

Models of Consciousness 5

Minimal Phenomenal Experience 3



Organising and Supporting Institutions:



上海科技大学
ShanghaiTech University



Programme

Table of Contents

Welcome	4
Everything you need to know	5
The conference team	6
Excursion	8
Conference dinner	9
Public evening event	10
Discussion sessions	11
Conference venue	12
Area map	16
Programme schedule	18
Talk titles and abstracts	26
Space for personal notes	82

Welcome!

It is a great pleasure to welcome you all to the fifth instalment of the *Models of Consciousness* conference series (MoC5) and the third *annual MPE meeting*. Following the conferences at *Oxford University* (2019, MoC1), online (2021, MoC2), *Stanford University* (2022, MoC3) and again *Oxford University* (2023, MoC4), it is a great pleasure to host this year's conference at *Bamberg University*, by kind invitation of Bamberg's *Institute for Psychology*.

Consciousness science is at a crossroads. The rapid development of Artificial Intelligence (AI) brings new questions and new methodological desiderata to the field. Mathematical structures and spaces of conscious experiences offer novel theoretical and empirical opportunities. And minimal phenomenal experience is introducing what could well be a new paradigm for conceptual and theoretical engineering.

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:

- ▶ Consciousness of Artificial Intelligence
- ▶ Mathematical Structures and Spaces of Conscious Experiences
- ▶ Minimal Phenomenal Experience

We hope to be able 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 Bamberg! — *The organisers*

Everything you need to know...

... initially :-)

Conference Venue

Buildings **U2** and **U5** of the **University of Bamberg**

Address: An der Universität 2
96047 Bamberg, Germany

[Google Maps Link](#) - [OpenStreetMap Link](#)



Venue of Consciousness of AI Evening Event:

Building **AULA** of the **University of Bamberg**

Address: Dominikanerstraße 2A
96049 Bamberg, Germany

[Google Maps Link](#) - [OpenStreetMap Link](#)



Live Updates...

... regarding the schedule, discussion sessions and socializing will be available at:

<https://twitter.com/moc5conference>



Wi-Fi Access:

SSID: **ModelsofConsciousness2024**

Password: **MoC@uniba**



Conference Team

Organisers



Johannes Kleiner

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

Don't hesitate to let us know in case you have any questions or are in any need of assistance!



Jonathan Mason

Oxford Mathematics of Consciousness and Applications Network



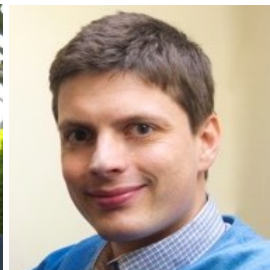
Sophie Taylor

Queensland University of Technology



Wanja Wiese

Institute for Philosophy II, Ruhr University Bochum



Robert Prentner

Institute of Humanities, ShanghaiTech University



Keenan Down

Cambridge Consciousness & Cognition Lab
Queen Mary University of London



Robin Lorenz

Quantinum



Sascha Benjamin Fink

Centre for Philosophy and AI Research {PAIR}
Friedrich-Alexander University Erlangen-Nürnberg

Advisory Board



Ian Durham

Department of Physics, Saint Anselm College

Lenore Blum

Distinguished Career Professor of Computer Science Emerita
Carnegie Mellon University



Claus H. Carstensen

Head of Psychological Methods of Educational Research
University of Bamberg

Vincent C. Müller

Director of the Centre for Philosophy and AI Research {PAIR}
Friedrich-Alexander University Erlangen-Nürnberg



Yakov Kremnitzer

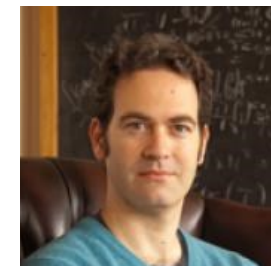
Oxford Mathematics of Consciousness and Applications Network
Mathematical Institute, University of Oxford

Thomas Metzinger

Director, Minimal Phenomenal Experience Research Network
Institute for Theoretical Philosophy, Johannes Gutenberg
University of Mainz



Adjunct Fellow, Frankfurt Institute for Advanced Studies



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.

For MoC5, we are excited to host an excursion that features technology, modern art and local culture:



Morning: Guided tour of the German Museum of Masterpieces of Science and Technology in Munich.

Afternoon: Participants can choose to either visit the iconic Oktoberfest in Munich, or visit the famous Pinakothek der Moderne museum of modern art.

Image © München Tourismus, Annika Mittelmeier



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 in one of Bamberg's most romantic restaurants, located within the historic *Michaelsberg Abbey*. The restaurant has been selected to feature both haut cuisine and affordable options for students.

Location: **Ristorante da Francesco Bamberg**

Address: Am Michelsberg 10f
96049 Bamberg



Day: **Wednesday, October 2nd**
Time: **Starting 19:00**

Payment: *Please note that the conference dinner is **not** included in the conference fee. The dinner is paid per person at the restaurant at the end of the evening.*

Registration: *Please note that due to space constraints, the conference dinner is **by registration only**.*



Public Evening Event


The conference features a public evening event

Consciousness of Artificial Intelligence

Everyone is cordially invited to attend! — The programme is still in the making and will be distributed separately.

Day: **Tuesday, October 1st**
 Time: **19:00 - 21:00**
 Location: **Aula of Bamberg University**
 Address: **Dominikanerstraße 2A,
 96049 Bamberg**



PS: As a technical note to finding the evening event location, this  is one of Bamberg's most iconic breweries. It serves a particular kind of smoked beer, which is customarily consumed while standing in the street in front of the brewery, next to the other N breweries in the same street. The brewery, called "Schlenkerla", is right next to the evening event location and can be helpful as a point of reference for finding the evening event.



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

#MoC5

for "Models of Consciousness 5".



Joining as a participant?

MoC5



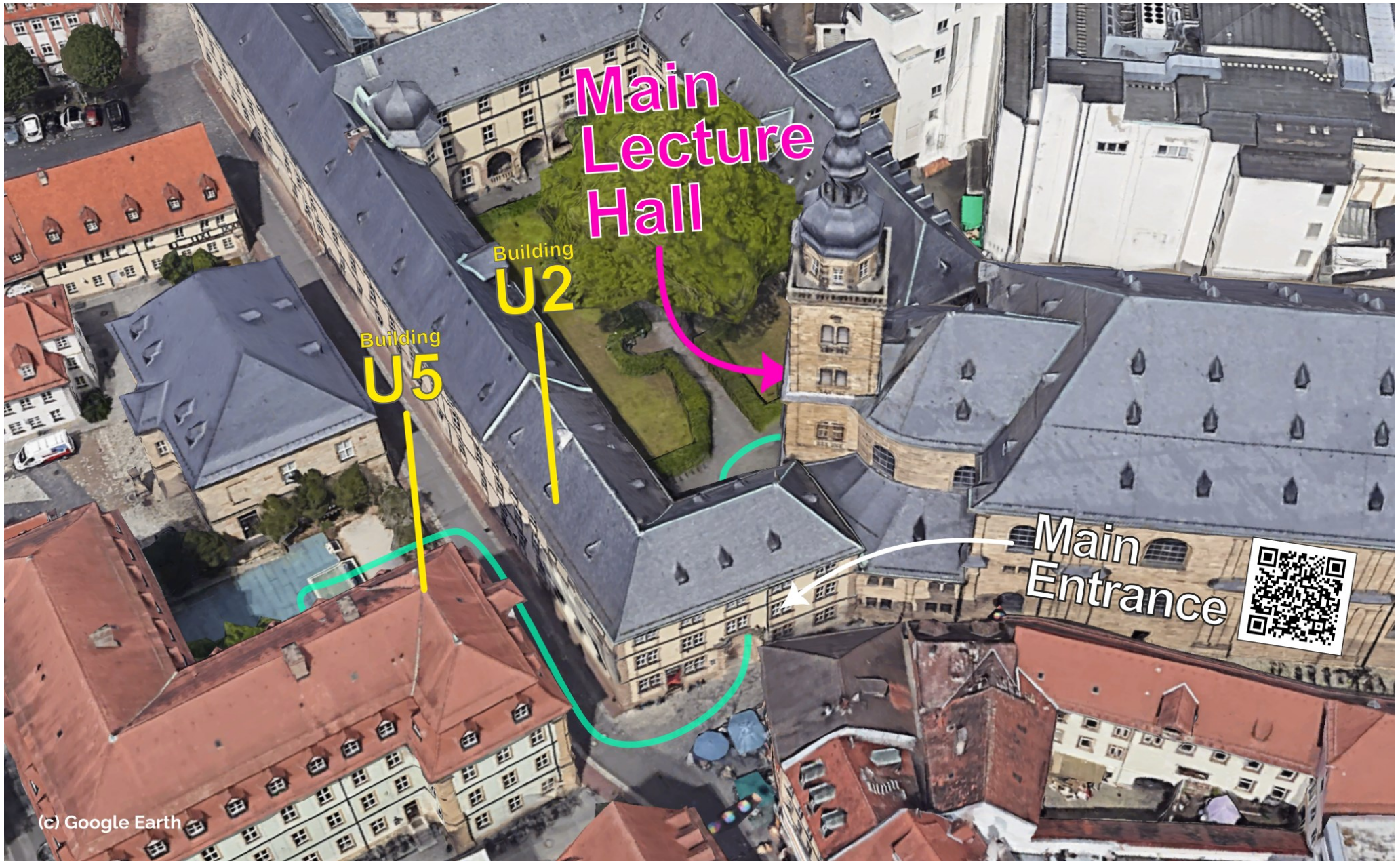
We hope discussion sessions at the conference are enjoyable and fruitful! Ultimately, this requires an atmosphere of trust and tolerance supported by all attendees of a session, much like described in the two quotes by Isaac Asimov and a collaborator of Werner Heisenberg on the back of this booklet.



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".

