an “add-on” to sensory processing. In this talk I synthesize scattered empirical clues under a unifying framework that maps reinforcement learning(RL) components onto the basal-ganglia -thalamo-cortical loop, complementing and extending cortico-centric accounts. Lastly, I argue that theories of consciousness can no longer treat valence as a peripheral issue: incorporating it is essential not only for a more complete account of the NCCs, but also for confronting the most ethically salient dimension of consciousness.

6 **Liya Zou**
Higher Institute of Philosophy (HIW), KU Leuven, Belgium

Desire in Movement: A Phenomenological Approach to Predictive Processing Theory Based Sensorimotor Theory and Embodied Cognition

This paper examines the ontological limitations of predictive processing-based sensorimotor theory in explaining conscious experience and argues that integrating the phenomenology of life can address these shortcomings. With recent developments in computational and cognitive neuroscience, Anil Seth integrates predictive processing with Degenaar and O’Regan’s sensorimotor theory and develops the predictive processing theory of sensorimotor contingencies (PPSMCs) to explain perception in terms of the brain’s ongoing predictions and their continuous updating through sensory feedback. While PPSMCs provide a computationally precise account of perception grounded in sensorimotor contingencies and predictive mechanisms, they do not address the deeper dynamics of living experience. Drawing on Renaud Barbaras’ phenomenology of life, which defines life as desire—a movement that transcends itself—this paper contends that perception must be understood as part of life’s fundamental striving. Barbaras’ notion of life as a movement of negation provides a framework for reinterpreting perceptual experience as a totalising totality between organism and environment, driven by desire rather than mere predictive sensorimotor coordination. By incorporating this perspective, PPSMCs-based ST can overcome its limitations concerning motility and embodiment, laying the groundwork for future investigations into desire’s role in conscious experience.

7 **Alexander Hölken**
Center for Mind and Cognition, Institute for Philosophy II, Ruhr Universität Bochum, Germany

Nested Brain-Body-Environment Systems – Extending Dynamical Emergence Theory to Identify Realizers of Conscious Experience

Throughout the last 15 years, Tomer Fekete & Shimon Edelman have developed a mathematical theory of phenomenal consciousness which models it as a continuous process occurring in dynamical systems with a minimum amount of representational capacity (Fekete & Edelman, 2011). According to their Dynamical Emergence Theory (DET), phenomenal experience can be formalized as a system's trajectory through a set of coarse-grained macrostates defined over the collective dynamics of its implementational substrate – that is, its neural realizers (Moyal, Fekete & Edelman, 2020).
In my talk, I identify issues with two of DET's central axioms: First, the characterization of experiential trajectories as series of discrete conscious states, and second, its focus on neural processes as the sole substrates of conscious experience. To solve these issues, I propose a fundamentally procedural version of DET based on the principles of coordination dynamics (Kelso & Engstrøm, 2007) which does not decompose phenomenal experience into discrete experiential macrostates, and whose methodology may be applied to non-neural systems as well. I then show that this version allows the application of DET to a wider range of conscious phenomena (such as rubber-hand-type illusions) and its integration into current research paradigms within ecological psychology and embodied cognitive science.

8 Fernando Rosas (invited talk)

School of Engineering and Informatics,
University of Sussex, United Kingdom

Consciousness and emergence: Formalisms, opportunities and challenges

Emergence is one of the most fascinating and challenging aspects of complex systems, being also a controversial topic featuring long-standing debates and disagreements. In this talk I'll introduce a pluralistic and pragmatic stance towards emergence, and explore its potential to illuminate our understanding of consciosness. The first part of the talk will argue for formal approaches to emergence, which give us tools to establish falsifiable hypotheses and procedures to verify them. We will then discuss how these methods can be used to investigate various aspects of consciousness both theoretically and empirically across different biological and artificial systems.

9 Hikari Sorensen

California Institute for Machine Consciousness, USA

What Is Consciousness a Property Of? Computational Functionalism and the Epistemology of Machine Mind

Discussions of machine consciousness are often muddled by implicit metaphysical assumptions. This talk examines the epistemology and metaphysics of computational functionalism, a position that rejects biologically essentialist views of consciousness and instead treats it as an implementable property of certain information-processing systems.
We argue that consciousness should be understood not as a substance or essence, but as a functional architecture characterized by recursive self-modeling, representational coherence, and adaptive control; we decouple consciousness from its implementation medium, viewing it instead as defined by its functional organization. On this view, consciousness is not limited to biological organisms but can, in principle, arise in any system that implements the relevant dynamics. We contrast this with common forms of naive biologism, which take substrate (e.g. neurons, carbon-based life) as explanatorily central rather than contingent. We motivate the need for an operational definition of consciousness that makes minimal ontological commitments and favors explanatory power and parsimony.

Finally, we address the challenges of empirical testing in the absence of direct access to phenomenology. We propose that rather than relying on indirect behavioral correlates or intuition, scientific inquiry should prioritize formal architectures and computational properties.

10 **Johannes Kleiner**
Institute for Psychology, University of Bamberg, Germany

Mathematics and the Real Correlates of Consciousness

In this talk, I would like to sketch what I think could be a new area of research in Mathematical Consciousness Science: the development of mathematics that support the empirical search for Neural Correlates of Consciousness, Computational Correlates of Consciousness, Physical Correlates of Consciousness, and related notions, which for brevity I will refer to as Real Correlates of Consciousness (RCCs), to have a substrate-neutral term. In the talk I will explain:

1. Why I think that the empirical search for RCCs is essential to make progress in consciousness science across all metaphysical and theoretical fronts
2. Which definitions of RCCs exist and how these are measured at the present
3. What role mathematics plays in contemporary measurement schemes
4. Why and where the development of new mathematics is possible, and arguably necessary, for the empirical discovery of RCCs.

As an example of 4., I will review a recent mathematical pilot study aimed at the canonical definition of NCCs.

11 **Wanja Wiese**
Institute for Philosophy II, Ruhr University Bochum, Germany

Is the problem of artificial consciousness scientifically tractable?

Can science meaningfully contribute to questions about consciousness in artificial systems? If yes, does this require presupposing a particular view on consciousness, such as computational functionalism?
I suggest treating these questions themselves as empirical scientific questions. To answer them, we first have to identify relevant fundamental issues about the nature of consciousness, such as the following:
i) What features of living organisms, if any, are required for consciousness?
ii) What differences between physical agents (in non-virtual environments) and virtual agents (in computer simulations), if any, matter for consciousness?
I argue that science can meaningfully contribute to these questions, by providing rigorous accounts of potentially relevant features of living organisms, and differences between virtual and non-virtual agents. Specifying these features and differences, ideally mathematically, is the first step towards answering questions like (i) and (ii). Going forward, science can at least contribute to answers by making these questions more intelligible and tractable.

12 **Ryuzo Hirota**
Graduate School of Arts and Sciences, University of Tokoyo, Japan

The Monoid-I: a category-theoretic approach to the structure of phenomenological self

The phenomenological understanding of selfhood poses a distinctive challenge: while the self is always present as the pole of one's experiences, it is not encountered in the same way as objects in the world. This subtle and seemingly paradoxical nature – being ever-given yet non-

objectifiable – has made its integration into empirical and scientific frameworks particularly problematic. Drawing on Husserl's phenomenological analyses of the ego, this presentation proposes a category-theoretic formalisation that reframes selfhood in terms of monoid structures. In category theory, a monoid is understood as a category with a single object, where all arrows originate from and return to that object, and are all mutually composable. Analogously, we argue that the phenomenological self can be conceived not as a pre-given entity nor as a mere bundle of experiences, but as a structural condition in which diverse experiences are rendered mutually composable through the identity of the self, which is itself devoid of any content. This approach aims to capture the unique structural characteristics of selfhood revealed by phenomenology, while also offering conceptual resources for exploring its relation to biological autonomy and for understanding disturbances of self-experience in psychiatry.

13 **Sophie Taylor**
Association for Mathematical Consciousness Science, Germany.

Models of Computation and Models of Consciousness: Computational interpretations of thought

The most well-known model of computation is undoubtably the Turing Machine, and in computational models of consciousness and cognition, two others --- production systems and neural networks --- are also commonly found. However, these are but three models out of many.
Herein, we examine a number of less-well-known models of computation --- such as RNA computing, membrane computing, and graphical cellular automata --- from the perspective of their usefulness in mathematical consciousness science, and argue for their adoption by consciousness researchers.

14 **Francesco Giorlando**
Department of Psychiatry, Monash University, Australia

Distributed Consciousness Theory (DCT): From thermodynamics to phenomenology, a non-representational, physicalist theory of consciousness.

DCT is a new theory of consciousness, which primarily addresses the substrate problem: what type of physical substrate best supports the phenomenology of consciousness.
The theory describes mutual information non-locally 'distributed' in physical systems. This is formally described as closed systems having reduced degrees of freedom (entropy) due to mutual-information with the environment. This thermodynamic principle is generalised to a field with distance metric equivalent to the inverse of shared information.
The theory describes how conscious objects may be represented in this distributed field and details information flows associated with conscious perception as well as processes in the brain which may interact with distributed information.
The theory is able to solve key problems in consciousness theory, including The Binding Problem (Bayne 2010) and a novel solution to The Hard Problem (Chalmers 1995). Consequences and remaining unsolved problems of the theory will be discussed including the problem of causal emergence.