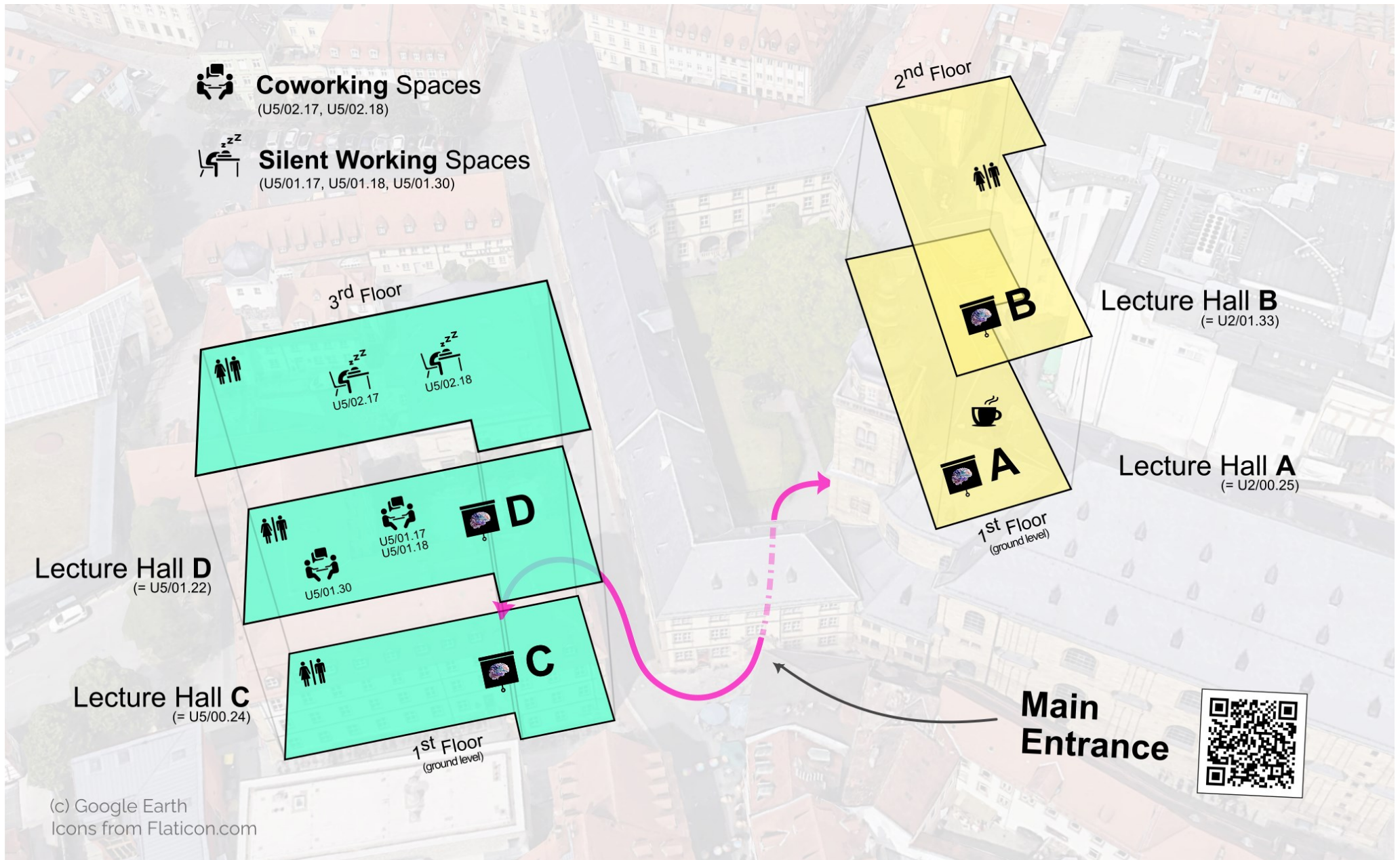
Conference Venue

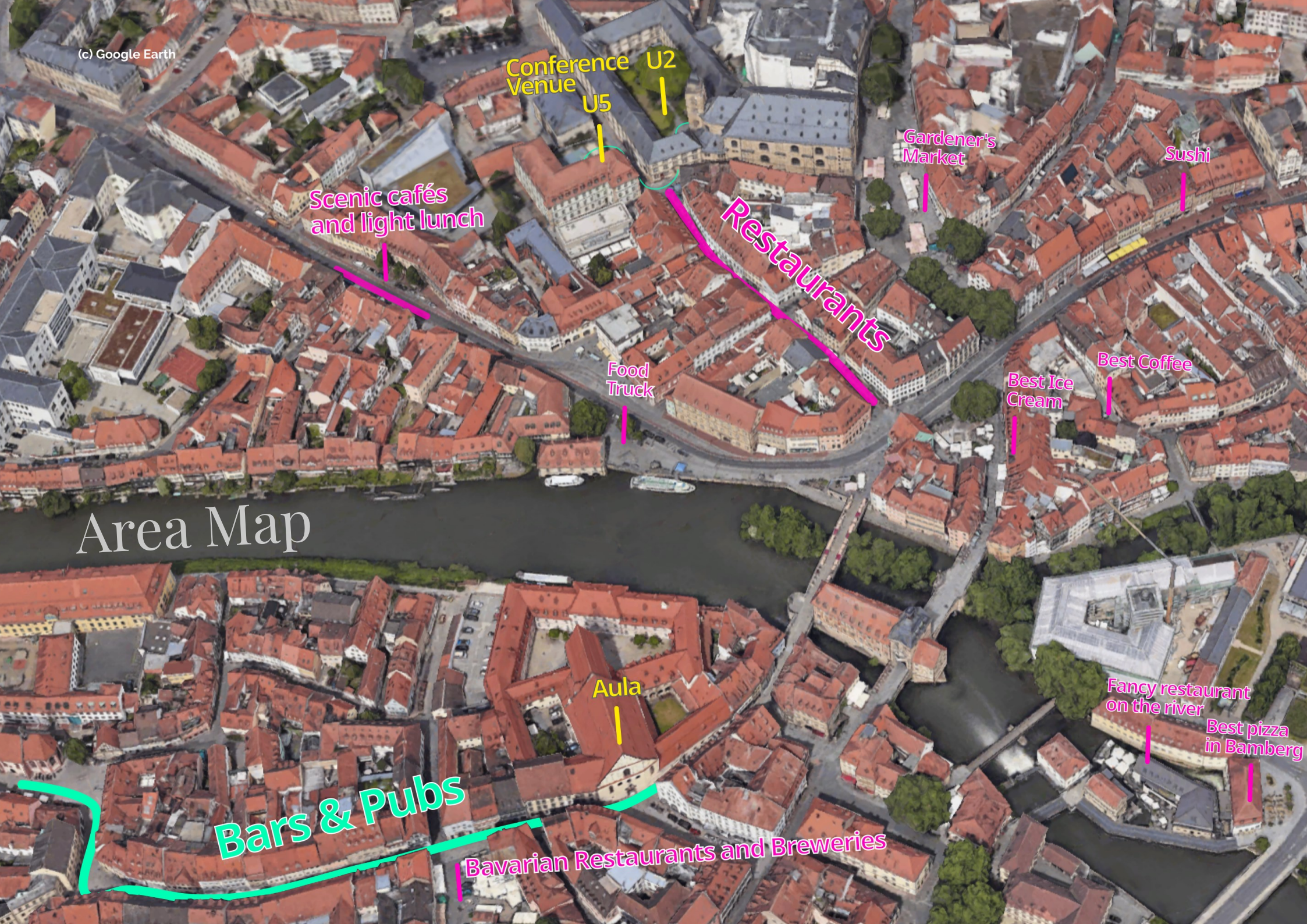


Address: An der Universität 2, 96047 Bamberg, Germany

Main Lecture Hall: U2/00.25 (Room 00.25 on the ground floor of Building U2)

Conference Venue





Conference Venue

U2

U5

Scenic cafés and light lunch

Gardener's Market

Sushi

Restaurants

Food Truck

Best Coffee

Best Ice Cream

Area Map

Aula

Fancy restaurant on the river

Best pizza in Bamberg

Bars & Pubs

Bavarian Restaurants and Breweries

	Tuesday		Tuesday (cont.)	
8:00 - 8:30	Registration (A)			
8:30 - 8:45	Welcome & Opening (A + B)			
8:45 - 9:45	Keynote (A + B) 1 Henry Shevlin			
9:45 - 10:15	Coffee (Courtyard)			
10:15 - 10:40	2 Joscha Bach (A)	5 Ian Durham (B)	8 Aureli Soria-Frisch (C)	11 Miguel Ángel Sebastián (D)
10:40 - 11:05	3 Wanja Wiese (A)	6 Vincent C. Müller (B)	9 David Silverstein (C)	12 Dezhi Luo (D)
11:05 - 11:30	4 Aida Elamrani (A)	7 Niccolo Negro (B)	10 Federico Seragnoli (C)	13 John Sanfey (D)
11:30 - 12:30	Discussion Sessions (A, B, C, D)			
12:30 - 13:30	Lunch			
13:30 - 14:30	Keynote (A + B) 14 Charlotte Gauvry			
14:30 - 15:00	Coffee (Courtyard)			
15:00 - 15:25	15 Kenneth Williford (A)	18 M. Kriegleder & M. Noichl (B)	21 Kobi Kremnizer (C)	24 HyeYoung Kim (D)
15:25 - 15:50	16 David Rudrauf (A)	19 Renee Ye (B)	22 Yair Pinto (C)	25 Nicolas Loerbroks (D)
15:50 - 16:15	17 Grégoire Sergeant-Perthuis (A)	20 Oisín Clancy (B)	23 Andrew Knight (C)	26 Graham Findlay (D)
16:15 - 17:30	Discussion Sessions (A, B, C, D)			
19:00 - 21:00	Public Evening Event Consciousness of AI (Aula)			

	Wednesday	Wednesday (cont.)
8:30 - 9:30	Keynote (A + B) 27 Heleen A. Slagter	
9:30 - 10:00	Coffee (Courtyard)	
10:00 - 10:05	MPE-1 Yizhi Li (A + B) Lightning Talk	28 TBA (C) Lightning Talk
10:05 - 10:30	MPE-2 David Higgins (A + B)	29 Sascha Benjamin Fink (C)
10:30 - 10:55	MPE-3 Ema Demšar (A + B)	30 Lukas Kob (C)
10:55 - 11:20	MPE-4 Yair Dor-Ziderman (A + B)	31 Daniel Weger (C)
11:20 - 12:30	Discussion Sessions (A, B, C, D)	
12:30 - 13:30	Lunch	
13:30 - 13:55	MPE-5 Toby Woods (A + B)	32 Tim Ludwig (C)
13:55 - 14:20	MPE-6 Cyril Costines (A + B)	33 Anna Eiserbeck (C)
14:20 - 14:45	MPE-7 Christopher Timmermann (A + B)	34 Rony Hirschhorn (C)
14:45 - 15:15	Coffee (Courtyard)	
15:15 - 15:40	MPE-8 Ronald Sladky (A + B)	35 Matthias Borgstede (C)
15:40 - 16:05	MPE-9 Anand Prasad (A + B)	36 Gaspard Fougea (C)
16:05 - 16:30	MPE-10 Romy Beauté & Catriona Osborn Moar (A + B)	37 Daniel Müller (C)
16:30 - 17:30	Discussion Sessions (A, B, C, D)	
19:00	Conference Dinner	

	Thursday	Thursday (cont.)	
8:30 - 9:00	Keynote (A + B) 38 Ute Schmid		
9:00 - 9:30	Keynote (A + B) 39 Christoph Benzmüller		
9:30 - 10:00	Coffee (Courtyard)		
10:00 - 10:25	MPE-11 Edmundo Lopez-Sola (A)	40 Keenan Down (B)	43 Paul Skokowski (C)
10:25 - 10:50	MPE-12 Ishan Singhal (A)	41 Eric Elmoznino (B)	44 John Barnden (C)
10:50 - 11:15	MPE-13 Lars Sandved-Smith (A)	42 Michael Timothy Bennett (B)	45 Nicolas Kuske (C)
11:15 - 12:30	Discussion Sessions (A, B, C, D)		
12:30 - 13:30	Lunch		
13:30 - 14:30	Keynote (A + B) 46 Manuel Blum		
14:30 - 15:00	47 Leonard Dung - Barbara-Wengeler-Prize (A + B)		
15:00 - 15:30	Coffee (Courtyard)		
15:30 - 15:35	48 TBA (A) Lightning Talk	52 Etienne Jacques (B) Lightning Talk	56 Robert Tromm (C) Lightning Talk
15:35 - 16:00	49 Lenore Blum (A)	53 Michael Schnabel (B)	57 Paweł Motyka (C)
16:00 - 16:25	50 Albert Newen (A)	54 Sylvia Liu (B)	58 Borjan Milinkovic (C)
16:25 - 16:50	51 Jonathan Mason (A)	55 Chetan Prakash (B)	59 Jakub Vohryzek (C)
16:50 - 17:00	Short Break		
17:00 - 18:00	Keynote (A + B) 60 Joel Frohlich		
18:00 - 19:00	Discussion Sessions (in case of interest) (A, B, C, D)		

	Friday		Friday (cont.)	
8:30 - 9:30	Keynote (A + B) 61 Nihat Ay			
9:30 - 10:00	Coffee (Courtyard)			
10:00 - 10:05	62 Francesca Castaldo (A) Lightning T.	66 Adel Chaibi (B) Lightning Talk	70 Iulia-Maria Comşa (C) Lightning T.	74 Uziel Awret (D) Lightning Talk
10:05 - 10:30	63 Giulio Ruffini (A)	67 Marc Heimann (B)	71 Robert Prentner (C)	75 Hans-Jürgen Graf Grote (D)
10:30 - 10:55	64 Carlotta Langer (A)	68 Igor Balaz (B)	72 Camilo Miguel Signorelli (C)	76 Serena Doria (D)
10:55 - 11:20	65 Martin Biehl (A)	69 C. Turner & S. G. Mayson (B)	73 Jan Erik Bellingrath (C)	77 Sophie Taylor (D)
11:20 - 12:30	Discussion Sessions (A, B, C, D)			
12:30 - 13:30	Lunch			
13:30 - 14:00	Showcase Talk (A + B) 78 Andrew Lee			
14:00 - 14:30	Showcase Talk (A + B) 79 Computational Phenomenology of Pure Awareness Prize			
14:30 - 15:00	Coffee (Courtyard)			
15:00 - 15:25	80 Maxine Collard (A)	83 Joachim Keppler (B)	86 Vishal Kapoor (C)	
15:25 - 15:50	81 Johannes Kleiner (A)	84 Marc Ebner (B)	87 Abhilash Dwarakanath (C)	
15:50 - 16:15	82 Johannes Brinz (A)	85 Ouri Wolfson (B)	88 Will Moorfoot (C)	
16:45 - 17:15	Closing Discussion (A)			

Friday, October 18	
11:00 - 12:00	Keynote (Online) 90 Martine Nida-Rümelin

Please note that due to personal reasons, the keynote of Martine Nida-Rümelin had to be postponed. It will be held online on Friday, October 18, at 11am CET. A link to join the meeting will be distributed by email shortly in advance.

Titles and Abstracts

1 Henry Shevlin (Keynote)

University of Cambridge
Leverhulme Centre for the Future of Intelligence
Institute for Technology & Humanity

AI and the future of consciousness science

Recent rapid progress in artificial intelligence has prompted renewed interest in the possibility of consciousness in artificial systems. This talk argues that this question forces us to confront troubling methodological challenges for consciousness science. The surprising capabilities of large language models provide reason to think that many, if not all, cognitive capabilities will soon be within reach of artificial systems. However, these advancements do not help us resolve strictly metaphysical questions concerning substrate-independence, multiple realizability, or the connection between consciousness and life. Ultimately, I suggest that these questions are likely to be settled not by philosophical argument or scientific experimentation, but by patterns of interactions between humans and machines. As we form valuable and affectively-laden relationships with ever more intelligent machines, it will become progressively harder to treat them as non-conscious entities. Whether this shift will amount to a vindication of AI consciousness or a form of mass delusion remains far from obvious.

2 Joscha Bach

LiquidAI
California Institute for Machine Consciousness

What does it take to build conscious Artificial Intelligence?

Artificial Intelligence is not just an engineering discipline, but also a philosophical project, aimed at the naturalization of the mind. By allowing to build testable models, AI offers a metaphysical framework and a methodology for defining and exploring mental representations, perception, agency, self modeling, attention and systemic models of psychology. At the same time, very little practical AI research is concerned with understanding consciousness and the mind. Starting from the epistemological position of computationalist functionalism, we will discuss the phenomenology of consciousness (especially second order perception and "nowness") in conjunction with hypotheses about the functionality of consciousness in a cognitive system from an AI perspective. What does a test for consciousness look like? Can we determine whether

existing AI systems possess conscious phenomenology? Is consciousness functionally a learning algorithm for self organizing systems? Is consciousness simple enough to discover it via a systematic search, or will it turn out to be as complex as eg. a biological cell? Can insights from AI experiments inform our understanding of consciousness in biological systems? What are the ethical and practical implications of building conscious AI?