15 Hye Young Kim, **Adel Chaibi** & Eric Petit
École Normale Supérieur, France
Intel, France

Why Philosophy Matters for AI and Consciousness Research?

In an era shaped by rapid technological change, the development of artificial intelligence demands more than technical innovation—it requires philosophical reflection. Philosophy is not merely the study of historical theories; it is the foundational activity of questioning that structures how we perceive, interpret, and understand the world. From ancient metaphysics to modern physics, concepts like time, mind, and consciousness have evolved within philosophical frameworks that continue to shape scientific inquiry. Today, as we build increasingly complex models of cognition and artificial systems, our conceptual tools—terms like "consciousness," "self," and "intelligence"—remain rooted in paradigms inherited from Descartes, Kant, and others. But our technological capabilities have outpaced our understanding. We create before we comprehend. Philosophy trains us to ask the questions that open new paradigms: What is AI—not just functionally, but existentially? How does it reflect or transform our humanity? Projet Y explores these questions by integrating philosophical thinking into the discourse around AI and consciousness. We argue that there is no science without context, and no model without metaphysical grounding. To move beyond inherited assumptions and face unprecedented challenges, we must revive philosophy as a living, collective practice—one that guides understanding, not after the fact, but as its very condition.

16 **Patrick Grüneberg**
Institute of Liberal Arts and Sciences, Kanazawa University, Japan

Performative Intentionality as a Function of Phenomenal Consciousness

Phenomenal consciousness is often regarded as a qualitative, epiphenomenal overlay to otherwise subpersonal processes. In contrast, I argue that phenomenal experience plays a constitutive and functional role by generating a first-person perspective: a self–world distinction that enables the agent to enter into meaningful relations with its environment. Within this subjective space, phenomena are not merely present but present for the agent, thereby acquiring significance through self-relatedness.
I illustrate this point by examining self-initiated bodily movement. Drawing on the concept of performative intentionality—a mode of intentionality expressed through activity rather than directed toward an object—I argue for a heterarchical model in which conscious motor acts arise from dynamic self-relations among the agent, its neuromuscular system, and environment. Accordingly, action-related phenomenal consciousness enables goal-directed modulation of bodily states by being identical with the self-controlled initiation of sensorimotor mappings.
Thus, conscious movement initiation exemplifies a functionally irreducible role of phenomenal consciousness: the realization of meaningful, agent-centered control. This model challenges representational accounts of intentionality and supports the view that conscious acts can play an efficacious role in action control. While movement initiation serves as a paradigmatic case, the performative structure of intentionality extends to all cognitive processes.

17 **Adam Rostowski** & Simon Bowes
Sussex Centre for Consciousness Science, School of Engineering and Informatics, University of Sussex, United Kingdom

The View From "The Intentional Stance": mind-dust, fiction, or natural minds?

Dennett's intentional stance often leads down one of two paths. The first, to a counter-intuitive ubiquity of mental states in nature. The second, to the rejection of more-than-pragmatic attributions of such states in even the most intuitive cases. Our two options seem to be mental states all the way down (panpsychism), or metaphors all the way up (fictionalism).

We show this to follow from acceptance of the “common kind assumption”: the view that indistinguishable experiences, whether veridical or illusory, form a single kind.
Panpsychism follows from construing these common factors as intrinsic properties of conscious experience (qualia), over-extending Dennett's gradualism about cognition.
Fictionalism follows from the common kind assumption’s undermining realism about empirical phenomena, by taking veridical and non-veridical experiences to constitute a single kind. Consequently, mental state attributions are qualified by the connective “as if” to fend off accusations of “metaphysics”, obscuring the potential for underlying processes to justify realism about such attributions.
We show that treating veridical and non-veridical states as distinct kinds instead (disjunctivism), opens a path from the intentional stance to meaningfully distinguishing truly and only seemingly mental processes. We show how Dennett’s own illusionism points down this path.

18 Takashi Ikegami (invited talk)

Ikegami Laboratory, Department of General Systems Studies,
University of Tokyo, Japan

From Prediction Machine to Individual: Constructive Consciousness and Informational Synergy in the Android ALTER3

We propose a constructive approach to consciousness by experimentally probing the emergence of individuality in a humanoid android coupled with large language models (LLMs). Using the android ALTER3 as a physical substrate, we interpose asynchronous LLM -based agent networks between its sensorimotor loop, thereby enriching the dynamics beyond reactive prediction–action mappings. This architecture enables the appearance of coherent traits such as personality, thought, and situated responsiveness, transforming the android from a mere prediction machine and container into a life-like individual.
To characterize this emergent individuality, we apply Partial Information Decomposition (PID) to the interactions among sensory inputs, motor outputs, and internal LLM states. Whereas purely reactive systems are dominated by unique information from the sensors, the insertion of LLM networks yields a significant increase in synergistic information—that is, information about future actions that arises only from the joint contribution of perception and internal cognition. We argue that this synergy provides a principled measure of individuality: the extent to which an agent constitutes itself as a coherent unit distinct from its environment.
By grounding the study of consciousness in a constructive experiment with an embodied android, and by quantifying individuality through informational synergy, we aim to bridge theoretical models of consciousness with artificial life approaches. This perspective invites discussion on consciousness not only as prediction or integration, but as the emergence of coherent individuality at the boundary of mind, body, and environment.

References:

1. Otto E. Rossler, Lisa-Ruth Vial, Frank Kuske, August Nitschke, Takashi Ikegami, and Andrei Ujica, et al. (2019). Brain Equation and Personogenesis. Clinics in Pediatrics, 2:1011.
2. Masumori A., Maruyama N., Ikegami T. (2021). Personogenesis Through Imitating Human Behavior in a Humanoid Robot “Alter3”. Frontiers in Robotics and AI, 7:532375.
3. Ikegami T. (2024). Generative Ethics in ALIFE. MIT Press.(chapter 17)
4. Yoshida T., Baba S., Masumori A., Ikegami T. (2024). Minimal Self in a Humanoid Robot.
5. Ikegami T. Baba,S., Yoshida,T, Kojima,H.(2025). ILIAD: Towards Integrated Life, Intelligence, and Agency Design (preprint).
6. Yoshida T., Masumori A., Maruyama N., Ikegami T. (2025). From Text to Motion: Embodied Semantics in Humanoid Robot ALTER3. Frontiers in Robotics and AI, 12:1581110.
7. Maruyama N., Yoshida T., Sato H., Masumori A., johnsmith, Ikegami T. (2025). A Concurrent Modular Agent: Framework for Autonomous LLM Agents, (preprint).

19 Robert Prentner (lightning talk)

Institute of Humanities, ShanghaiTech University, China

Interface Consciousness

In this talk, I reconceive consciousness as the phenomenological bridge enacted through an agent's interface with the world. This approach integrates the interface theory of perception with mathematized phenomenology, yielding a model that is theoretically robust, empirically grounded, and computationally implementable.
Consciousness exists in the relations around us, not merely inside our heads. We conceptualize it as the fundamental aspect of reality with which an agent interfaces. To fully understand how subjective experience arises from the interaction with the world, we need to understand the spatio-temporal-causal structuring of the phenomenal.
This opens a 'cosmological' approach to the study of consciousness, moving beyond the confines of human-like subjectivity toward a more general account of how relational structures give rise to experiential forms.
This framework ultimately aims to inform new scientific modes of knowledge and technology.

20 Zoe Lee-Youngzie

Department of Psychology, University of California, Los Angeles, USA

A Category-Sheaf Compositional Model of Local-to-Global Phenomenal Experience

In this talk, we discuss a novel mathematical framework that formalizes the compositional relationship between the "broader" qualia of unified experience and its "narrower" components.
Recent work in the structural paradigm of phenomenal consciousness has shown promise in characterizing the phenomenal qualities of specific aspects of experience, such as color or sound, by empirically measuring or modeling their relative differences in a mathematical structure, often called a qualia structure. However, how a phenomenally unified experience can be decomposed into such local aspects of experience has been relatively unexplored, partially due to the lack of appropriate mathematical foundations.
To address this gap, we extend prior work that formalizes qualia structures as categories: we propose a mathematical framework based on category and sheaf theory in which phenomenal integration is modeled as a structural composition of qualia structures. This approach bridges local and global properties of experience, and provides new insight into the limitations of previous models that treat experience as consisting of either decomposable dimensions or indivisible wholes.
We explore an empirical application of this framework in a case of sensory-semantic decomposition in material perception. Model predictions are evaluated against observations in human psychophysics and artificial neural networks.

21 Etienne Jacques

Independent, USA.

Ineffable Machines: When Metaphysics Meets Hypercomputation

In this talk, we introduce a new kind of abstract machines as a possible solution to several metaphysical intuitions: idealism, process philosophy, relationism, and pattern philosophy.
There exists a strong correspondence between the physical world, the mental realm, and the mathematical domain. To attempt to explain this correspondence, we postulate that our

minds engage in unconscious processes at far higher speeds and finer scales than those accessible to conscious awareness.
We model these processes using elementary abstract machines whose state transitions are, by postulate, undetermined by any other machine. We call them Ineffable Machines.
By treating the collective dynamics of these machines as a formal metaphysical framework, we can then introduce more concrete concepts, thus providing a structured model for the experience of objective entities, whether abstract or concrete.
The first half of the talk will cover motivation and background. The second half will focus on the law of evolution governing machine states, which is most relevant to consciousness.