3 Wanja Wiese

Ruhr University Bochum
Institute for Philosophy II

Negative indicators of consciousness in artificial systems

In the absence of an empirically confirmed theory of consciousness that can be applied to artificial systems, it is promising to explore indicators that provide at least some evidence for the presence of consciousness in artificial systems. This approach looks for probability-raising indicators of consciousness that make it more likely that a given system is conscious.

I argue that this approach also needs an explicit account of negative indicators of consciousness. A negative indicator is a factor that is itself at most slightly probability-raising; its absence, however, can negatively affect the probability that a system is conscious substantially. I propose a distinction between two types of negative indicators: indirect and direct negative indicators. While the absence of indirect negative indicators influences the extent to which other indicators are probability-raising, the absence of direct negative indicators lowers the probability that an artificial system is conscious, regardless of which other indicators are present.

4 Aida Elamrani

ENS-PSL
IJN, DEC

Understanding Artificial Consciousness: Clarifying Concepts and Advancing Research

Artificial Consciousness is a burgeoning field, promising significant advancements in modelling both artificial intelligence and human cognition. However, this interdisciplinary domain grapples with nascent and ambiguous terminology, often leading to misunderstandings and apprehensions. The first aim of this talk is to clarify the discourse by proposing key definitions and elucidating their historical and scientific context. The second aim is to enhance technical understanding of current research in the field, particularly focusing on implementations of the Attention Schema Theory and Transformer architectures. Building on this foundation, the last part of the talk seeks to alleviate fears surrounding AC and promote further research by highlighting the key advantages these models offer.

5 Ian Durham

Saint Anselm College
Department of Physics

Quo vadis consciousness science?

On the importance of mathematical approaches

Are mathematical approaches to consciousness science really necessary? Is a mathematical theory of consciousness even possible? The answers to those two questions may not actually be the same. It is entirely plausible that a mathematically sound theory of consciousness will never be found while, at the same time, rigorous mathematical models of certain aspects of conscious experience open new and fruitful areas of research. Regardless of one's belief or non-belief in physicalism, anything purporting to call itself 'science' should be well-grounded in a mathematically consistent framework. But the form of that framework is equally important because it can help clarify, often in non-obvious ways, the nature of the system it purports to model. In this talk I will present several examples to illustrate this point and will discuss what mathematical approaches can tell us about physicalist claims on consciousness.

6 Vincent C. Müller

Friedrich-Alexander-Universität Erlangen-Nürnberg
Centre for Philosophy and AI Research {PAIR}

Syneidisi allergy: symptoms, etiology, therapy

Unlike everyone else at this conference, I have this allergy; I get tearing eyes and spells of dizziness when people talk about Syneidisi, the deepest philosophical problem of all ... their eyes glaze over and they cannot express how important it is, indeed they love discussing what the nature of the mysterious problem is, or whether it even exists at all. Syneidisi seems to be a thing I have, or fleeting states of what it is like, or something something, but it is the only thing I really know, though no-one else can know it, and it cannot be described. In addition, it is fundamental to epistemology, the philosophy of mind, the philosophy of AI, almost everything. So mysterious. - In the presentation, I will discuss what causes this allergy of mine, whether there is a cure, or whether I should just stay clear of the allergen. Perhaps you should, too?

7 Niccolo Negro

Tel Aviv University
School of Psychological Sciences

Towards a Lakatosian science of consciousness

The neuroscience of consciousness is undergoing a significant empirical

acceleration thanks to several adversarial collaborations that intend to test different predictions of rival theories of consciousness. In this context, it is important to pair consciousness science with confirmation theory, the philosophical discipline that explores the interaction between evidence and hypotheses, in order to understand how experiments challenge or validate theories of consciousness.

In this paper, I examine this relationship by adopting a Lakatosian lens and propose that Lakatos' philosophy of science can aid consciousness scientists to better interpret adversarial collaborations in consciousness science and, more generally, to develop a confirmation-theoretic model of theory-appraisal in this field.

I do so by suggesting that such a model be built upon three Lakatos-inspired criteria for assessing the relationship between empirical evidence and theoretical predictions: a) the model should represent the distinction between prediction and accommodation; b) the model should represent the structural relevance of predictions; c) the model should represent the boldness of the predictions. I argue that this model has both normative and descriptive virtues, and can move the debate forward by acknowledging that theory-appraisal needs to consider the diachronic development of theories and their logical structure.

8 Aureli Soria-Frisch

Starlab Barcelona
Neuroscience

The two-dimensional characterization of consciousness through Graph Neural Networks

The development of metrics that can be used as neural correlates of consciousness (NCC) in clinical settings is an open question. The experimental apparatus and time-consuming recording procedures of existing ones (Cassini2013) limit its clinical applicability. Moreover the existent metrics do not reflect the continuous nature of consciousness state changes as represented in the Awareness-Wakefulness theoretical space (Laureys 2005). We propose an alternative to ease the clinical application of NCCs.

We have used an EEG dataset including recordings of 20 subjects in 2 sessions of 8h per subject with 60 EEG channels. We have modified the deep graph neural network (GNN) denoted as BrainNet (Kawahara2017), which uses connectivity features as input data, to allow the prediction of sleep stages in the aforementioned two-dimensional space. We have performed a classical three-way holdout validation procedure for hyper-parameter tuning and performance evaluation. The modified BrainNet successfully represents consciousness in the Awareness-Wakefulness space. We found on average a successful RMSE of 2.42 in the Awareness component, whereas we achieved an RSME of 1.78 in the Wakefulness one. The results show the first success in the continuous characterization of consciousness states based on spontaneous EEG, which facilitates its usage in the clinical setting.

9 David Silverstein

Independent Researcher

Proximal intention and hypothesized generators of the readiness potential

The readiness potential (RP) has been observed to occur with EEG/MEG immediately prior to self-paced decision making and motor action, but the cause of this neural correlate is under debate. Some have suggested that the presence of a RP before conscious recognition and report of a decision to act indicates an absence of conscious free will. EEG measures RP signals at the Pre-SMA, SMA and motor cortices. Additionally, human electrode studies have shown evoked responses correlated with the RP in the basal ganglia and thalamus. A hypothesis is proposed that subcortical structures and cortico-basal ganglia-thalamo-cortical (CBGTC) loop dynamics play important roles for inhibition and control during RP generation as well as motor preparation. To investigate this, a computational dynamical systems model was developed with relevant brain areas as neural masses and the connections between them. Deciding when to act includes cognitive control, response inhibition and selection. The model predicts possible abstract neural dynamics of brain areas during the early and late RP and the transition between them. Simulations suggest that the early RP having a variable time duration indicates preparation with intention to act prior to decision commitment, but does not represent action initiation in the absence of free will.

10 Federico Seragnoli

Lausanne University, Faculty of Psychology, Psychiatry
Geneva University Hospital, Psychedelic Assisted Psychotherapy Unit

Metacognitive feelings of meaning in psychedelic induced altered states of consciousness

Metacognition can be used to study various non-ordinary states of consciousness like meditation, hypnosis, lucid dreaming and ecstatic epilepsy. Metacognition presents a declarative facet, defined as the declarative knowledge on how cognition works (e.g. epistemic beliefs), and a procedural facet, defined as monitoring and control functions for cognition to modulate itself. Within the procedural facet, the concept of metacognitive feelings (e.g. tip of the tongue feeling) related to cognitive actions (Proust, 2013) is described, giving special attention to metacognitive outcome-oriented feelings, such as the "eureka" or "aha Erlebnis" experience, which reflect a profound feeling of meaning. In the context of the modern resurging interest in Psychedelic Assisted Therapy (PAT), metacognitive feelings are linked to psychedelic induced non-ordinary states of consciousness by proposing a neuro-cognitive framework to study their interaction. In particular, a parallel between the

"metacognitive feeling of meaning" and the noetic quality of the mystical-type experience potentially occurring in PAT is described. By questioning the interplay between metacognition processes and psychedelic-induced non-ordinary states of consciousness, we can gain understanding into fundamental mechanisms of PAT and into the nature of consciousness.

11 Miguel Ángel Sebastián

National Autonomous University of Mexico
Institute for Philosophical Research

Minimal Self Consciousness as Perspectival Information

This paper develops the idea that Minimal Self-Consciousness (MSC) corresponds to perspectival de se information, underlying our capacity to self-locate in logical space. Phenomenal information, even in its minimal form, is more fine-grained than possible world information, constraining uncertainty not only about the world we inhabit but also about how, where, and when we are within it.

Awareness in experience is subjective and naturally expressed using the first-person pronoun. However, some experiences seem to defy this because the self is not an object of awareness, and they do not involve any specific time or spatial location. I argue that awareness in experience is a form of first-personal information (perspectival de se) concerning the state we are in, offering a semantic characterization in terms of self-location in logical space. This proposal reconciles the view that experience is subjective, and thus expressed with the first-person pronoun, with the idea that the pronoun's reference is not constitutively part of phenomenal information (the experience can be "selfless"). It also aligns with the idea that the experience can be timeless and spaceless: representations of the self, time, and spatial location are not necessary conditions for consciousness to occur.

12 Dezhi Luo

University of Michigan
Weinberg Institute for Cognitive Science

Perceptual Metacognition, Memory, and Machine Self-Consciousness

Self-consciousness is central to various considerations regarding consciousness in artificial intelligence. In this talk I adopt computational functionalism and interpretive sensory-access theory as working hypotheses to offer a mechanistic proposal which supports the computational realizability of self-consciousness. To begin, I distinguish two kinds of self-consciousness: the sense of personal ownership and introspection. I argue that, contrary to traditional phenomenological views, they may rely on a generic perceptual metacognitive process while differentiated by their mnemonic contents. In

accordance with the computationally-informed perceptual reality monitoring (PRM) theory, the subjectiveness of perceptual awareness is attributed to memories retrieved following source monitoring and represented in the quality space via medial prefrontal centered networks. Ventromedial PFC has been shown to be responsible for evaluating self-importance of perceptual contents and subsequently guiding conceptual autobiographical retrieval in decision-making, a process proposedly could be implemented by predictive world-modeling architectures. In light with recent findings documenting vmPFC in task-irrelevant perceptual mnemonics, I suggest that such self-referential processes may be part of the implicit perceptual metacognitive mechanism postulated by PRM and is responsible for assessing and rendering familiarities in perceptual awareness underlying personal ownership. The same process, if conducted explicitly and involves episodic contents, would allow introspective access to self-knowledge.

13 John Sanfey

Independent

The simultaneity between experiencing and awareness of awareness

If nothing existed except consciousness, then being unaware of its existence would be identical to no awareness. There is no 'something it is like' to not experiencing the only experienceable thing. Contentless awareness is impossible, the minimum content is some concept of awareness. But if only consciousness existed, we could still imagine idealist and realist explanations for its existence. Either something else caused it, or we did and forgot. We know we are not now causing consciousness deliberately, because if we were consciously causing it, we would be conscious of doing so. The simultaneity between experiencing and knowing that we are experiencing makes both idealist and realist explanations logically possible, which has causal consequences.

Simultaneous causation is not possible in physics, but there is certainly no time interval between experiencer and the experience, and this does have causal consequences. What makes this possible within the laws of physics, is that 'the experienced' is always a model, whether sensory or conceptual. Consciousness only interacts with models, not the unknowable stuff being modelled. When the meaning of a model changes, human behaviour changes, but the physical matter being modelled remains unknowable, and behaves as it has done since the big bang.

14 Charlotte Gauvry (Keynote)

University of Bonn
CST

Hemispherotomy as a borderline case of synthetic consciousness

Recent advancements in neurosurgery, biomedical research, and technological innovations have introduced new candidates for synthetic consciousness, including atypical brain entities such as isolated hemispheres after hemispherotomy, brain organoids, and advanced AI systems. My presentation will focus on the first of these entities: the isolated hemisphere. Unlike the much more famous split-brain case, hemispherotomy, in my view, has been underexplored in consciousness research, despite its significant philosophical implications.

Recent findings from fMRI and EEG studies (e.g., Hawasli et al., 2017; Blauwblomme et al., 2020; Bauer, Gauvry, Rüber et al., preprint; Colombo, Massimini et al. preprint), however, have shown sleep cycles in the isolated hemisphere and increased cortical activity in its default mode network. This activity, coupled with the lesional hemisphere's preserved olfactory connection with the external environment, has sparked intriguing philosophical discussions regarding its potential sentience.

What if the isolated hemisphere is sentient? How can we tell? What theoretical conclusions could be drawn of the observation of a so-called "island of awareness" (Bayne, Seth, Massimini, 2022)? What ethical stance should we adopt? I will argue that the methodological framework introduced by Birch 2024, which I refer to as "valence sentientism," offers a promising approach for addressing this borderline case.