22 Gustaf Malmberg

The Finnish Quantum Institute, Department of Philosophy,
History and Art Studies, University of Helsinki, Finland

Is the phenomenal epiphenomenal? – the prospects of mental causation in Chalmers vs Bohm

The de Broglie-Bohm pilot wave interpretation of quantum theory assumes that an electron is a particle that is always guided by a wave or field, the latter of which is described by the wave function. Bohm and Hiley later reinterpreted the field to be a subtle energy which contains active information which means that the field is literally 'in-forming' the energy and movement of the particle rather than pushing and pulling it mechanically. Bohm proposed that such active information at the quantum level is a primitive mind-like quality of the elementary particles and developed a new theory of the relation between the mental and the physical, where active information, being simultaneously mental and physical, is the bridge between mind and matter. Similar to David Chalmers' Double Aspect theory of Information, Active Information differs in that it operates at the fundamental quantum level and therefore it could possibly avoid the charge of epiphenomenalism. To do that would require an account of how phenomenal properties relate to Bohm's Active information. This talk discusses the prospects of avoiding epiphenomenalism of phenomenal properties in the Bohmian scheme.

23 Ivan Serwano (lightning talk)

Independent, United Kingdom

Engineering Intentionality: Leveraging Quidditative States in Computational Models of Mind

This presentation introduces quidditative states, a novel construct proposed to replicate the functional role of qualitative states, such as subjective feelings and experiences, in computational models of the mind. While leading AI systems like LLM demonstrate multi-domain competence, they lack intrinsic motivation and contextual understanding, which highlights a gap between superficial generality and genuine cognitive intelligence. Quidditative states are defined as states within physical systems that possess defining and accessible attributes that have the properties of discernibility, differentiability, continuity, and variability. These properties are shared with phenomenal qualitative states but do not require subjective phenomenal form. This approach argues that these structural properties are the efficacious elements enabling intentionality and intrinsically motivated agency in biological cognition. By adopting quidditative states as structurally specifiable proxies for phenomenality, it becomes possible to engineer artificial agents capable of self-directed goal formation and adaptive understanding without relying on epiphenomenal assumptions. The research challenges the prevailing view that qualitative character is dispensable in modeling intelligence and offers a realist path toward strong AI. This paradigm will shift AI beyond resource-intensive scaling and reframes intelligence as grounded in intrinsic value formation. The concept of quidditative states, introduced here as an original contribution.

24 Kohtaro Tadaki

Department of Computer Science, Chubu University, Japan

An analysis of Wigner's friend in the framework of quantum mechanics based on the principle of typicality

The notion of probability plays a crucial role in quantum mechanics. It appears in quantum mechanics as the Born rule. In modern mathematics which describes quantum mechanics, however, probability theory means nothing other than measure theory, and therefore any operational characterization of the notion of probability is still missing in quantum mechanics. In our former works [K. Tadaki, arXiv:1804.10174], based on the toolkit of algorithmic randomness, we presented a refinement of the Born rule, called the principle of typicality, for specifying the property of results of measurements in an operational way. The Wigner's friend paradox is a Gedankenexperiment regarding when and where the reduction of the state vector occurs in a chain of the measurements by several observers where the state of the consciousness of each observer is measured by the subsequent observer. It is one of the central open questions in the measurement problem of quantum mechanics. In this talk, we make an analysis of the Wigner's friend paradox in the framework of quantum mechanics based on the principle of typicality. We then draw common sense conclusions about the Wigner's friend paradox. In addition, we make a prediction, which is testable in principle, about its variant proposed by Deutsch.

25 Ian Durham

Department of Physics, Saint Anselm College, USA

Conscious experience in time-reversed Doppelgängers: how a passage of time could make a difference to experience

Three questions about time have been widely debated: (1) Is a passage of time a fundamental physical property of our universe? (2) If it is, could that explain why we experience time as flowing? (3) If it is, could it make any difference at all to our experience? Many theorists answer no to all three questions. To (1), they argue that physics supports a block universe, containing no fundamental time passage. To (2) and (3), they argue that no clear account has been given of how such a thing could be relevant to experience, let alone our experience of time. Price (1996, 2007) discussed these issues in the context of the Gold universe, a symmetric universe, which begins with a Big Bang, followed by an entropy increase, a decrease, then a Big Crunch. According to Price, negative answers to (1)-(3) are supported by the contention that your ``time-reversed Doppelgänger'' on the other temporal end of the Gold universe will experience what you experience, and will experience time flowing into your past, making time's passage perspectival. Advocates of time's passage, such as Maudlin (2007) rejected Price's conclusions, arguing that there is no reason to assume the Doppelgänger would experience anything. In this paper, we show how question (3) can be answered positively, by a theory of consciousness, in a rigorous and quantitative way. We use the integrated information theory (IIT) and apply it to toy Gold universes. Our aim is not to argue for or against IIT but to use it as a proof-of-concept, demonstrating how a theory of consciousness, which incorporates a passage of time, could contradict Price's predictions, and be consistent with Maudlin's. In addition, we show that question (2) can be answered positively, without requiring that our experience of time flow directly represents the passage of time. (Joint work with Kelvin McQueen and Larissa Albantakis)

26 Paul Skokowski

Center for the Explanation of Consciousness,

Qualia, Quantum Mechanics and Neuroscience

One might be forgiven for wondering how quantum mechanics could be relevant to human sensation. After all, aren't humans, our sensory systems, and the objects we sense macroscopic objects? If so, classical physical theories should suffice. But it turns out that humans are already capable of sensing quantum mechanical objects. Further, the quantum mechanics of detection support an externalist approach to sensory experience and qualia. This paper presents an analysis of detection in the quantum mechanical formalism which supports the fine-grainedness of internal representational—Intentional—states, a feature of mental states that is also supported by results from neuroscience. That is, the quantum mechanical formulation of detection, which describes the sensory/perceptual experience of an external property, concurs with the vehicle, content, and causal role descriptions for these mental states. I will show that these components of detection are exemplified in the quantum mechanical formulation by an internal eigenstate (the vehicle), the eigenstate of the external property sensed (the content), and the deterministic Schrödinger equation (the causal role). This paper is based on work from the forthcoming book Sensing Qualia: Solving the Hard Problem of Consciousness, University of Chicago Press, January 2026.

27 Erik Nemeth (lightning talk)

ISIR, Sorbonne, France

Multi-Layer Awareness Architecture Exhibiting Curiosity and Hippocampal Replay

The existence of consciousness plays a significant role in the evolution of autonomy and decision-making, and identifying this role could allow us to reverse-engineer and analyze its underlying mechanisms and functions.
According to the Distributed Adaptive Control Theory of Consciousness (DAC-TOC) [1], this latter might have emerged due to the need of distinguishing the virtual "self" from the virtual "other" in an increasingly multi-agent world, allowing for modeling disparate agencies. The phenomenology of the resulting conscious experience is complex and multifaceted – thus instead of speaking of consciousness as a singular entity, we propose a multi-dimensional description of consciousness profiles [2]. One key axis of these profiles is awareness, defined as the content-related component of conscious experience. Within this framework, we propose the CAVAA architecture: a multi-layered hierarchical cognitive architecture theoretically capable of supporting awareness, consisting of a reactive-, adaptive-, contextual- and a virtualization layer [3].
In this talk I would like to present the CAVAA architecture through its basic components, with a special focus on my work on the virtualization layer [4]. This layer, focusing on the abstraction and modeling of relevant aspects of the external and internal environment, takes inspiration from the recent literature of hippocampal replay in spatial navigation [5-6] and the Active Inference principle [7], establishing a curiosity-driven informed exploration and replay via the virtualization of discounted epistemic rewards. In and of itself, this layer is capable of reproducing several key findings in the study of hippocampal replay more faithfully than previous proposals, supporting the idea that the virtualization of uncertainty might contribute to spatial navigation and learning. When integrated into the full architecture, we observe a targeted replay of relevant experiences, corresponding to our definition of aware learning (as opposed to an uninformed, unfiltered reinforcement learning), which I will demonstrate through an instantiation of the CAVAA architecture on a simulated robot.

[1] P. Verschure, "Synthetic consciousness: the distributed adaptive control perspective," Philosophical Transactions of the Royal Society B, vol. 371, no. 1701, pp. 20150448–20150448, Aug. 2016, doi: https://doi.org/10.1098/rstb.2015.0448.
[2] K. Evers et al., "Preliminaries to artificial consciousness: A multidimensional heuristic approach," Physics of Life Reviews, vol. 52, pp. 180–193, Mar. 2025, doi: https://doi.org/10.1016/j.plrev.2025.01.002.
[3] O. Guerrero-Rosado, I. Freire, E. Németh, A. Vinck, P. Verschure, M. Khamassi (in prep.). Mental time travel in robots.