15 Kenneth Williford

University of Texas at Arlington
Department of Philosophy & Humanities

Projective Consciousness Model Research Group: David Rudrauf,

Towards a Mathematical Theory of Pre-Reflective Self-Consciousness: From Metamathematics to the Projective Consciousness Model

Pre-reflective self-consciousness (PRSC) is consciousness's acquaintance with itself, prior to introspection. This talk, the first of a series, lays out some of the history of philosophical theories of PRSC along with the relevant phenomenological and psychological motivating considerations, including the issue of multimodal and diachronic integration (further developed in the second talk in the series), and then sets the stage for the last talk by briefly canvassing the history of key attempts at mathematically modeling PRSC, in particular attempts based on metamathematical self-reference (Hofstadter), circular

structures (Varela, Khromov, Williford), and projective geometry (Bennequin, Rudrauf, Williford, Sergeant-Perthuis). Connections between the three modeling strategies are discussed, and the topic of duality, key to the projective geometrical strategy and to the generalization to homogenous G-spaces presented by G. Sergeant-Perthuis, is highlighted. Models based on non-wellfounded structures capture certain key features of PRSC and subjective character, and these features can be related to the PCM and Sergeant-Perthuis' proposal in a way that merits further exploration. Together, the models capture five key features of subjective character: the sense of individuality (or de se awareness), the capacity for reflection and mental self-attribution, the synchronic unity of multimodal consciousness, point-of-view, and social self-awareness.

16 David Rudrauf
University Paris Saclay
CIAMS
PCM Group

Integrating Multimodal Sensory Information through the Projective Consciousness Model: A Framework for Spatial Perception and Action Planning

The Projective Consciousness Model (PCM) is often mistaken for a model of visual awareness. In fact, PCM provides a framework for understanding how multimodal sensory information, as influenced by embodied constraints, is flexibly integrated with action planning in a unified manner in spatial consciousness. This talk, part of a series, discusses the psychophysical, cognitive, and behavioral motivation for this claim. PCM posits that multimodal perception and imagery are structured according to 3D projective geometry, which dynamically adapts to sensorimotor contexts.

Spatial perception and cognition involve continuous calibration and inference, adapting to sensory ambiguity. Multimodal spatial experiences are consistent with a structuring role of projective geometry. Overall, visual space is organized in a 3D perspectival manner. The PCM explains phenomena like the Moon Illusion and Ames Room Illusion. Pictorial relief can be explained using PCM's underlying projective transformation group. Studies show that congenitally and early blind individuals use spatial imagery and perspective-taking for motion planning and theory of mind. PCM also addresses cross-modal interactions and multisensory integration, explaining phenomena like the Rubber Hand Illusion and full-body illusions. Such integrative process supports optimal stimulus location, eye movement coordination, and goal-directed actions, crucial for perceptual stability and navigation.

17 Grégoire Sergeant-Perthuis
Sorbonne Université

What can a homogeneous space say about PRSC?

The PCM is a particular instance in which the actions of an agent are translated into group actions on its world model. In the case of the PCM, the actions are projective transformations. Our focus is on the properties of spaces endowed with a group action (G-spaces) to capture aspects of pre-reflective self-consciousness (PRSC). If the world model is a homogeneous space, any point in that space can be seen as a potential action of the agent if a reference point is chosen. Despite the simplicity of this observation, it could nevertheless hint towards understanding how an agent could singularize itself in a pre-reflective manner, i.e., find a direct trace of its own presence and existence through the representation of the environment it creates. This reference point can be seen, informally, as a frame of the agent. Furthermore, the Yoneda lemma for G-spaces suggests that these frames correspond to points in space. This identification might underlie the possibility for an agent to rewire its beliefs about the world into beliefs about its frame, thereby enabling reflection and introspection. Finally, G-spaces may be too general to adequately account for the phenomenology of internal representations; however, fiber bundles might provide a solution.

18 Moritz Kriegleder
University of Vienna
Cognitive Science Hub
Maximilian Noichl
Utrecht University
Department of Philosophy
Theoretical Philosophy Group

Mapping Consciousness Science: A Big Data Analysis

The scientific study of consciousness has always been an interdisciplinary endeavour, combining theories and tools from many fields such as psychology, philosophy and computer science. However, the multidisciplinary nature of consciousness science leads to divergent strands of research. With a plethora of models of consciousness available, a critical analysis of overlaps and tensions becomes necessary to map different approaches in the field. The aptly named multiple explanandum problem posits that different models of consciousness actually have different explanatory goals (Vilas et al. 2022). To better understand the landscape and goals of consciousness science, it is necessary to map the relationships between recent publications. Towards this goal, we analysed more than 60,000 papers covering the main strands of theory as defined in recent reviews (Seth & Bayne 2022, Camillo et al. 2021,

Butlin et al. 2023). We use a SOTA-language model to assess similarities in the abstracts of these papers, and on this basis map the current strands of research. Despite claims of interdisciplinary exchange, we find several research clusters that are not well integrated directly into the science of consciousness.

19 Renee Ye

Ruhr-Universität Bochum

Anthropocentrism: The Intimate Framework

In this talk, I challenge the dominant questions surrounding consciousness: 'Which entities besides humans are conscious?' and 'Is X conscious?'. Instead, I advocate prioritizing the relative question 'How is X conscious with respect to human consciousness?' as a more fruitful avenue of inquiry that fosters genuine progress in understanding consciousness. Therefore, I propose a paradigm shift in our approach to human and nonhuman consciousness research by prioritizing the investigation of the relative question.

Anthropocentrism, the deeply-ingrained and elusive perspective inherent to humans, significantly influences all aspects of their existence, including consciousness. I redefine anthropocentrism as the fundamental human cognitive framework and explore its significance and potential biases. While often considered unhelpful and harmful, I argue that anthropocentrism can also serve as a benign and necessary feature of comparative consciousness research. To achieve this, I present a nuanced taxonomy of anthropocentrism and suggest methods to address its challenges. I distinguish between Pernicious Anthropocentrism and Benign Anthropocentrism, encouraging researchers to embrace the latter while avoiding the pitfalls of the former. By adopting the relative question and understanding the role of anthropocentrism, I present new models for studying human and nonhuman consciousness.

20 Oisín Clancy

Compassion Institute

Space of Possible Minds: The landscape of consciousness

The space of possible minds is vast. There are over two million biological species on planet earth currently identified, along with an unknown amount more, all of which may have consciousness. Humans have a great diversity of minds within a single species. Artificial digital minds are being created around the world and will soon populate society. Out in the universe we can envisage aliens who may possess states of being vastly different from those in our own solar system. There may even be entities that inhabit the world which are simply beyond our current understandings. Spirits, DMT entities, human-AI hybrids, chimeras, xenobots, etc. are all contenders for other types of exotic minds that have already been posited. There is a need to deeply reflect upon

this rich tapestry of minds in order to develop an ontologically sincere understanding of consciousness. The nature of the exploration is full of abstract objects, structures, and relationships; all things that mathematics provides strong conceptual definitions for. Branches such as differential geometry, topology, category theory may provide a system of formal concepts that are beneficial in allowing us to fully research these landscapes of consciousness.

21 Kobi Kremnizer

University of Oxford
Mathematical Institute
OMCAN

Quantum-Classical transitions, collapse models, and possible applications to modelling consciousness

It is very possible that there is new physics in the quantum-classical transition. Recent research by Contera, Ghosh and collaborators is pointing at new behaviour of matter in certain scales and conditions which are relevant to biological matter. A promising way of modelling this is by using quantum collapse models. These combine a quantum wave function and a classical stochastic field interacting with each other. The interaction might be controlled not only by mass but also by structural properties of matter. The structure of matter determines an effective mass which in turn controls the strength of the classical stochastic field. This controls the rate of decoherence of the quantum wave function. Such models can be relevant to consciousness. It is possible to argue from both scientific and philosophical perspectives that consciousness cannot emerge but requires new physics. These quantum collapse models can be used to model consciousness with the classical stochastic field being a field of (proto-)awareness. The strength and structure of the field is controlled by the structure of matter: brains cause very structured experience. The field in turn ensures that structured matter behaves mostly classically: brains are mostly classical. This also points to new computational abilities of matter.

22 Yair Pinto

University of Amsterdam
Cognitive Psychology

How can we test if consciousness collapses the wavefunction?

First, I will outline that there is a measurement problem and that the notion that consciousness causes the collapse of the wavefunction follows directly from the empirical and theoretical foundations of quantum physics. I will argue that there are three possible explanations of the measurement problem. Denialism (which entails super-determinism and the many worlds theory), objective collapse and subjective collapse theories. I will outline that denialism is counter-factual. Moreover, objective collapse theories are not in line with the theory of

quantum physics, and have no empirical support so far.

Second, and the main focus of the talk, I will discuss how this controversial notion of 'conscious causes collapse' can be investigated. I will first outline that ideas along the lines of 'delayed choice quantum erasure' will not work. Subsequently, there will be an interactive discussion to explore which paths are open to test if consciousness collapses the wavefunction. Here the main focus will be on ideas of multipartite versus sequential bipartite entanglements, along the lines of Renou et al.'s (2021) paper.

The main goal of this talk is to outline how "consciousness collapses" can be investigated. This idea should not be taboo, but should be empirically confronted head-on.

23 Andrew Knight

Independent

Artificial Intelligence Cannot Be Conscious

The possibility of algorithmic consciousness depends on the assumption that conscious states can be copied or repeated by sufficiently duplicating their underlying physical states, leading to a variety of paradoxes, including the problems of duplication, teleportation, simulation, self-location, the Boltzmann brain, and Wigner's Friend. In an effort to further elucidate the physical nature of consciousness, I challenge these assumptions by analyzing the implications of special relativity on evolutions of identical copies of a mental state, particularly the divergence of these evolutions due to quantum fluctuations. By assuming the supervenience of a conscious state on some sufficient underlying physical state, I show that the existence of two or more instances, whether spacelike or timelike, of the same conscious state leads to a logical contradiction, ultimately refuting the assumption that a conscious state can be physically reset to an earlier state or duplicated by any physical means. Several explanatory hypotheses and implications are addressed, particularly the relationships between consciousness, locality, physical irreversibility, and quantum no-cloning.

24 HyeYoung Kim

Ecole Normale Supérieure
Husserl Archive

Can a machine think?

This paper explores various ways of viewing artificial consciousness and poses again the fundamental question of whether a machine can be conscious and what that might mean for us. To answer this question, it is required to reflect on the idea of consciousness, i.e. what it means to be 'conscious', as well as the human-machine relation. In the era of artificial intelligence, which is 'more' than a washing machine that serves as a tool that functions basically as a 'slave', it

seems to be necessary to understand the 'meaning' of a machine and our relationship to it from a different perspective which will encourage us to understand ourselves from a new perspective as well.

25 Nicolas Loerbroks

Ruhr University Bochum

Computation, Consciousness, and Teleology

Kleiner (2024) shows that consciousness requires mortal computation. As a first approximation, a mortal computation is characterized by not being defined with respect to an Instruction Set Architecture (ISA), as opposed to contemporary digital computers. It appears reasonable to ask what it means for the brain to implement a computation, i.e. to put this result in the context of a philosophical account of physical computation. According to Piccinini & Anderson (2024), a physical system implements a computation robustly if there are transitions between physical states that map onto the transitions between the abstract computational states, and the physical states bear the same amount of information about the computational trajectory of the system as those computational states. A physical system implements a computation strongly if the physical system is, in addition, usable by agents as means to their computational ends (like a PC). I argue that a mortal computation is characterizable as being implemented robustly, but not strongly. Conversely, Anderson & Piccinini argue that nervous systems implement those computations strongly that correspond to the organisms' teleological functions. And, indeed, behaviours associated with consciousness such as actions perceived as deliberate or creative appear to be characterizable as being not teleological.

26 Graham Findlay

University of Wisconsin-Madison

Dissociating Artificial Intelligence from Artificial Consciousness

Recent developments in machine learning raise the question of artificial consciousness: if a computer were functionally equivalent to a human, having the same cognitive abilities, would it experience sights, sounds, and thoughts, as we do when we are conscious? Answering this question in a principled manner can only be done on the basis of a theory of consciousness that is grounded in phenomenology and its essential properties, translates them into measurable quantities, can be validated on humans, and can be extrapolated to any physical system. Integrated Information Theory (IIT) fits these criteria, and provides principled tools to determine whether a system is conscious, to what degree, and the content of its experience. We consider pairs of systems constituted of simple Boolean units, one of which - a basic computer - simulates the other with full functional equivalence. By applying the principles

of IIT, we demonstrate that (i) two systems can be functionally equivalent without being phenomenally equivalent; (ii) that this conclusion applies no matter how one ‘black-boxes’ the computer’s units; and (iii) that even certain Turing-complete systems, which could theoretically pass the Turing test and simulate a human brain in detail, would be negligibly conscious.

27 Heleen A. Slagter (Keynote)

Vrije Universiteit Amsterdam

Institute for Brain and Behavior

Department of Experimental and Applied Psychology

Meditation and the scientific study of consciousness: from consciousness to pure awareness to complete cessation of awareness

In this talk, I will explore the value of meditation as a tool in the scientific study of consciousness. I will first introduce three main meditation styles that have been examined most in scientific research – focused attention, open monitoring and non-dual awareness meditation – and present an overarching theoretical framework rooted in the concept of the predictive brain. This framework offers a novel explanation for not only the progressive effects of meditation on standard cognitive processes but also for the more atypical experiences reported by practitioners, including the experience of non-dual or pure awareness, a state of awareness characterized by the loss of sense of self, space, and time. I will then discuss the intriguing phenomenon of the complete cessation of consciousness, which some advanced meditators report they can voluntarily induce for extended periods of time. This state, devoid of any conscious experience, goes beyond even the state of pure awareness and has only recently attracted scientific scrutiny. I will address how this state of cessation may be accommodated within the predictive processing framework, present first data from physiological and neural assessments of this state and outline important avenues for future research. Altogether, I hope to demonstrate how the scientific study of meditation can provide unique insights into consciousness and the nature of the human mind.