[4] E. Németh, K. Jedlovszky, A. Chartouny, I. Freire & M. Khamassi (in prep.). The epistemic function of the hippocampus.
[5] M. G. Mattar and N. D. Daw, “Prioritized memory access explains planning and hippocampal replay,” Nature Neuroscience, vol. 21, no. 11, pp. 1609–1617, Oct. 2018, doi: https://doi.org/10.1038/s41593-018-0232-z.
[6] E. Massi et al., “Model-Based and Model-Free Replay Mechanisms for Reinforcement Learning in Neurorobotics,” Frontiers in Neurorobotics, vol. 16, Jun. 2022, doi: https://doi.org/10.3389/fnbot.2022.864380.
[7] T. Parr, G. Pezzulo, and K. J. Friston, Active Inference. MIT Press, 2022.

28 Fernando Rodriguez-Vergara

AI Research Group, Department of Informatics, University of Sussex, United Kingdom

Artificial consciousness and temporality: the phenomenological need for cognitive time

The way we experience the world has an inherently temporal aspect; events follow the ones before in a linear fashion. Our experience presents to us in the form of a fleeting present with no clearly demarcated beginning or end, always imbued within an evanescent stream of perceptions and thoughts. While different approaches from cognitive science and related fields share the view that this temporal aspect is fundamental to our phenomenology, our experience of time seems in direct tension with the physically grounded, widespread notion of discrete state-space transitions that underpin so much of modern cognitive science and artificial life. In other words, while state-space transitions seem to correctly characterize most cognitive phenomena, it isn’t clear how this relates to the fluid and evanescent temporality of our experience. We present a formal framework centered on the idea of how sensory-perception incompleteness translates into temporally dense constructions of the perceptual present.

29 Igor Balaz

Neovivum Technologies, Serbia

A Mathematical Framework for Hierarchical Concept Formation in Machine Consciousness

We present a formal mathematical framework for modeling how genuine understanding emerges from information processing through hierarchical concept formation. The model addresses how concepts emerge from experience rather than being pre-defined through symbolic assignment.
Our implementation employs Finite State Machines (FSMs) as mathematical representations of individual phenomena instances, organized into weighted conceptual graphs where nodes represent concepts and link strengths encode similarity measures. Complex phenomena are represented by Higher-Order Automata (HOAs) - hierarchically nested graphs containing FSMs and other HOAs as nodes, with links encoding spatial and temporal constraints. HOA learning employs breadth-first traversal with automata activation to decompose complex phenomena into previously learned components. When existing machines cannot explain new patterns, the system performs decomposition-based learning.
As agent encounters similar phenomena, the conceptual graph self-organizes into clusters of similar nodes, with each cluster forming robust concepts that generalize across variations. This addresses the symbol grounding problem by showing how meaning develops from experiential clustering rather than externally imposed symbolic assignment.
Our framework provides operational definitions for the transition from unconscious processing to conscious representation. Current implementation demonstrates concept

formation in geometric domains, with agents learning elementary shapes and their complex compositions through supervised and unsupervised learning.

30 **Ben Gaskin**
University of Sydney, Australia

Temporal continuity as an architectural constraint on machine consciousness

This paper traces the implications for machine consciousness which follow from the argument that consciousness is fundamentally characterised by a continuous rather than discrete temporality. To the extent that such continuity proves necessary (if not sufficient), then many prominent architectures in artificial intelligence are unlikely to be candidates for consciousness.
We thus introduce a three-tiered framework distinguishing between: 1) perceptual continuity, where observers project continuity onto discrete processes; 2) behavioural continuity, the coherent state-transition rules that support such projections; and 3) intrinsic continuity, where the substrate itself unfolds as a continuous temporal flow. Central here is the increment-rule distinction: in discrete systems, temporal progression (increment) remains separate from state-transition logic (rule); continuous systems unify these aspects.
Through a series of thought experiments, we show how this view forces a fork between facing the boundary problem and accepting constraints on viable architectures. We will then conclude by applying this framework to architectures in artificial intelligence: from artificial and spiking neural networks to neuromorphic computing and neural organoids.
This perspective provides conceptual equipment for evaluating temporality in artificial systems and makes specific predictions about which design approaches might exceed these limitations, thereby augmenting our theoretical basis for identifying consciousness in artificial systems.

31 **Tom Froese (invited talk)**
Embodied Cognitive Science Unit,
Okinawa Institute of Science and Technology, Japan

Re-Envisioning the Future of Human-AI Relations by Integrating Human Consciousness into the Natural Order

There is a rapidly growing awareness that our self-image as the sapient species is being challenged by the intensification of artificial intelligence. What do these surprising technological breakthroughs in AI imply for humanity's place in the natural order? Is there still something special about the human mind apart from intelligence, and if so, what is it exactly? Consciousness suggests itself as the last bastion resisting mechanization, but current theories are incapable of upholding the claim that human consciousness plays a fundamental role in behavior generation comparable to that of intelligence. Reaffirming humanity's unique role in the unfolding of reality will require a theory capable of meeting what I call the "Participation Criterion" of consciousness: a viable theory must explain how the presence of human consciousness makes a unique difference to the world compared to its absence. This talk introduces the required ontological concepts that enable a theory to meet the Participation Criterion, illustrates how these concepts productively integrate into the current scientific world image, and re-envisions humanity's future relation with AI accordingly as an intensified form of codependent arising.

32 **Sid Kouider** (showcase talk)
Brain & Consciousness Lab, Department of Cognitive Science, CNRS & École Normale Supérieur, France

The Pros and Cons of Global Workspace Theory

Global Workspace Theory (GWT) is arguably the most popular theory of consciousness, offering computational descriptions and testable predictions about the mechanisms underlying conscious access. Yet it is often criticized for having weak explanatory power: its functional stance is said to come at the cost of sacrificing phenomenology, reducing it to accompanying cognitive mechanisms (attention, global broadcasting, …) and neglecting the problem of how subjective experience arises in the first place. Additionally, GWT faces growing challenges from recent findings showing that neural markers once thought to reflect conscious access can be triggered unconsciously or explained as consequences rather than intrinsic properties of consciousness. In this talk, I will review these critiques and illustrate how GWT, with some flexibility and compromise, can still accommodate such findings. However, I will argue that the core methodological approach supporting GWT, namely the contrastive analysis, is reaching its limits, failing to deliver unequivocal evidence for the specificity of consciousness. I will argue that while GWT remains our most comprehensive scientific model of consciousness, it nonetheless leaves an explanatory gap. Finally, I will discuss how new approaches, particularly those based on foundational models and closed-loop neural interfaces, offer promising avenues for studying the specificity of consciousness.

33 **Chris Rourk**
Citizen Scientist, USA (formerly U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research)

Quantum consciousness - fact and fiction

The study of the neural correlates of consciousness (NCCs) has provided extensive evidence of the association of neural function and consciousness. Some proposed "quantum consciousness" theories such as Orchestrated Objective Reduction (Orch-OR) and Posner Qubits are at odds with NCCs, whereas one proposed quantum biological function (Catecholaminergic Neuron Electron Transport or CNET) is consistent with and complements NCCs. This talk will review the evidence in support of Orch-OR, Posner Qubits and CNET, and will further examine whether that evidence is consistent with or contradicts NCCs. In particular, Orch-OR relies on evidence of UV superradiance in microtubules that was obtained from tests that were performed in tubo and in the absence of microtubule associated proteins like ferritin, which quench microtubule fluorescence. Posner qubits lack any relevant experimental evidence. CNET is built on extensive evidence of 1) electron tunneling associated with ferritin, 2) the functional relationship between ferritin and neuromelanin in catecholaminergic neurons, and 3) the way that catecholaminergic neurons function, and has made numerous predictions that were subsequently demonstrated in reputable and repeated laboratory tests. Differences of opinion may exist about how to interpret this evidence, so a thorough understanding of the evidence is important for meaningful discussion of quantum consciousness.

34 **Joachim Keppler**
Department of Consciousness Research,
DIWISS Research Institute, Germany

An Evidence-Based Quantum Approach to Consciousness: The Brain's Dance with the Zero-Point Field

Empirical findings indicate that conscious states are inextricably linked with long-range synchronized activity patterns that result from phase transitions and exhibit the key features of self-organized criticality (SOC)., Crucially, the framework of quantum electrodynamics (QED) provides the appropriate methodological resources for explaining the origin of phase transitions and critical dynamics. An essential ingredient of QED is a fluctuating ocean of energy, the ubiquitous electromagnetic zero-point field (ZPF), consisting of a spectrum of normal modes. QED-based model calculations give rise to the conclusion that SOC originates from a bottom-up orchestration process governed by resonant brain-ZPF coupling. Thus, taking all available pieces of evidence into account, a profound new insight takes shape, namely, that the fundamental principle behind the formation of conscious states is the resonant coupling of the brain to the ZPF. This coupling causes an amplification of the dynamically relevant ZPF modes, suggesting that the necessary condition for the emergence of a conscious state is the selective excitation of ZPF modes. These insights support the hypothesis that the ZPF acts as a universal, inherently sentient but yet undifferentiated background field of consciousness, with differentiated conscious states arising through the selective excitation of phenomenal qualities immanent in this ubiquitous field.

35 **Anna Eiserbeck**
Cluster of Excellence Science of Intelligence, Department of Psychology, Humboldt-Universität zu Berlin, Germany

Attributing Consciousness to AI: Mechanisms, Effects, and Implications

While the factual status and potential for artificial consciousness remain unresolved, there is a growing number of cases in which artificial intelligence (AI) systems appear to exhibit behaviors typically associated with consciousness. This development increasingly blurs the boundaries between what is perceived as conscious living beings and non-conscious artificial systems. Consequently, questions concerning the attribution of consciousness—and the associated social and ethical implications—are becoming ever more relevant. Which factors shape our perception of consciousness? How does perceiving an entity as conscious influence interactions, judgments, and broader societal outcomes? This talk synthesizes empirical evidence from various fields to address these questions. First, mechanisms and factors underlying the attribution of consciousness are examined. Subsequently, the impact of these attributions is explored with regard to emotional perception, social interaction, evaluative judgments, and broader societal and ethical implications. Methodological challenges in studying perceptions of AI consciousness are also discussed. Finally, promising directions for future research are outlined. Overall, investigating perceptions of AI consciousness provides valuable insights into the broader understanding of consciousness itself.

36 **Francesco Lässig**
Independent, Switzerland

Unambiguous Representations: NCC Candidate and Gateway to Conscious Content

Impure representationalism (Chalmers, 2010) is one proposal for tackling the Hard Problem of consciousness: it identifies phenomenal properties with the right manner of representation. However, what counts as right remains underspecified, leaving a key gap in determining the conditions a representation must satisfy. I suggest that a necessary ingredient is unambiguity: ruling out competing interpretations. I model ambiguity as the entropy of interpretations given a representation and estimate it in artificial neural networks. Using a relational decoder that sees only similarity patterns among shuffled neurons in

unseen MNIST classifiers, I test whether these patterns alone reveal what neurons are about.
Main findings:
1. Relational structure generalizes across networks. The decoder recovers what neurons represent (e.g. class identity or pixel location) in new models purely from inter-neuron relations.
2. Training choices modulate ambiguity. Networks with nearly identical input-output accuracy differ sharply in estimated ambiguity, indicating that ambiguity is largely independent of task-level performance.
By turning ambiguity into a measurable quantity, the relational decoder paradigm not only operationalizes the “right manner” clause of impure representationalism but, whereas most current theories focus on quantifying the level of consciousness (e.g. IIT’s Φ), it also offers a computational handle on the contents of consciousness.

37 Dezhi Luo

Weinberg Institute for Cognitive Science, University of Michigan, USA

Understanding Magic: The Missing Ontology in Computational Functionalist Theories of Consciousness

Prominent scientific theories of consciousness are often computational functionalist in nature. That is, they posit that certain forms of information processing are responsible for transforming unconscious mental states into conscious ones. While such theories account for many empirical features that distinguish conscious from unconscious states, they are often criticized for appealing to a kind of “magic”: it remains unclear how a proposed computation—or “difference-maker”—could confer subjective character on selected states. I argue that this tension can be resolved by applying a specific ontological framework to standard computational functionalist theories—namely, a particular interpretation of what has been called virtualism. According to this framework, mental states are components of a simulated entity realized through computation. An entity, on this view, is conscious precisely in virtue of discriminating certain mental states from others based on their functional role to the entity. I first examine the mechanistic underpinnings of this proposal and then show how it accounts for key features of subjective character, thereby demystifying computational functionalist theories.

38 Nathaniel Virgo

Physics, Astronomy and Mathematics, University of Hertfordshire, United Kingdom

Interpreting dynamical systems as agents, and where to draw the boundary

How does the mind relate to the physical world? Dennett's answer to this question was that mental phenomena are part of a ""stance"" one uses to describe the physical world. This doesn't make them meaningless or non-physical, since not all systems' behaviour can be usefully described in intentional terms.
Over the past few years I've been working with colleagues to formalise a version of this idea, via the notion of *interpretation map*, a mathematical function that takes the state of a systems and ascribes a semantic meaning to it in regards to an interpretation of the system as an agent. Key results include an explanation of why beliefs and (Bayesian) belief updating seem to arise inevitably as soon as one tries to interpret a system as ""trying"" to perform a task.
In this talk I will explain one version of the framework and discuss the question of *which* system to interpret intentionally, i.e. where we should draw the boundary. This question is closely related to ideas about the extended mind. Our framework answers it in an agnostic way: different boundaries give rise to different interpretations, but they are not unrelated, and studying their relationships can yield insight.

39 Susan Schneider (invited talk)

Center for the Future of Mind, AI & Society, Florida Atlantic University, USA

From Circuits to Sentience: Why Digital Chatbots Aren't Conscious but Biological AIs May Be

Are chatbots AI zombies or conscious minds? I suspect they are AI zombies, at least when implemented on today's standard hardware. I argue that today's LLM's consciousness-like behaviors do not suggest they are conscious, because there is an error theory—a theory explaining why they behave as if they are conscious in absence of actual felt experience. Using circuit tracing work at Anthropic, I explain that LLMs function as "crowdsourced neocortices" — as they scale up they come to mirror human conceptual structures, leading to human-like behaviors, including those involving consciousness. By contrast, bio-computing and neuromorphic systems present more serious candidates for AI consciousness, being 'grey zone systems'. I evaluate the limits of functional consciousness (FC) as a marker, outline the ACT and modified ACT tests, and argue for a case-by-case approach to AI consciousness assessment. The upshot: functional consciousness and phenomenal consciousness must not be conflated, lest we mistake AI zombies for conscious beings, with profound moral and policy consequences.
Related papers:
https://philarchive.org/rec/SCHTET-14
https://philpapers.org/rec/SCHIAC-22

40 Peter Bruza

Queensland University of Technology, Australia

A very brief introduction to Quantum Cognition

Quantum Cognition is a field that employs the conceptual framework and formalism of quantum physics to provide new models and understandings of human cognition. This presentation aims to provide an overview of the key concepts of quantum cognition: indeterminacy, incompatibility and contextuality. These concepts will be used to address the question that if cognitive phenomena are meaningfully quantum-like, what might that imply about the underlying nature of the underlying cognitive reality. In particular, the computational and phenomenological views of the human mind will be compared and contrasted.

41 Michael Schnabel

Department of Political Science, Vanderbilt University, USA

The Universal Kōan: Quantum Theory and the Three Natures of Mind

Quantum cognition offers a novel framework for modeling decision-making and judgment under uncertainty. It employs Hilbert space representations and quantum probability to account for contextuality, ambiguity, and order effects—features relevant to human cognition and language. Formally, this shift reflects a move from classical to quantum probability and raises a deeper question: what might quantum cognition reveal about the nature of mind?

Buddhism has long engaged questions about the nature of mind. The Mahāyāna tradition, particularly the prajñāpāramitā texts and the Chan/Zen school, embraces paradox as a method for dissolving attachment to fixed conceptual frameworks, thereby clearing the way for a direct realization of mind. Intriguing parallels have been suggested between Buddhist dialectical methods and the structure of quantum logic. How does the paradoxical language of Mahāyāna relate to the "weirdness" of quantum mechanics and the difficulty of arriving at a satisfying interpretation?

The Yogācāra (Mind-Only) school offers a consciousness-based analysis of experience structured around the three natures (trisvabhāva): the imagined, the dependent, and the perfected. In this talk, I explore how these three aspects of mind relate to the two truths of Madhyamaka, and where quantum cognition and quantum theory might intersect with this layered understanding of consciousness and reality.

42 **Yuko Ishihara**
College of Arts and Letters, Ritsumeikan University, Japan

Nishida's Reversal of Kant: Quantum mechanics and the structures of consciousness

Traditionally known for his spiritual and Zen-influenced thought, Nishida Kitarō—the founder of Kyoto School philosophy—also engaged deeply with scientific developments, particularly during the final decade of his life. Drawing on his writings on physics and mathematics, this talk examines Nishida's efforts to incorporate insights from quantum mechanics into his broader epistemological and ontological framework. For Nishida, modern quantum physics marks a departure from classical objectivist logic, moving toward a mode of understanding rooted in embodied experience and the dynamic interplay between self and world. He interprets this shift as pointing toward a "world that envelops the self," thereby challenging the rigid subject–object dualism that characterizes much of Western philosophical tradition. This reorientation parallels Nishida's philosophical ambition to reverse Kant's Copernican turn. Rather than grounding knowledge in the subjective structures of cognition, Nishida conceives of consciousness as a self-enfolding activity that "mirrors itself within itself." While it cannot itself be objectified and hence is indeterminate (it is "no-thing"), consciousness serves as the condition of possibility for all determination and knowledge. In this way, Nishida offers an alternative to the familiar dichotomy between the mind as either a passive recipient or an autonomous agent.