MPE-1 Yizhi Li (Lightning Talk)

Ruhr-University Bochum

Institute of Philosophy II

Research Training Group “Situated Cognition”

Pure Awareness in Neo-Confucianism

Much of the debate in Chinese Neo-Confucianism from the Song-Ming era focuses on the Self-Cultivation Theory (Kung-fu Theory), specifically on how to recognize and realize the “original mind” and integrate it into secular life. Neo-Confucian philosophers of the Song-Ming period documented numerous

experiences related to the so-called “original mind,” characterizing it as a Confucian version of the experience of pure awareness (Chen, 2015; Metzinger, 2024; Stace, 1960).

However, within Neo-Confucianism, there are two main streams—the School of Principle (理学) and the School of Mind (心学)—that have very different understandings of how we can introspectively access pure awareness. Philosophers inclined towards the School of Mind believe that pure awareness or the original mind can be recognized in daily consciousness alongside other experiential contents. In contrast, philosophers of the School of Principle believe that pure awareness can only occur as a full-absorption episode (Metzinger, 2020), meaning it cannot coexist with other conscious content.

In this lightning talk, I will first introduce the Neo-Confucian characterization of pure awareness, then discuss the differences between the School of Principle and the School of Mind on pure awareness, and lastly, briefly discuss the implications and practical ramifications of their differing views on this issue.

MPE-2 David Higgins

MPE Project

Tsadra Foundation, New York

The nature and characteristics of pure awareness (rig pa) according to the Great Perfection (Dzogchen) Tradition of Tibetan Buddhism

Between the 8th and 14th centuries, a succession of scholar-practitioners in the syncretistic Tibetan Buddhist tradition known as Great Perfection (Dzogchen) articulated and defended a distinction between pure awareness (rig pa) and dualistic mind (sems) mind that was considered indispensable to understanding its complex systems of thought and practice. By helping the adept differentiate unconditioned mental states characterized by minimal conceptual and affective distortion from conditioned dualistic states, the distinction was used to clarify the conditions of delusion and enlightenment as well as the disclosive path leading from the former to the latter. This talk traces the development of this distinction within its historical and doctrinal contexts, with a particular focus on the meaning, nature, and characteristics of pure awareness as contrasted with those of dualistic mind. It concludes with a brief assessment of the contemporary relevance of the distinction, suggesting some implications its normative-ethical orientation could have for investigations of consciousness going forward.

MPE-3 Ema Demšar

Monash University
Department of Philosophy
Monash Centre for Consciousness and Contemplative Studies

Co-authors: Ema Demšar, Ruben E. Laukkonen, Heleen A. Slagter, Claire Petitmengin, Thomas Andrillon, Jennifer M. Windt

Using micro-phenomenology for the study of minimal phenomenal experience: Examples from investigating mind blanking and cessation of consciousness in meditation

One of the challenges encountered in studying minimal phenomenal experience (MPE) is acquiring detailed phenomenological descriptions of MPE-related processes and states. In this talk, we focus on phenomenological interviews as one of the approaches that could help address this challenge. We present preliminary findings from applying micro-phenomenological interviews (Petitmengin, 2006) to investigate two types of experiences that are plausibly related to MPE but haven't yet been systematically phenomenologically explored: mind blanking, in which individuals report momentary absence of awareness of conscious experience or of contents of conscious experience in wakefulness (Andrillon et al., 2019), and the cessation of consciousness in meditation (Nirodha Samapatti; Laukkonen et al., 2023). Study 1 investigated moments of spontaneous, involuntary mind blanking reported by healthy and ADHD individuals during the Sustained Attention to Response Task. Study 2 investigated the experience immediately preceding and immediately following volitionally induced suspension of conscious experience by an expert meditator. We present research designs and preliminary findings from both studies, discuss the implications of our research for understanding the phenomena of mind blanking and cessation, and consider the relation of both phenomena to MPE. We conclude by discussing the challenges and prospects of applying micro-phenomenology to the study of MPE-related experiences.

MPE-4 Yair Dor-Ziderman

University of Haifa
Integrated Brain and Behavior Research Center (IBBRC)
Contemplative science and neurophenomenology lab

Holding on to existence: self-models, mortality denial, and the phenomenology of meditative self-dissolution

Self-models have evolved in tandem with human development, increasing its adaptive strength by extending its predictive horizons. However, this predictive capacity can also generate existential terror as it lays bare the reality of death, whose characteristics of imminent and certain non-existence coupled with

temporal vagueness run counter to the prime evolutionary directive of securing the organism's, and by proxy, its self-models', existence. Summarizing a series of studies, we introduce a neurophysiological visual mismatch-response paradigm indexing a prediction-based neural mechanism associating death-related stimuli to the 'non-self/other/world' part of the experiential field (termed death-denial). We then provide evidence that this sub-personal defense mechanism is absent in meditators, but not in ayahuasca veterans. On the other hand, we show that higher level (self-report/behavioral) mortality defenses are equally lower in meditator and ayahuasca samples relative to controls. Finally, we present neurophenomenological evidence linking the valence of meditative self-world boundary dissolution experiences—a training of self-models in non-existence—to computational equanimity regarding the coupling of death- and self-related stimuli (termed death-acceptance). We argue that as the training path to pure awareness passes through the down-regulation to cessation of minimal self-modelling, preliminary practices training death acceptance may be crucial for inculcating a predictive horizon of non-existence.

MPE-5 Toby Woods

Monash University
Monash Centre for Consciousness and Contemplative Studies (M3CS)

Do Nothing meditation as a potential shortcut to minimal phenomenal experience

In classical accounts of traditional forms of meditation said to lead to minimal phenomenal experience or a complete stillness of mind, that experience is typically presented as coming at advanced stages of meditation that often require years of intensive practice to reach. By contrast, in the modern, secular form of Do Nothing meditation developed by the late Australian psychiatrist, Ainslie Meares, it is said that meditators typically achieve a complete stillness of mind with just a few weeks or months practice of 20-40 minutes per day. In this presentation I will provide an overview of my 7 years of full-time research comparing the Do Nothing meditation with traditional practices. I will focus in particular on two recent lab-based studies which involved a total of more than 320 participants and which examined what happens in mind and body in a single session of either Do Nothing or focused attention practice. The presentation will explore what the research as a whole reveals about whether the Do Nothing technique can be regarded as a path, and perhaps even a shortcut, to minimal phenomenal experience.

MPE-6 Cyril Costines

University of Freiburg
Department of Psychosomatic Medicine and Psychotherapy
Medical Center

Inducing MPE Using Floatation-REST: A Phenomenological Pilot Study

The investigation of altered states of consciousness induced by consciousness-modifying techniques is crucial for advancing our understanding of consciousness, with minimal phenomenal experience (MPE) during full-absorption episodes in meditation being particularly significant. However, inducing stand-alone MPE in laboratory settings has proven challenging, typically requiring highly skilled meditators. This pilot study addresses this limitation by evaluating the potential of the floatation-restricted environmental stimulation technique (floatation-REST) to induce MPE in a broader participant pool, including non-meditators. Participants with extensive meditation experience were subjected to two conditions: floatation-REST alone and floatation-REST combined with nondual (ND) meditation. Post-intervention assessments employed semi-structured interviews, guided by micro-phenomenological methods, and thematic analysis based on MPE-M92 questionnaire factors. Results indicate that floatation-REST can induce MPE phenomenology, with greater depth and consistency when combined with ND meditation. Participants reported dissolution of body boundaries, altered sense of time and space, and profound reduction or elimination of self-consciousness, all indicative of MPE. This study suggests that floatation-REST may serve as a reliable method for inducing MPE, offering a potential alternative to traditional meditation practices. These findings lay the groundwork for future research using micro-phenomenological interviews, psychometric questionnaires, and neurophysiological measurements to further contrast ND meditation-induced versus floatation-REST-induced experiences.

MPE-7 Christopher Timmermann

Imperial College London
Brain Sciences
DMT Research Group

Preliminary findings of the utility of employing 5-MeO-DMT to study MPEs

Anecdotal reports of 5-MeO-DMT – a powerful short-acting psychedelic – suggest this compound can induce experiences that can eliminate the sense of self, thoughts and sensory contents while preserving wakefulness, resembling minimal phenomenal experiences (MPEs). In this neurophenomenological study of 5-MeO-DMT ingestion in naturalistic contexts, measures of brain activity (measured with EEG) and micro-phenomenological interviews were

combined to assess the utility of 5-MeO-DMT as a tool to study MPEs. Results revealed that the 5-MeO-DMT experience follows a temporal pattern of enhanced emotions and bodily sensations, followed by a progressive elimination of contents of experience, which on some occasions resulted in a ‘void’ experience resembling an MPE. EEG findings resulted in a global reduction of alpha power for all participants. However, beta power reductions appeared to be specific to individuals reporting stronger MPE-like phenomena, and this biological marker was absent from individuals reporting lower MPE-like phenomena. These preliminary findings suggest that 5-MeO-DMT is a reliable means to study MPE. However, detailed forms of phenomenological interviewing are required to adequately characterise this state and its neural correlates, as not all participants achieve it. Future controlled experiments combining real-time experience sampling of phenomenologically informed participants are required to adequately study pharmacologically induced MPEs.

MPE-8 Ronald Sladky

University of Vienna
Faculty of Psychology
Department of Cognition, Emotion, and
Methods in Psychology
SCAN-Unit

Of hidden springs and endless oceans

In the wake of the active inference framework two popular theories of consciousness highlight the relevance of insular cortex for interoceptive self modeling [Seth-2021] and subcortical brain regions for qualitative experience [Solms-2021]. Both provide a compelling ecological argument for integrated conscious experience, i.e., self organization of complex organisms with different parallel and hard to reconcile optimization goals. Computationally, both theories require some kind of self model as basis, which could contradict credible reports of MPE [Metzinger-2024].

However, duality of conscious experience could be explained by a neuroscientific theory of two distinct brain networks emerging from different evolutionary history, leading to marked differences in cytoarchitecture and function [Sanides-1962, Luu-Tucker-Friston-2024]. Speculatively, System-I, originating from an olfactory system and amygdala-centered expansion gradient towards ventro/medial cortex, could enable interoceptive self modeling for everyday, habitual interactions with the body and the world. System-II, hippocampus-centered towards dorso/lateral cortex, could enable less egocentric forms of cognition and experience. In this sense, MPE could be a less salient/habitual form of experience, when neither interoceptive/ exteroceptive prediction errors nor spontaneous episodic memory reactivation provide self-referential engagement of System-I. System-II, if not operating on content, could give rise to self-less experience of the world as such.

MPE-9 Anand Prasad

Indian Institute of Technology, Kanpur, India
Cognitive Science

Minimal Phenomenal Experience (MPE) and the Nature of Subjectivity and Time

In the context of minimal phenomenal experience (MPE), the prevailing assumption is that subjectivity is entirely absent in pure awareness. This conclusion is based on the dissolution of specific properties of subjectivity, such as the first-person perspective and self-localization in space. However, while these properties are integral to subjectivity, their absence does not negate the existence of subjectivity itself. Some individuals report experiencing a bare witness or a sense of presence that might be a default property of consciousness, with other properties (FPP) being content-induced. Similarly, MPE is often considered timeless due to the lack of change (zero content). We propose that the very persistence of awareness—being aware of itself as the only content—could serve as a rudimentary marker for the passage of time. Imagine an opera singer holding a note: while there's no pitch change, the experience of the sustained note creates persistence of same experience and duration. Likewise, the persistence of awareness in MPE might provide a minimal sense of time.

MPE-10 Romy Beauté

University of Sussex
School of Engineering and Informatics
Sussex Center for Consciousness Science

Catriona Osborn Moar

University of Sussex
School of Engineering and Informatics
Sussex Centre for Consciousness Science

Complexity measures during brain dying stages : epistemological and methodological discussion on the study of the dying process and associated experiences

We investigated the complexity of spontaneous EEG data during the brain-dying process, following the study by Xu et al. (2023). We considered two separate complexity measures: Lempel-Ziv complexity (LZc), based on the compressibility of data (Schartner et al., 2015), and Statistical complexity (Sc), which calculates the diversity of statistical interactions. In this talk, we will first expose the limitations of the LZc measure, especially in a context where the brain becomes progressively less stationary through the stages of the dying process. We will show how alternative measures of complexity may be more appropriate in such contexts. Secondly, we will raise questions about whether

measures computed from brain activity (eg complexity measures) could inform us of the presence of any ongoing subjective experiences, such as Near-Death Experiences (NDE) or Minimal Phenomenological Experiences (MPE). Finally, we aim to open a discussion on a possible analogical pure-awareness experience as a primary, last conscious experience during the dying process, in parallel with the hypothesis of pure-awareness experience as a primary, first conscious experience in the womb (Metzinger, 2024). This discussion will invite epistemological reflection on the definition of the dying process and methodological questions on how to test this hypothesis from EEG data.