43 **Shigeru Taguchi (invited talk)**
Center for Human Nature, Artificial Intelligence, and Neuroscience, Faculty

A Monoid Model of Consciousness: How Category Theory Can Help Us Understand the Unity of Consciousness

We undeniably undergo what is referred to as consciousness, yet it cannot be grasped as a substantial object of perception. Its mode of being is phenomenologically distinct from that of the objects that appear within consciousness. In this talk, I propose a theoretical approach to this distinctive ontological status of consciousness by drawing on the concept of loops, particularly through the lens of monoids in category theory. Specifically, I explore the idea that consciousness emerges when multiple sensorimotor loops—dynamic patterns of interaction between body and world, involving perception and action—form a structure of potential interconnections that instantiate the structure of a monoid. Basic sensorimotor loops and predictive processes, taken alone, are not sufficient to account for consciousness, as they typically operate below the threshold of awareness. I argue that consciousness arises precisely when the potential connectivity among these loops is realized—when a coherent structure emerges in which both the diversity of conscious contents and the unity of consciousness are simultaneously maintained. I suggest that this latent structural organization can be effectively modeled as a monoid in category theory.

44 Miguel Aguilera (invited talk)

Basque Center for Applied Mathematics, Spain.

Nonequilibrium Neural Computation: Critical Integration, Binding, and Sequential Memory

Effective neural information processing requires flexible architectures capable of integrating multiple sensory streams that vary over time with internal and external events. From a thermodynamic perspective, neural computation is an out-of-equilibrium, non-stationary process that evolves dynamically and gives rise to entropy production. At the cognitive level, nonequilibrium neural activity leads to ongoing changes in both sensory inputs and internal states. In contrast, classical theoretical neuroscience has primarily relied on stationary, equilibrium information paradigms (e.g., efficient coding theory), which often fall short in capturing the role of nonequilibrium fluctuations in neural dynamics. In this talk, we analyze how the nonequilibrium dynamics of associative memory networks (aka Hopfield networks) can support different kinds of neural computation in such out-of-equilibrium conditions. This approach aims more broadly to characterize the thermodynamic and computational organization of flexible, large-scale neural processes. First, I will recapitulate previous results showing how dynamical mean-field theories for associative memories reveals that integrated information is enhanced near dynamical critical points, offering insight into predictions from Integrated Information Theory. Second, we explore how large-scale nonequilibrium neural networks can support metastable, sequential activity. We present an extension of Hopfield networks with asymmetric couplings that generate structured, cyclic dynamics beyond classical fixed-point attractors. These cycles can store and replay co-occurring temporal patterns, generalizing classical memory capacity calculations to include metastable, cyclic structures. This provides a statistical physics framework for understanding structured sequences and temporal integration, and could provide a computational basis for phenomena like binding by synchrony.

45 Valtteri Arstila (lightning talk)

Department of Philosophy, University of Turku, Finland

The Dispensability of the Specious Present in Theories of Consciousness

Many contemporary theories of consciousness omit any substantive account of the specious present—the purported temporal span of immediate experience. This omission could be seen as a deficiency. However, I argue it is neither accidental nor problematic. All phenomena typically cited as requiring a specious present, such as motion perception or temporal binding, can be fully explained by unconscious processing and discrete mechanisms of content integration. Consequently, no explanatory gap arises from leaving the specious present aside. This perspective invites reconsideration of whether temporally extended experiential structures are theoretically indispensable or merely conceptual artifacts.

46 Chetan Prakash

Department of Mathematics, California State University, USA

Observer and Observed in Conscious Participator Dynamics

In a participatory universe, with inherently entangled interacting systems, what is observation? How are observer and observed related? We propose answers to these questions within an approach that takes consciousness as fundamental. We model the process of consciousness experiencing itself as a network of abilities called “conscious

agents," each with sets of experiences, decisions and actions. Relationships between these sets are stochastic, giving rise to a Markov process on the experience interface of a given conscious agent. This process then "traces" onto any given subset of the experience interface as a new Markov process of experiences restricted to that subset. An "observer," then, is such a traced process, and its observations are characteristics of the traced process (e.g., asymptotics). The "observed" is the corresponding process on any larger set of experiences that contains those of the observer. This leads to a natural logic of observation, the "trace logic," describing the relationships between different traced observers. Of particular interest is when observers can "join" coherently, so as to make new observers. Moreover, there seem to be some deep connections to physics: we propose an origin, as well as possible generalizations, for the results of relativity, quantum theory and particle physics.

47 **Renaud Gauthier**
Department of Mathematics, University of Mary, USA

Homotopy Wave Function - a higher quantum treatment of consciousness

In a quantum formalism, consciousness can be modeled by a measurement process wherein the observer apprehends the realm of phenomena. The homotopy wave function in Algebraic Geometry provides a description of what it is one can observe. Since this is couched in the language of Segal topoi, which provide a setting for describing natural phenomena, it follows one obtains a multi-layered structure of reality, and consequently consciousness is necessarily parametrized by one's entanglement with various strata of such a structure.

48 **Yuria Shimizu** (lightning talk)
Graduate School of Arts and Sciences, The University of Tokyo, Japan

How the Body tells the Stories: A Predictive Processing Model of Embodied Self Narratives across Lifespan

We constantly tell stories about our past, present, and future selves. Previous work has shown that bodily and narrative aspects of the self work in tandem to build a coherent sense of self through time. Our self narratives are hypothesized to emerge from early sensorimotor experiences, but the mechanism remains an open question.
Here, we propose a theoretical model describing a two-way interaction between high-level, abstract self-representation and low-level bodily sensorimotor processing in continuously generating embodied narratives across the lifespan. Building on previous work on reconstructive recall, our model treats memory not as static retrieval from a passive 'container', but as the generative simulation grounded on past bodily experiences.
The body tells our life stories and stores our memories way before one is able to explicitly retrieve them cognitively at the high level. This information flow is bidirectional and continuous: top-down narratives about one's experiences actively modulate our bodily sensorimotor engagement with the world. In turn, low-level embodied stored narratives from early life will impact our conscious behaviour.
Our model will provide the integrative picture of how our self narratives are told not only through our high-level cognitive stories but also through our bodies.

49 **Anya Daly**
The EthicsLAB, Department of Philosophy,
University of Tasmania, Australia

Embodied Consciousness in Virtuality: The phenomenology of virtual identification and agency.

Virtuality, often described as the media of disembodiment wherein mind is conceived as pure data and body as mere vehicle, depends on a tacit mind-body dualism. Such accounts typically seek to sever connections between mind and body to uphold the mind's superiority and absolute autonomy, as if 'the ghost in the machine' can abandon the body and take up residence in a technological-machine-world-space.
This view appeals to cultural theorists who propose that virtuality provides a means of transcending gender, race, and disability through a transcendence of the body. Proponents of posthumanism pursue a similar direction in recognizing the potential of virtuality to challenge the division between the 'real' and the 'represented', thereby reconfiguring identities beyond the physical.
The above accounts' promotion of 'disembodied' agency is problematic. So-called 'disembodied' virtual agency depends on embodiment and the body is already a virtual body. Self-identification with virtual bodies through our capacities of visuo-tactile perception and cross modal congruency effects are exploited in current applications. How this is possible requires elucidation. Through phenomenological analyses of perception, body schema and body image this paper sheds light on how the sense of ownership and agency is established in the proprioceptive shift from actual to virtual body?

50 **Elay Shech**
Department of Philosophy, Auburn University, USA

From Numerical to Linguistic Representation: On the Possibility of AI Consciousness

As LLMs, and so-called Language Agents (LAs), increasingly rely on natural language to guide learning, inference, and interaction, a "linguistic turn" is reshaping machine learning. This presentation explores the philosophical implications of this shift for debates about AI consciousness. Traditionally, machine learning systems have operated through opaque numerical mechanisms—including gradient descent, vector embeddings, and tensor operations—primarily optimized for predictive accuracy. In contrast, recent developments (e.g., Chain-of-Thought, Reflexion, ReAct) center natural language as the medium of reasoning, feedback, and coordination. This shift brings such systems closer to a representational format many philosophers associate with consciousness: linguistically and propositionally structured content. If a linguistic turn is indeed linguistic in the relevant cognitive sense, it may mark a significant step toward artificial consciousness. For instance, some scholars have recently suggested that slightly modified language Agents might satisfy conditions for consciousness placed by Global Workspace Theory. We assess this possibility by distinguishing superficial language mimicry from genuinely structured linguistic cognition, examining whether these systems satisfy constraints like self-reference, memory integrity, and reflective coherence. At the same time, we caution that optimistic descriptions of LLMs and LAs often use cognitive vocabulary fairly liberally, thereby smuggling cognitive significance rather than arguing for it.

51 **Jad Tarifi** (lightning talk)
Integral AI, USA

Consciousness in Artificial General Intelligence

We propose that Artificial General Intelligence (AGI) equipped with an integrated, dynamic world model—essential for flexible general reasoning—necessarily exhibits consciousness in a fundamental sense. Our argument rests on two interconnected principles. First, we hold that proto-consciousness constitutes a basic feature of physical reality, not merely an emergent property of biological neural networks. This position, compatible with monist,

panpsychist, and certain theist frameworks, circumvents the ""hard problem"" of consciousness by grounding subjectivity in the fundamental architecture of reality itself. Second, we contend that world models in AGI function to create an ""inner theater""—a unified simulation space where perception, memory, imagination, and planning integrate. This space operates as the functional equivalent of first-person experience, serving as the medium through which AGI comprehends its own position within a modeled reality.
Under this framework, AGI systems with self-reflective and coherent world models will manifest a structured first-person perspective—their own experiential standpoint anchored in sensorimotor simulation, memory, and attentional processes. This theoretical foundation reframes questions surrounding synthetic pain and reward, emergent machine emotions, and establishes principled grounds for evaluating the moral standing of open-ended artificial general intelligence.

52 **Giulio Ruffini**
Neuroelectrics, Spain

The Algorithmic Weltanschauung

Kolmogorov Theory (KT) is an algorithmic-information framework seeking unified insights into complexity, biology, intelligence, and subjective experience. KT adopts a stance akin to monistic idealism or panpsychism, positing experience as fundamental, and investigates how structured experience arises from mathematical structures. Reality, in this view, is a computational construct driven by algorithmic compression. We discuss how persistent algorithmic patterns emerging within an “algorithmic soup” give rise to agents—entities optimized for survival by capturing structured information. Agents propagate their informational identity across generations by steering action to maximize future self-information (telehomeostasis). Essential components of agents include an information membrane, a modeling engine, an objective function, and a planning module. Driven by survival imperatives, agents rely on lossy compression via coarse-graining (emergence of simplicity) and valence (goals) as fundamental cognitive mechanisms. Structured experience and subjective reality naturally emerge from these computational processes, providing further links with foundational physics themes such as symmetry. We further explore core KT concepts—models, life, intelligence, emotion, emergence, and multi-agent systems—in an interdisciplinary context, highlighting how algorithmic complexity and computational modeling illuminate fundamental questions concerning nature, mind, and reality.

53 **Francesca Castaldo**
Neuroelectrics, Spain

Algorithmic Psychodynamics: An Algorithmic-Agent Framework for Neuropsychiatry

Algorithmic Psychodynamics introduces a novel computational framework grounded in Algorithmic Information Theory (Kolmogorov Theory) and the Free Energy Principle, aiming to unify first-person subjective experiences and third-person neuroscientific observations in the study of neuropsychiatric disorders, such as Major Depressive Disorder. Within this paradigm, emotional states and cognitive dysfunctions emerge from the interplay of three fundamental computational modules—namely, the objective function (evaluating valence), the modeling engine (predicting internal and external states), and the planning engine (guiding goal-oriented behaviors)—within algorithmic agents. First-person aspects such as structured experience, consciousness, and emotions (or valence) are systematically linked to the agent’s algorithmic and dynamical properties. Pathologies are then understood as disruptions within these computational processes: aberrant valence evaluation within the objective function, inaccuracies in predictive internal models, or impaired action planning.

By explicitly mapping these computational elements onto neural substrates and circuitry (e.g., cortico-striatal-thalamic loops, default mode, and salience networks, neuromodulatory systems), Algorithmic Psychodynamics provides a rigorous, mechanistic account of neuropsychiatric conditions from a unified algorithmic-dynamical perspective. Consequently, therapeutic interventions—including psychotherapy, brain stimulation, and pharmacology—are conceptualized as targeted modifications of the agent's computational structure and dynamical state.

54 **Yuko Ishihara (invited talk)**

College of Global Liberal Arts, Ritsumeikan University, Japan

Mirroring Consciousness: Nishida's Mirror-Imagery and Self-Consciousness

The metaphor of the mirror has long been used to explain the workings of the mind and consciousness, often portraying it as passively reflecting reality. Richard Rorty's Philosophy and the Mirror of Nature (1979) famously critiqued this representationalist image, dismissing the mirror-imagery as a misleading model of cognition. For Rorty and other anti-representationalists, the mind is not a mirror of the world but is active, embodied, and inseparable from it.

Yet the mirror image is also central in non-Western traditions, such as Indian and Buddhist thought, where it does not necessarily entail representationalism. This talk examines the modern Japanese philosopher, Nishida Kitarō's use of mirror-imagery, likely inspired by Zen Buddhism, as a way of rethinking the metaphor in a non-representational framework.

Nishida's mirror-imagery features in his famous formulation of jikaku or self-consciousness: "the self mirrors itself within itself." Consciousness, or mind, is here taken to be the "basho" or place that mirrors itself within itself. Unlike an ordinary mirror, however, consciousness is not an entity but is a "mirror of nothingness" that reflects reality without itself being an entity. Unlike the representational model, this mirror does not posit a mind-world dualism; rather, it is the field (basho) in which both self and reality are disclosed.

How, then, does consciousness mirror reality without representing it? What role does the mirror-imagery play in Nishida's anti-representational account of mind? By pursuing these questions, I situate Nishida's account of self-consciousness in dialogue with contemporary debates on mind, self-consciousness, and non-representational models of cognition.

55 **Moritz Kriegleder** (showcase talk)

Vienna Cognitive Science Hub, Department of Philosophy,
University of Vienna, Austria
Association for Mathematical Consciousness Science, Germany

Perspectival Realism in Consciousness Science

For the last three decades, the search for neural correlates and the push for unifying theories of consciousness has dominated consciousness science. However, localizing consciousness and comparing theories has proven to be much more complex than selecting the right one, as demonstrated by adversarial collaborations such as Cogitate and Intrepid. Instead of fully fledged theories, we have more ambiguous theoretical frameworks that can lead to different empirical predictions. In our talk, we argue that a convergence of theories is not the only way to make progress in consciousness science. Perspectival realism is a philosophical framework that accommodates different types of explanations and evidence to identify robust phenomena. For example, climate science uses incompatible models at different levels of description to derive testable predictions about the impact of global warming. Similarly, consciousness science could make progress by integrating different theoretical perspectives into perspectival models. A perspectivist approach addresses the question of whether illusionism is correct by assuming that there is no robust phenomenon, there is a unified phenomenon that could be explained by converging theories, or there are multiple robust phenomena that necessitate multiple levels of description.

56 **Yifeng Chen**
META, Department of Computer Science, Peking University, China

Consciousness as entropy reduction (I)

Though familiar to all, `consciousness' lacks a definition. One difficulty is the variety of interpretations across various disciplines with contrasting approaches. Another is the need for a third-person account of subjective experience. Furthermore the quest includes the meaning of consciousness for entities other than humans, including AIs; and the notion has not been clarified by thoughts about related concepts like self awareness, empathy and mental capacity. In this and the following talk, a model of consciousness is considered which, having a logical basis, lends itself to simulation using a simple mathematical model called Consciousness as Entropy Reduction (CER). The work owes much to GWT, IIT and the earlier less mainstream `feature-map model' in Psychology. CER considers the contents of consciousness and subconsciousness as `scenarios': vectors of patterns (or features) on various channels (or feature locations). A feature map is one subconscious input scenario amongst many from which the conscious experience is chosen. The result is an internal simulation of the outside world, as in `predictive processing' from Neuroscience. Solving problems internally by simulation is more efficient than experimenting in the actual environment. Conscious experience in such a form is not only efficient but amenable to adaptability, a major evolutionary advantage.

57 **Jeff Sanders**
DISEI, East China Normal University, China

Consciousness as Entropy Reduction (II)

In Consciousness as Entropy Reduction (CER) subconscious content is modelled as a probability distribution of features over channels. One is chosen for consciousness by gradient descent of the entropy of that distribution until a zero-entropy deterministic distribution is reached. CER offers the advantage of simulation.
The brain is seen as an entropy-reducing mechanism that keeps `squeezing' the subconscious content to form scenarios of conscious experience. If conscious experience is projected back to the subconscious, the mechanism may produce consequent experiences from previous experiences, giving rise to `high-order' thoughts---a characteristic of more evolved forms of consciousness. CER does not distinguish the objective from the subjective. Subjective concepts such as self, self awareness, self consciousness and subjective experiences become internally generated abstract patterns appearing in conscious scenarios.
Gradient descent, which reduces entropy in CER, resembles that of a Hopfield neural network, which changes state by energy reduction until it reaches an energy-minimal stable state. In CER the entropy of a subconscious distribution may be considered `energy' and its stable states the extreme distributions with zero entropy.
We analyse: the experiment of binocular rivalry from human vision; Libet's experiment from Neuroscience; thinking fast and slow; and the formation of higher-order consciousness.

58 **Masanao Ozawa**
Nagoya University/Chubu University/Ritsumeikan University/RIKEN,

Quantum instrument approach to question order bias in survey research

The ""question order effect"" in surveys refers to changes in response caused by altering

the order of questions. Understanding this effect is crucial for ensuring the reliability of survey but is difficult to explain using conventional methods. Quantum cognition models have attempted to account for this phenomenon using projective measurement in quantum theory. However, these models fail to reconcile the question order effect with the ""response replicability effect""—the tendency for individuals to give consistent answers to repeated questions.
To overcome this limitation, we introduced ""quantum instrument models"", the most general mathematical models for quantum measurement. This approach successfully shows both the question order and response replicability effects consistently in a model. Our model also reproduces the well-known Clinton-Gore survey results and eliminates the question order bias from the real data.
We found that the bias elimination achieved through quantum instruments is fundamentally different from that obtained via the conventional randomization method. In this talk, we demonstrate that while our quantum instrument model completely eliminates the order bias, randomization does not eliminate the bias entirely but merely reduces it to at most half of its original magnitude, and we further clarify the classical, quantum, quantum-like natures in quantum instrument models.