MPE-11 Edmundo Lopez-Sola

Neuroelectrics Barcelona

Understanding pure(r) experience: algorithmic and neurobiological perspectives

The aim of this talk is to provide new theoretical and computational perspectives on the study of Minimal Phenomenal Experience (MPE). We propose that a deeper understanding of the process of de-fabrication of experience (leading to “purer” forms of experience) is essential for the scientific study of consciousness and ultimately for reducing suffering in sentient beings. We use as a starting point an algorithmic agent framework rooted in the Kolmogorov Theory of consciousness (Ruffini 2017), which posits that agents generate compressive models of the world to navigate and predict their environment. This theory encompasses relevant ideas for our topic, such as the structure of experience, valence, and model confidence. This framework, which can be viewed as a generalization of active inference, is applicable to studying both biological and artificial systems and serves as a foundation for understanding de-fabrication in contexts such as meditative practice or psychedelics.

A key aspect of the proposed framework is that the disruption of the comparison between model predictions and input data is closely linked to the process of de-fabrication. We will propose a neurobiological implementation of this mechanism using a neural mass model of physiological activity (Sanchez-Todo 2023) aligned with predictive processing mechanisms.

MPE-12 Ishan Singhal

IIT Kanpur
CDIS

Minimal phenomenal experiences are not timeless

Temporal properties of experience have been the cornerstone of phenomenology. It is argued that all experiences irrespective of modality, content, or state, possess temporal properties. Therefore, in phenomenology, time is considered as one of the fundamental organizing properties of

experience. In deviating from this, exceptional experiences in deep meditative states, psychedelics, and states of flow are argued to be ‘timeless’. This talk will contend that these claims stem from distortions of only two dimensions of phenomenal time (namely flow and duration). The distortions or absence of these properties are not enough to claim that experience can occur without any reference to a dimension of time. In this talk, I will propose that these experiences are not timeless in virtue of them having temporal properties of persistence and endurance. To support these claims, we will consider first-person reports, computational models, a related temporal illusion, and a thought experiment. It will be argued that to register an experience as occurring, it minimally requires a phenomenal property of persistence.

MPE-13 Lars Sandved-Smith

Monash University

Monash Center for Consciousness and Contemplative Studies

A deep computational phenomenology of MPE as metacognitive free energy optimisation

This short talk presents a proposed computational model of the inferential processes that may be at play during momentary and ongoing experiences of so-called minimal phenomenal experience (MPE). This model assumes that phenomenological dynamics can be formalised by the tools of Bayesian mechanics and the free energy principle, specifically relating experience to belief dynamics. The architecture employed evinces a hierarchical depth, whereby higher level inferences are about the parameters of lower level inferential processes. I relate MPE to the mathematically derived possibility of a free energy minimum that emerges, at the metacognitive level, as the individual develops a generative model of the parameters that define the space of beliefs.

28 TBA (Lightning Talk)

For details, please refer to the online version of this booklet.

29 Sascha Benjamin Fink

Friedrich-Alexander-Universität Erlangen-Nürnberg

Centre for Philosophy and AI Research

What structure is there in consciousness? Methodological vs. Metaphysical Structuralism

Intra-phenomenal structures play a crucial role in investigating consciousness, e.g., the degree to which turquoise differs from oxblood red, in which way a C#-

chord played on a piano resembles a D7-chord played by a string ensemble, how the flow of felt time can be characterised, or how to multi-sensory objects are formed. However, it is a matter of debate which of these structures is there in phenomenality itself and which are imposed on it methodologically. Here, I will reconstruct this problem, present three different solutions, and discuss their merits: Metaphysical Hypo-structuralism, Metaphysical Hyper-Structuralism, and Methodological Structuralism.

30 Lukas Kob

OvGU Magdeburg

Philosophy

Structuralism and consciousness: the two-factor interaction view

There is an urgent need to design more experiments that pit theories of consciousness (ToCs) against each other. In this talk, I argue that developments in the relationship between general mechanisms of consciousness and neural encodings of the structure of phenomenal experience (Fink et al. 2021) potentially provide an additional area for testing theories of consciousness. I argue that the major ToCs differ in their implications for the relationship between the NCC and phenomenal structure, and that this provides further grounds for testing ToCs. While some theories argue that explanations of phenomenal structure and general mechanisms of consciousness can be separated, others deliberately explain consciousness and phenomenal structure through a single account (Marvan and Polák 2020). After outlining the one-factor and two-factor views, I propose a two-factor interaction view as a metatheoretical framework that covers a middle ground between these extremes. While the mechanisms underlying phenomenal structure are likely to be distinct from general mechanisms of consciousness, there is evidence that general mechanisms of consciousness modulate content structure. I propose experimental designs to further investigate this issue with analyses of population code structures, such as representational similarity analysis, and outline the implications for the major ToCs.

31 Daniel Weger

Goethe University Frankfurt

Structuralist representationalism

The hard problem of consciousness is the problem of explaining how and why physical states and processes give rise to conscious experience. Representationalism claims that the phenomenal character of an experience can be explained in terms of its representational content. However, tracking representationalism, which combines the representationalist idea with the claim that representation is a matter of tracking relations, faces several substantial objections. The phenomenal intentionality theory, which turns the

representationalist idea upside down and seeks to explain intentionality in terms of phenomenal consciousness, is also flawed because it does not even attempt to address the hard problem of consciousness. Instead of abandoning representationalism altogether, I propose to combine the representationalist idea with the structuralist view of representation, which holds that representation is based on an exploitable structural correspondence. According to structuralist representationalism, phenomenally conscious representation involves an exploitable structural correspondence between a neural activation space and a quality space. It thus claims that the phenomenal character of experience is one and the same as the representation of the quality to which the underlying neural state structurally corresponds. This view provides an account of phenomenal consciousness in physicalistically respectable terms while avoiding the shortcomings of tracking representationalism.

32 Tim Ludwig

Karlsruhe Institute of Technology

Institute of Technology Futures, Department of Philosophy

Interplay between concepts, structure, and empirical practice: an inspiration from physics

The emerging field of mathematical consciousness science is still in its early stages with many competing views and approaches. And the specific challenges of consciousness science, which originate from the special role played by consciousness, are clearly distinct from any other science. Nevertheless, it can be fruitful to ask what we can learn from other sciences and their development.

Considering physics as an established science, we focus on Newtonian mechanics, which marks an important step in the historical development of physics. More explicitly, we present a methodological reconstruction of Newtonian mechanics and, on that basis, discuss the interplay between concepts, structure, and empirical practice. In particular, we show how the interplay between Newton's concepts (specified in definitions) and structure (specified in laws) can be used to design an empirical practice.

Taking the example of Newtonian mechanics as inspiration, we finally present a few ideas for how the interplay between concepts, structure, and empirical practice could be made use of in mathematical consciousness science.

33 Anna Eiserbeck

Humboldt-Universität zu Berlin

Department of Psychology

Needed: Scientific investigations of the effects of consciousness assumptions

The question of whether artificial intelligence (AI) can achieve consciousness or

already possesses it is currently a topic of intense debate among both experts and laypeople. Opinions on this matter vary widely, reflecting the complexity of the issue. Beyond the question of AI consciousness itself, it is crucial to investigate the implications of such beliefs or claims. This is particularly relevant as the distinction between human and non-human, real and AI-generated, is becoming increasingly blurred in contexts such as virtual reality, interactions with robots and large language models, and media that can present both authentic and AI-generated content. In this talk, I will present experimental studies demonstrating how assumptions or information about the authenticity or animacy of an agent, despite identical external behaviors, influence the processing and evaluation of emotional information. The findings underscore the importance of subjective assumptions about the mind behind identical behaviors and highlight the necessity of examining the effects of beliefs about consciousness, which represent an essential basis for ethical considerations.

34 Rony Hirschhorn

Tel Aviv University

Sagol School of Neuroscience

More than words - can free reports adequately measure the richness of perception?

How rich is our conscious experience? We feel that we experience the world in great detail, yet whether this is indeed the case is widely debated. While some researchers hypothesize that conscious experience is rich and detailed, others claim that experience is coarse and sparse. Despite the increasing diversity of paradigms in experimental cognitive psychology, this question remains unanswered. Recent works attempting to study spaces and structures to describe phenomenology in experimental research have begun to use freely reported words as means of experimental report. I challenge the interpretation that freely reported words can be used to represent observers' visual experiences, or to prove that conscious experience is highly rich and detailed. Through a series of experiments using degraded and intact versions of images used in a previously published work (Chuyin et al., 2022), I show that the intact and degraded stimuli yielded experiences that differ in their perceived richness; however, both intact and degraded experiences were described in highly-specific words. Across six experiments, I found a consistent relationship between words' specificity to the image and their general frequency of use in the English language.

35 Matthias Borgstede

University of Bamberg

A theory of psychological variables and their measurability based on relational responding

Empirical findings suggest that relational responding is operant behavior and, as such, selected by its environmental consequences. Following the formal requirements of representational measurement, I propose that a measurable psychological variable emerges, if and only if relational responses are selected such that the resulting structure is isomorphic to a subset of the real numbers. The theory of operant selection further implies that selection is a multilevel process that shapes behavior with regard to its expected fitness consequences. Since fitness itself has a quantitative structure, quantitative psychological variables should emerge if and only if the environmental contingencies allow for a reliable prediction of expected amount of fitness gain. Consequently, measurable psychological variables are context-dependent fitness predictors that may function as discriminative stimuli for expected fitness gains given a certain action is taken in that context. The theory clarifies what psychological variables are, why they are sometimes quantitative and sometimes not, how they can be measured and why we should expect quantity in the realm of psychology in the first place. Moreover, the theory is empirically rich, yielding several unique hypotheses that can be tested experimentally.

36 Gaspard Fougea

ENS Paris-Saclay
Computer sciences
LMF

Finite automata as a tool for modeling psychological phenomena

Using computational frameworks to model cognition has been standard since the cognitive revolution in the 1950s. In our research, we provide the use of a theoretical computer science tool which has the advantage of being simple and accessible for the modeling of psychological phenomena: Finite Automata. Our research leverages finite automata's capacity to represent complex theories in a simple and clear way, translating established psychological and consciousness theories into finite-state models. Some advantages of this methodology are:

1. Educational utility via the clarity and simplicity of finite automata diagrams
2. Transdisciplinarity
3. Mathematical formalization of theories which facilitates the identification of ambiguities within theories
4. Comparative analysis of theories via the comparison of automata

As a proof of concept of this approach, we present an enriched finite automata model of Lazarus and Folkman's Transactional Theory of Stress (Lazarus &

Folkman, 1984), focusing on key elements such as cognitive appraisal and coping. While an exhaustive model of the entire theory is beyond the scope of this study, we attempted to model some of its essential components. Lazarus, R. S., and Folkman, S. (1984). *Stress, Appraisal, and Coping*. New York: Springer.

37 Daniel Müller

University of Osnabrück
Institute of Philosophy

Two Paradigms of The Unconscious - And What They Might Tell Us About Consciousness

Much effort has been put into understanding consciousness, however, we might have to take a different angle and direct our attention to the unconscious as the complementary concept which helps us delineate and better understand consciousness.

Two paradigms are currently fashionable in the philosophy of the unconscious: the first is the notion of the unconscious of Neo-Freudian scholars which holds that repressed desires drive many of our actions, navigating us from our unconscious with us only in charge of rationalising our actions after the fact. The second is one notion from dual-process theories of cognition so-called type 1 thinking is automatic, quick, unconscious cognition which in turn is error-prone and leads to biases.

Using the intersection of these different paradigms as a minimal theory of the unconscious might give us some novel insights fruitful for the framework of the scientific study of consciousness. In particular, they might sharpen the concept of consciousness and give us a better grip on what sets apart conscious and unconscious processing and might inform us on the preconscious in between. Particularly since the contrastive analysis is a frequently employed paradigm in the (neuro-)science of consciousness, an investigation into what constitutes the unconscious might be helpful.

38 Ute Schmid (Keynote)

University of Bamberg
Cognitive Systems Group
Bamberg Center for Artificial Intelligence (BaCAI)

Can AI Systems Simulate or Have Consciousness?

Whether it is possible for artificial, digital systems to have consciousness, is an ongoing debate since the beginning of AI research. It has been discussed for symbolic AI systems, artificial neural networks, hybrid cognitive architecture, as well as most recently in the context of generative AI, such as Google's LaMBDA chatbot. In the talk, I will present core aspects of this debate, starting with the physical symbol hypothesis of Newell and Simon. I will discuss the

possibility of testing the artificial consciousness hypothesis with respect to the Turing test and to psychometric AI. Furthermore, Artificial General Intelligence (AGI) will be introduced as a line of AI research which contrasts to standard narrow approaches designed for specific tasks.

39 Christoph Benzmüller (Keynote)

University of Bamberg
Information Systems and Applied Computer Sciences
Chair for AI Systems Engineering (AISE)
Free University of Berlin

AI Consciousness — Notes from a Logician's Perspective

Starting with a short description of my basic research interests, which lie between computer science, AI, and metaphysical and mathematical foundations, I will focus my talk on the following question: To what extent can logical formalisms step outside themselves and, e.g., reflect on their consistency? I will show that within the framework of LogiKEy, a metatheoretical, logico-pluralistic knowledge representation framework that I co-developed, logical self-reflection is indeed possible to a certain extent with relatively simple means. The approach I present uses shallow semantic embeddings of object logics in a sufficiently expressive meta-logic (the latter being by default classical higher-order logic HOL), and, as I will show, it supports nested embeddings of HOL in HOL itself in a fractal-like way, so that consistency claims about the inner logics can be made and proved at the outer layers.