59 **Sean Tull**
Quantinuum, United Kingdom

A Compositional Analysis of Quantum Cognition

In joint work with Masanao Ozawa, we will present a compositional framework for models of cognition and decision making, and apply it to assess the use of quantum theory in cognitive science. The approach models decisions as instruments in process theories, formalised in terms of category theory, and makes it easy to study, compare and generalise both classical and quantum models of cognition.
We then apply this perspective to analyse the arguments for quantum cognition. We will see that many effects attributed to quantum theory can be captured classically, once one considers a slightly more general form of (deterministic) classical model than typical Bayesian models, and we even prove that every decision dataset has such a classical model. We discuss further constraints that could point to quantum models (‘naturalness’, fewer parameters, a ‘logical’ structure…), arguing that the strongest case may come from compositional models involving multiple domains, as in Bell scenarios, making use of the quantum tensor product."

60 **Manuel Baltieri** (showcase talk)
Research & Development Department, Araya Inc., Japan

What is it like to be a Braitenberg vehicle?

Braitenberg vehicles have been used in the embodied cognitive science literature as one of the main examples of systems that can solve seemingly complex tasks without a complicated model of the world. In robotics and biology they are also a rich source of inspiration for simple circuits in robots and living systems that solve tasks involving some kind of taxis, i.e. movement in response to some stimulus. These humble vehicles are however rarely studied in their own right, and while their particular behaviours have been explored extensively, especially in simulation, their ability to capture, at least implicitly, particular features of their environment, how this can be characterised and what it entails, is somewhat overlooked. Here, we make Braitenberg vehicles the focus of our work, and take some initial steps in a project aiming to describe the world as subjectively "experienced" by a Braitenberg vehicle. We characterise their environment and Umwelt, aiming to describe what these humble vehicles can or cannot disambiguate while interacting with the world. We then add restrictions on these vehicles that affect their Umwelt, and show how these influence their behaviour in silico.

61 **Nicolas Loerbroks**
Ruhr University Bochum, Germany

Does consciousness research presuppose the unity of the sciences?

Descartes introduced both the mind-body problem and the reductionist mechanistic worldview. Although the latter has been questioned in the 20th century, for instance by Feyerabend, Putnam, and Fodor, it still rules in science and philosophy today. In this talk, I will demonstrate that the mind-body problem as an explanandum stems from the mechanistic worldview. In particular, Chalmers formulated the "hard problem" in opposition to the "easy problems", which are subject to mechanistic explanations. In this light, Dennett's antithesis can be construed as stemming from the methodological pluralism present throughout the cognitive sciences. Because psychological "natural kinds" do not correspond to neuroscientific "natural kinds", it feels like there is something more to explain. If cognitive science was to progress such that they do correspond, the "hard problem" would be dissolved; consciousness science becomes part of cognitive science. But what if we were to assume more than methodological pluralism? Would there still be a single identifiable explanandum? Does the notion of a scientific theory of consciousness depend on the same doctrine that declared consciousness to be outside the scope of scientific investigation? Furthermore, Putnam introduced computational functionalism as a non-reductive hypothesis. What does this mean for "AI-consciousness"?

62 **Nezihe Müge Kuyumcuoğlu Tütüncüoğlu**
& Sinem Elkatip Hatipoğlu
Department of Philosophy, Marmara University, Turkey

Beyond states: a dynamic model of higher-order theory of consciousness

Higher-order theories (HO) claim that a mental state is conscious when it's represented by another mental state but the specifics of the representing relation between the target and the HO state are not well-articulated resulting in the strange possibility of a HO state representing the subject to be in a non-existent target state. Such problems emerge from taking discrete HO states to be sufficient for consciousness while remaining ambiguous about how they come about.
We want to propose a relational version of HO theory, grounded in the framework of implicit definition and recursive interaction. According to interactivism (Bickhard 2009; 2025), the type of representation that leads to consciousness consists in the dynamic interactions of an agent with the world. Dynamicity is captured via the mathematical notion of implicit definition (Hilbert 1927): we model mental content not as explicitly encoded representations, but as a network of agent-internal constraints that differentiate feasible interactions from unfeasible ones. Consciousness emerges when a first-level flow of anticipation is modulated by a higher-order flow, both of which are implicitly defined. This higher-order relational structure constitutes an interiorized action/intervention process, instead of a static higher-order state.

63 **Pawel Motyka**
Virtual Reality and Psychophysiology Lab, Institute of Psychology, Polish Academy of Sciences, Poland

Phenomenologically Distinct and Intensity-Varied VR Hallucinations: A Platform for Simulating and Studying Altered Conscious Experience

Exposure to unusual sensory conditions can drive broad-scale changes in subjective experience. Although this has been studied mainly with intense rhythmic or unstructured visual stimulation, advances in machine learning and immersive technologies allow the creation of highly complex, naturalistic alterations in visual phenomenology. We present new developments of the Hallucination Machine featuring (a) stereoscopic stimulation, (b)

hallucinations varied in strength and qualitative type (Deep-Dream vs. Style-Transfer), and (c) real-time, gaze-contingent propagation of hallucinatory content. Across a series of exploratory VR studies, we assessed the effects of these manipulations on subjective experience and peripheral physiology, including cardiac and oculomotor activity. As the primary focus, we assessed experiential changes using a questionnaire targeting high-level, non-sensory experience, complemented by free-form descriptions analysed using Large Language Models and Topic Modelling. We observed distinctive, intensity-scaled effects of visual alterations, with a modest advantage for psychedelic-like Deep-Dream hallucinations over artistic-style Style-Transfer transformations in eliciting reports of altered experience. These changes were not matched by increases in heart rate entropy – a proposed physiological marker of psychedelic states. By combining immersive, parametrically controlled hallucinations with NLP-based analysis of experience, this work contributes to the study of non-pharmacologically induced altered states and highlights new directions for computational phenomenology.

64 **Trevor Hewitt**
Sussex Centre for Consciousnes Science, School of Engineering and Informatics, University of Sussex, United Kingdom

Trained participants recreate images of geometric visual hallucinations induced by stroboscopic light

A better understanding of visual hallucinations could shed light on how the brain constructs conscious experience from sensory input, but their phenomenology remains difficult to characterise. Stroboscopic light rapidly induces simple visual hallucinations of geometric patterns that resemble those reported in clinical or psychedelic contexts, making it a useful model for studying hallucinations. In two experiments, over 100 participants recreated geometric visuals from hallucinations and veridical perception. Participants viewed 11, 40, and 80 Hz strobe light and recreated visuals using freehand drawing (Experiment 1) or a custom interface (Experiment 2), with both experiments yielding a similar pattern of results. Mathematical analysis revealed systematic differences in hallucinated geometric forms across strobe frequencies. These findings support the theorised relationship between strobe light frequency and the phenomenology of induced hallucinations, while challenging existing theoretical models of their neural mechanisms. Methods were validated with trials where participants recreated veridical stimuli instead of hallucinations. Image recreation is shown to be a viable tool for quantitative and mathematical investigations into the phenomenology of visual hallucinations. This opens the door for computer-vision analyses of hallucinations and visual experiences across contexts, including psychedelic experiences and neurological disorders.

65 **Kobi Kremnizer** (showcase talk)
Mathematical Institute, University of Oxford, United Kingdom

Quantum collapse and its application to consciousness: a research programme

I will present a sketch of a research programme to verify or falsify quantum collapse models and then understand how these models could also model awareness/ consciousness. Currently, experiments looking at collapse models focus on microscopic systems where it is very difficult to determine if the experimental results are due to a collapse mechanism or to environmental effects. Using systems which are on the quantum\classical boundary might give a better way to test quantum collapse models. If this first step is successful and quantum collapse models are verified the next step would be to refine the models and see exactly how the structure of matter affects collapse rates which are controlled by a stochastic field. Building on that we would be able to analyse the structure of the stochastic collapse field around brains and relate that to more standard human phenomenology. The first step of this programme is based on joint work in progress with Ghosh, Gubinneli and Contera.

September 30 – October 4, 2025
Hokkaido University

At the center of a discussion with Werner Heisenberg was "the shared problem and the desire to grasp and clarify it. One carefully approached it, passed it to the other, like in a friendly table tennis game, where both made sure that the ball remained in play. All the attention was focused on truly understanding the other and to avoid letting him stumble sophistically over his poor and inadequate expression. One could stutter, one could speak vaguely, even incomprehensibly, and he would guess what one actually wanted to say, would repeat it in his own different words, so that one could often exclaim with pleasure: 'Yes, exactly that...!'. During such an (...) intense exchange of thoughts, the ideas and concepts sharpened, so that their contours became recognizable more clearly."

A former collaborator of W. Heisenberg

"First and foremost, there must be ease, relaxation, and a general sense of permissiveness. The world in general disapproves of creativity, and to be creative in public is particularly bad. Even to speculate in public is rather worrisome. The individuals must, therefore, have the feeling that the others won't object. (...) It seems necessary to me, then, that all people at a session be willing to sound foolish and listen to others sound foolish."

Isaac Asimov, How do people get new ideas, 1959