40 Keenan Down

University of Cambridge & Queen Mary University of London
Department of Psychology
Consciousness and Cognition Lab

The Common Thread: Simplifying the study of consciousness with shared experience, shared identity and shared information

Many problems and thought experiments in the philosophy of mind depend on the relationship between shared experience and shared identity. Are the experiences of two people ever comparable? If two people could have inverted spectra, why couldn't I experience inverted spectra each time I look at the same apple? If I am cloned or killed and revived, am I the same individual? Does shared experience imply a shared identity? In an effort to streamline the search for viable theories of consciousness, we use the notions of shared experience and shared identity to construct two properties that a theory of

consciousness might be required to hold. We argue that these properties are intrinsically linked, allowing us to dismiss philosophical stances which hold contradictory views on these properties. We discuss the implications of these two properties, their philosophical burdens, and how they might be actualised to the neuroscience of consciousness using information theory.

41 Eric Elmoznino

Mila - Quebec AI Institute

Why can't we describe our conscious experiences? An information theoretic attractor dynamics perspective of ineffability

Conscious states seem both rich or full of detail and ineffable or hard to fully describe or recall. The problem of ineffability, in particular, is a longstanding issue in philosophy that partly motivates the explanatory gap: the belief that consciousness cannot be reduced to underlying physical processes. Here, I will provide an information theoretic dynamical systems perspective on the richness and ineffability of consciousness. Under this framework, the richness of conscious experience corresponds to the amount of information in a conscious state and ineffability corresponds to the amount of information lost at different stages of processing. I will describe how attractor dynamics in working memory would induce impoverished recollections of our original experiences, how the discrete symbolic nature of language is insufficient for describing the rich and high-dimensional structure of experiences, and how similarity in the cognitive function of two individuals relates to improved communicability of their experiences to each other. While the model may not settle all questions relating to the explanatory gap, it makes progress toward a fully physicalist explanation of the richness and ineffability of conscious experience—two important aspects that seem to be part of what makes qualitative character so puzzling.

42 Michael Timothy Bennett

Australian National University
College of Engineering, Computing and Cybernetics
School of Computing

Why Nature Does Not Like Zombies - Unifying Lower and Higher Order Theories of Consciousness: A Computational Approach

Why is there consciousness? To answer we construct a formalism unifying lower and higher order theories. We assume natural-selection learns organisms that learn cause. Cause is learned by constructing object X such that X causes Y, not by presupposing the world is divided into objects including X and then testing to see if X causes Y. If one cannot presuppose objects, then one must construct objects by being attracted to or repulsed from contentless quality (affect). Learning reduces quality to policies. Policies classify causes of affect, and so have quality. With scale, progressively higher orders of 'causal-identity'

are constructed. These facilitate refference, then self-awareness and access-consciousness. Refference lets an organism discern the consequences of its actions. A policy called a “1st-order-self” performs this function. It has a quality that accompanies everything an organism does. That quality is what it is like to be the organism (phenomenal-consciousness). An unconscious zombie would be less able to learn cause and adapt. Likewise, survival may demand organism-X infer organism-Y’s prediction of X’s interventions (“2nd-order-selves”). Access-conscious contents are communicable in the Gricean sense, so they must be of 2nd-order-selves. This ‘dissolves’ the hard-problem by going a level down to the fundamental drive: stay alive!

43 Paul Skokowski

Stanford University
Symbolic Systems and Philosophy
Center for the Explanation of Consciousness
Oxford University
Philosophy Faculty

Androids then and now

Elsewhere I have argued for a position called sensory naturalism, which takes seriously the approach that the senses are physical detectors of properties in the environment: qualia. This approach recognizes that sensory experiences include physical properties in the world external to the brain, and so mental states, should be seen as extending beyond the boundaries of neuroscience and into our external physical environment. If we now include androids/AIs as possible creatures with mental states, then mental states would need to be more broadly construed, since such creatures presumably are made of non-biological components. And if such creatures learn, form memories and experience (sensorially) the world by virtue of (and as a causal result of) the humans who design them, then androids requires a physical naturalism beyond the biological naturalism of humans and other life that has previously evolved on Earth. This talk will examine the possibility of a physical naturalism that makes room for the mental states of androids/AIs, including their experiencing qualia. We will start with a historical perspective from Descartes, which I suggest has some support from our everyday beliefs about androids, and then use sensory naturalism as a roadmap towards determining whether such creatures might experience qualia.

44 John Barnden

University of Birmingham
School of Computer Science

Machine Consciousness, Meta-Causation, and Mimicry

The recent MetaDyn theory claims that pre-reflective phenomenal consciousness is identical to suitably patterned meta-causation operating at a basic physical level. Meta-causation is causation that acts partly or wholly on causal instances themselves. Thus, if we consider an AI program implemented on a present-day computer, and assume that computer circuitry doesn’t involve (suitably patterned) meta-causation, the program’s running cannot involve consciousness, no matter how algorithmically sophisticated it is. And even if the implementing physical machine were suitably meta-causal, the consciousness thereby arising might well not be the experiencing of the program’s cognitive processing (reasoning, etc.). I will explore (1) what knitting-together of the implementation-level meta-causation and the computation-level processing might ensure such experiencing. Also:- (2) The physical operations that, according to MetaDyn, provide conscious processing of some entity C (organism or artefact) could in principle be numerically simulated by a computer program. If this program were implemented on a conventional computer, there would be no consciousness arising. I will consider philosophical and scientific, including evolutionary, issues raised by this. If, by contrast, the simulating program were run on a meta-causal machine as above, then consciousness might arise, but arguably it would plausibly not consist of experiencing C’s conscious cognitive operations.

45 Nicolas Kuske

Université de Toulouse
Artificial and Natural Intelligence Toulouse Institute
CerCo, CNRS UMR5549

Bridging Consciousness and AI: Evaluating the Global Latent Workspace in Embodied Agents

In the wake of the success of attention-based transformer networks, the discussion about consciousness in artificial systems has intensified. The global neuronal workspace theory (GNWT) models consciousness computationally, suggesting the brain has specialized modules connected by long-distance neuronal links. Depending on context, inputs, and tasks, content from one module is broadcasted to others, forming the global neuronal workspace representing conscious awareness.

The global latent workspace (GLW) model introduces a central latent representation around which multiple modules are built. A semi-supervised training process ensures cycle consistency, enabling content translation between modules with minimal loss. The central representation integrates

necessary information from each module, with access determined by transformer-like attention mechanisms.

We examine the dynamics of a virtual embodied reinforcement learning agent with a minimal GLW setup, involving deep visual sensory and motor modules. The augmented PPO agent exhibits complex goal-directed behavior in the Obstacle Tower Challenge 3D environment. Latent space representations cluster into sensorimotor affordance groups.

This study links GNWT with sensorimotor contingency theory, suggesting that changes in sensory input relative to motor output constitute the neuronal correlates of conscious experience. This convergence in a machine learning setup raises the question: Can such in silico representation suffice for phenomenal spatial perception?

46 Manuel Blum (Keynote)

Professor Emeritus at U.C. Berkeley
 Professor Emeritus at CMU
 Chair Professor at Peking University
 Dept of Computer Science at CMU

Insights on Consciousness from Theoretical Computer Science

This talk will do the following:

1. Describe the Conscious Turing Machine (CTM), a model of consciousness inspired by Baars' Theater Model, and formalized from the viewpoint of Theoretical Computer Science.
2. Explain how we check the correctness of the CTM as a model of consciousness, and
3. Give some insights from the CTM on the hard problem of pain.

This describes work of Lenore and Manuel Blum. Manuel will give the talk.

For details, see:

- M. Blum & L. Blum, (2021, March). A Theoretical Computer Science Perspective on Consciousness. JAIC, 8(1), 1-42. <https://bit.ly/3sUqC7d>
- L. Blum & M. Blum, (2022, May) A theory of consciousness from a theoretical computer science perspective: Insights from the Conscious Turing Machine. PNAS, 119(21). <https://bit.ly/38zAhf6>
- L. Blum & M. Blum, (2024) AI Consciousness is Inevitable: A Theoretical Computer Science Perspective. <https://arxiv.org/pdf/2403.17101>

47 Leonard Dung

Ruhr-University Bochum
 Institute of Philosophy II

Consciousness and Computational Implementation

According to computationalism about consciousness, implementing the right kinds of computation is sufficient for phenomenal consciousness. However, based on recent work – particularly by Derek Shiller – I explicate implementationalism about consciousness. Implementationalists believe that there are substantive constraints on implementing computations, rather than merely simulating them, that conventional, silicon-based artificial systems do not satisfy. So, implementationalism rules out consciousness in conventional AI. I first develop a dilemma for implementationalism: Implementationalism either conflicts with all independently plausible accounts of computational implementation or turns out to be ad-hoc, unsupported, and inferior to widely known non-computationalist alternative accounts. I also suggest that the mechanistic account of computation can be used to derive plausible constraints on computational implementation which support the possibility of consciousness in conventional AI.

48 TBA (Lightning Talk)

For details, please refer to the online version of this booklet.

49 Lenore Blum

Carnegie Mellon University
 School of Computer Science
 Center for AI Consciousness

How rCTM, a robot with a CTM brain, answers Kevin Mitchell's questions about consciousness

In Wiring the Brain (Mitchell, 2023), Kevin Mitchell presents “a non-exhaustive list of questions ... that a theory of consciousness should be able to encompass”. He declares that “even if such a theory can't currently answer all those questions, it should at least provide an overarching framework (i.e., what a theory really should be) in which they can be asked in a coherent way, without one question destabilizing what we think we know about the answer to another one.” Mitchell's questions are thoughtful, interesting, and important. We offer preliminary answers from the perspective of the Conscious Turing Machine Robot (rCTM), a simple formal machine model of consciousness. This is joint work with Manuel Blum. For background on CTM and rCTM, see (Blum & Blum, 2022) and (Blum & Blum, 2024).

50 Albert Newen

Ruhr-Universität Bochum
Institut für Philosophie II
Center for Mind and Cognition

The ALARM Theory of Phenomenal Consciousness: AI systems can only become conscious if they realize structures of basic biological systems

Phenomenal consciousness is still one of the biggest riddles. The aim is to characterize the core functional role of consciousness for biological systems and then to discuss what this means for AI systems. We start with outlining a two-level theory of consciousness to account for the relation between evolutionary old and more recently developed types of consciousness. The conceptual suggestion is that we should distinguish two levels of consciousness, namely bodily arousal and general alertness. This conceptual distinction is proven to be adequate and epistemically fruitful for three reasons which will be spelled out. The distinction is supported by 1. the evolutionary functional roles of phenomenal consciousness, 2. neuroscientific studies of the role of deep cortical and thalamic brain activities for consciousness as described by Halassa (Nakajima et al. 2019) and 3. systematic recent observations about animal consciousness. Furthermore, we argue that due to the functional roles of consciousness and their specific biological anchoring that AI systems with the cognitive architecture they is nowadays available cannot be expected to become conscious. Of course, we can produce a simulation of the relevant functional roles but with a simulation of the functional roles we do not receive the core phenomenon of consciousness.

51 Jonathan Mason

Association for Mathematical Consciousness Science

Additivity of measures of consciousness on sigma-algebras of independent systems: An example from Expected Float Entropy Minimisation

In the field of consciousness science, various formal measures of consciousness have been proposed over the last 20 years. Most notably, Integrated Information Theory (IIT) proposes that structure integrated information (big Phi) corresponds to the quantity of consciousness of a system. In this talk we consider when formal measures of consciousness need to abide by the mathematical definition of a measure, as formally defined in Measure Theory. In principle, formal measures of consciousness ought to be able to measure the consciousness of any proposed finite system, including systems with many independent subsystems such as a nest of ants, if the ants are indeed independent. For any finite system, one can generate a sigma-algebra

from its set of independent subsystems; take the power set. Over this sigma-algebra formal measures of consciousness need to be sigma-additive, non-negative and return a measure of zero on the empty system. As an example, we show that it is straight forward to obtain such a measure from the function μ used in the definition of model unity in the Expected Float Entropy Minimisation (EFE) model of consciousness. To what extent the measure may qualify as a measure of consciousness is as yet unknown.

52 Etienne Jacques (Lightning Talk)

Independent

Formal Metaphysics and the Subjective Construction of Objective Reality

We present a novel approach to the longstanding metaphysical question of what constitutes reality. By assuming the existence of subjective experiences as a foundational starting point, we derive a well-defined concept of Objects while avoiding the difficulties of ontological grounding. Specifically, we propose to define Objects as deriving from collective subjective experiences. Within our formal model, we establish a hierarchy of levels of abstraction, where any third-order or higher structure corresponds to an Object. As an immediate application of this framework, we classify natural objects, both concrete and abstract, in a manner reminiscent of 19th-century Naturphilosophie.

To further illustrate the utility of our approach, we show how it can be used to implement a form of ontological relativism. This leads to potential experimental predictions and practical applications in fields such as artificial consciousness and artificial reality.

Our approach suggests the existence of numerous subjective experiences beyond those we directly encounter. Among these, some are formally similar to our human experience of Objective Reality, indicating a comparable quality of Consciousness.

53 Michael Schnabel

Vanderbilt University
Department of Political Science

Can Buddhist Theory and Practice inform contemporary Models of Consciousness?

The growing field of mathematical models of consciousness could benefit from considering already existing models that derive from Buddhist traditions. The Abhidharma and the Yogācāra school understand consciousness as a dynamic process unfolding within the Mind-ground (tathāgatagarbha) according to a causal model described by the twelve links of dependent origination (pratītyasamutpāda).

The Chan school (precursor of Zen) absorbed these strains and integrated them into a contemplative practice that is aimed at “realizing how Mind works.” The Chan/Zen tradition also emphasizes that conceptual models of any kind, linguistic or mathematical, while providing useful signposts to guide the practice towards realization, may ultimately not be necessary and sometimes even obstruct it.

As we formulate models of consciousness in the 21st century, these ancient studies are both informative and cautionary. If realization truly “does not depend on words and letters”, as the Chan/Zen tradition asserts, how do we develop models of consciousness that are useful without neglecting context and limitations? I will discuss some key features of Buddhist models and how they bear on current developments in this field. I will then look towards how these aspects might inform assumptions on conscious agency, time, and our attitude towards model-building itself.

54 Sylvia Liu

Ludwig Maximilian University of Munich (LMU)
Faculty of Philosophy, Philosophy of Science and Religious Studies

Non-subjective' subjectivity? Non-self and Chinese Huayan Buddhism

Is it possible to lose one's sense of subjectivity? If so, are they still a person? Can a non-subjective perspective replace our subjective phenomenal experience? These are the questions I will address in my talk. My answers are: Yes, losing one's sense of subjectivity is possible; we remain persons even when we lack a sense of our subjectivity; and our subjective phenomenal experience can be replaced by a non-subjective perspective. Furthermore, I will bring the Chinese Huayan Buddhist concept of non-self to this debate. This ancient Buddhist school advocates for the possibility of transcending our personal subjective perspective in favor of a non-centered, omniscient view.

55 Chetan Prakash

CSUSB
Mathematics
Association for Mathematical Consciousness Science

Logic of Observation and the Physical Headset in Conscious Dynamics

We propose that acts of observation are acts of participation in a conscious stochastic dynamical process, generated by acts of perception, decision and action made within a network of mutually interacting conscious agents. An observer, in this view, is an agent whose perceptual space is contained in the perceptual space of the observed - and the same holds for their action spaces. Both observer and observed participate as dynamical parts of a larger

underlying conscious dynamics. We define these notions precisely and use them to develop a logic of observation, one that comports well with previously studied logics of belief. This observational, or “trace,” logic is more general than those describing classical or quantum logics. We further suggest how to view our current physical theories, in particular particle physics, as projections of very special instances of conscious dynamics. The projection involves identifying some communicating classes of the asymptotics of conscious dynamics as “particles” and specifying which properties of communicating classes correspond to physical properties of particles. Our goal is to establish these correspondences by recognizing the projection as one from conscious dynamics to the amplitudehedron, an aspect of the positive geometries currently being actively studied in high-energy physics.

56 Robert Tromm (Lightning Talk)

University of Oxford
Centre for Eudaimonia and Human Flourishing
Linacre College

Changes in functional hierarchical organization after acute and chronic use of psychedelics

Previous work has proposed that psychedelics flatten the brain's hierarchy, resulting in increased flexibility of brain states through disintegration of the default mode network. The present work leveraged novel measures to examine changes in the functional hierarchical organization of the brain under ayahuasca, DMT, and cannabis. Participants were imaged at baseline/placebo and after administration of ayahuasca (N=24), DMT (N=17), cannabis in chronic users (N=14), or cannabis in occasional users (N=12). Entropy production across the brain was estimated through the INSIDEOUT framework. Trophic coherence was evaluated on directed networks derived from a Hopf bifurcation model, allowing for evaluation of the coherence of the global brain network and regional changes in hierarchical influence. Irreversibility was found to decrease under ayahuasca and cannabis in chronic users, but not in DMT or cannabis in occasional users. Coherence was significantly decreased under both ayahuasca and DMT, while an increase was found for cannabis in occasional users. Regional hierarchy broadly and significantly decreased during ayahuasca and cannabis in chronic users, but not for DMT and cannabis in occasional users. Our findings are consistent with the Relaxed Beliefs Under Psychedelics (REBUS) model, lending support to the theory that psychedelics work principally by relaxing the brain's hierarchy.

57 Paweł Motyka

Polish Academy of Sciences
Institute of Psychology
Virtual Reality and Psychophysiology Lab

Mapping the space of altered states of consciousness

While states of consciousness are increasingly better understood across a spectrum from unconsciousness (e.g., coma or anesthesia) to the waking state, we lack a successful approach for capturing the qualitative diversity within wakefulness. A fundamental challenge seems to lie in identifying the dimensions that can effectively organize the coarse-grained representation of the experiential landscape. The present study aimed to create an empirically grounded classification of pharmacologically induced altered states using a bottom-up approach. Participants (n = 739) assessed the subjective dissimilarity between pairs of vividly remembered experiences associated with the use of various psychoactive substances. Multidimensional scaling (MDS) was employed to model these dissimilarities as distances in a geometric space. The primary dimension, interpreted as the 'intensity of mind alteration', positioned baseline/ordinary state alongside states evoked by depressants, and stimulants at one end, while situating psychedelic- and dissociative-induced states at the other. The second dimension covered the inhibition-stimulation spectrum, ranging from states evoked by depressants and opioids to classical stimulants. The MDS was complemented by other multidimensional reduction techniques (PCA, UMAP, t-SNE), preserving subject-level information. We propose that such an approach may eventually inform our understanding of the dimensionality and topology underlying the landscape of human-accessible consciousness states.

58 Borjan Milinkovic

University of Melbourne and Paris Brain Institute (ICM)
Melbourne School of Psychological Sciences

Discovering emergent dynamical structure in wakefulness, sleep and anaesthesia

The complex structure of the human brain gives rise to emergent patterns of activity through the dynamic interactions among its various subcomponents. Recently, there has been growing interest in describing higher-order interactions beyond pairwise measures in brain data, particularly to stratify global conscious states. A promising approach is provided by Dynamical Independence (DI) - a formal information-theoretic measure quantifying the emergence of macroscopic dynamics. We applied this method to high-density EEG recordings in wakefulness, sleep and anaesthesia to quantify and characterise emergent dynamical structure.

Applied to EEG, emergent dynamical structure is discovered by identifying

coarse-grained variables as lower-dimensional spaces whose future evolution is not enhanced by knowing the past dynamics of their microscopic constituents.

Our study reveals that emergent variables across all macroscopic scales in both propofol and xenon conditions show significantly higher degree of emergence than the wake condition. Conversely, ketamine shows significantly lower emergence than the wake condition across all macroscopic scales. Crucially we show that dynamical structure emerges in wake and ketamine-induced anaesthesia—where conscious report was recovered—across all high-order scales. Altogether, we present a new method to explore the emergent dynamical structure of global states of consciousness from EEG data, including sleep/wake transitions.

59 Jakub Vohryzek

University Pompey Fabra
Information and Communication Technologies
Centre for Brain and Cognition

Classical serotonergic psychedelics' acute effects induce changes to brain's functional hierarchy

Psychedelic intervention has garnered substantial attention in recent years for its ability to induce profound reorganisation within the human brain, both in the acute and long-term phases of its effects. Understanding these aspects is important not only for a comprehensive grasp of the underlying neurobiology and its therapeutic potential, but also of the changes to neurophenomenology it elicits.

Recently, we proposed that acute and long-term psychedelic action can be formalised in the context of three distinct timescales. The “fast time” neuronal dynamics are shaped by the underlying structure that itself changes over “slower time” modulated by (in)dependent plasticity mechanisms. Finally, the known long-term effects of psychedelics can be formalised as changes to the plasticity itself (metaplasticity) at “ultraslow” time scale.

In this talk, I would like to focus on the fast timescale and formalise a concrete definition of hierarchy, that quantifies the time asymmetry in the directionality of information flow. Using this approach, I show that the time functional hierarchy decreases under the classical serotonergic psychedelics (psilocybin, LSD and DMT). Moving on, I propose a neurobiological implementation of the slow and ultraslow action of psychedelics to motivate further empirical investigations.

60 Joel Frohlich (Keynote)

University of Tübingen
Helmholtz Center Munich at the University of Tübingen
Fetal MEG Center

The first experience: searching for the emergence of consciousness in infants and fetuses

Our earliest memories often date back to the toddler years. However, this may not tell the whole story of our first experiences. The possibility of earlier experiences, i.e., infant consciousness, has strong bioethical implications. While infants were sometimes operated on without anesthesia as recently as the 1980s, the vast majority of researchers today believe that infants are at least probably conscious. Infants, newborns, and possibly even late-term fetuses may possess inner lives that simply do not leave lasting memories. This talk will examine the evidence for three possibilities: consciousness arising before, during, or after birth. At 35 weeks gestation, fetuses begins to exhibit cortical responses, measured with MEG, suggestive of predictive processing, often associated with consciousness. At birth, the dramatic shift from the womb's sedating environment to the stimulating outside world may necessitate the development of awareness, and some resting state fMRI networks associated with consciousness are in place no later than full term birth. However, an important developmental milestone is the postnatal GABA shift, which raises questions about the balance of neural inhibition/excitation and the complexity of newborn brain activity, potentially relating to the emergence of consciousness. Because any given marker of consciousness that has been validated in adults might not translate to infants, a “cluster-based” approach is likely needed. This approach, which looks for a convergence of diverse markers, avoids relying on any single marker and remains agnostic to specific theories of consciousness.

61 Nihat Ay (Keynote)

Hamburg University of Technology
Institute for Data Science Foundations
Santa Fe Institute

Anticipation and Collective Surprise

A number of promising theories of consciousness are based on information theory and, in particular, the notion of surprise. A prominent candidate of such a theory is given by the Integrated Information Theory of Consciousness. Further approaches are based on the idea of the brain as a predictive machine. The brain is able to predict or anticipate its sensory data stream, if it is not surprised by it. Surprise, however, is encoded in the brain in a distributed way. On the other hand, the unity of the state of being surprised calls for a mechanism that encodes surprise in an integrated way. In this talk, I will outline a formal description of such a mechanism.

62 Francesca Castaldo (Lightning Talk)

Neuroelectrics
Brain Modeling Department

The algorithmic agent perspective and computational neuropsychiatry: from etiology to advanced therapy in major depressive disorder

Major Depressive Disorder (MDD) is a complex and heterogeneous condition impacting millions of individuals globally. Computational neuropsychiatry can potentially deliver significant breakthroughs through mechanistic modeling of this condition. To develop a computational framework from first principles, we start from the Kolmogorov Theory of consciousness (KT), where agents interact with the world driven by an objective function that evaluates valence. In this context, emotion is defined as the tuple [world model, valence], and depression is defined as a state characterized by persistently low valence. Since in KT valence is a function of the agent's world model, depression may be seen to originate from inaccurate world models (cognitive biases), a dysfunctional objective function (anhedonia, anxiety), deficient planning (executive deficits), or unfavorable environmental conditions – potentially linking the model with depression biotypes.

Building upon existing literature, we bridge algorithmic, dynamical systems, and neurobiological concepts, emphasizing the role of plasticity in maintaining psychological health – the ability of the agent to update elements such as their world model or objective function. Finally, we examine how combining brain stimulation, psychotherapy, and plasticity-enhancing compounds like psychedelics can repair neural circuits, and how personalized computational models can optimize these therapies.

63 Giulio Ruffini

Neuroelectrics
Brain Modeling Department

Kolmogorov Theory (KT): a novel algorithmic perspective linking structured experience, dynamics, and symmetry

We introduce a novel framework to study consciousness based on algorithmic information theory (KT). Starting from the premise that “there is experience,” and in the context of the algorithmic agent, we focus on the requirements for structured experience (S) – the spatial, temporal, and conceptual organization of first-person perspective. We argue that access to efficient (compressive) world models is crucial for both individual survival (homeostasis) and survival of kind (telehomeostasis). We posit that world models form the basis for S and that the successful comparison of world data with model-generated data provides “realism” (presence) and structure to the experience of natural agents and, by extension, to artificial systems. Qualitative aspects of S reflect data

structure and are linked to algorithmic features of the model program, such as program length or structure, associated with the complexity and efficiency of the model's representation. Using Lie group theory and employing a neural network model as an agent, we show that data tracking constrains the agent's network structure to mirror the symmetries of the generative model. This interdisciplinary approach offers a new perspective for identifying neural correlates of agenthood and structured experience in natural agents, offering avenues for advancing AI and computational brain modeling.

64 Carlotta Langer

Hamburg University of Technology
Institute for Data Science Foundations

Integrated Information in the Active Inference Framework

The active inference framework provides a principled approach to modeling sentient behavior. In this framework perception and action selection are treated in a unified way. The resulting agents form an internal generative model of the relevant dynamics of the world in order to infer their future observations, their internal states and to select actions. We combine this modeling framework with the Integrated Information Theory of consciousness and are therefore able to analyze the active inference agents from the perspective of integrated information. The Integrated Information Theory aims at quantifying the level of consciousness of a system by assessing its capability to integrate information. In this talk we define a measure of integrated information for the generative model by making an additional structural assumption. Experiments with simulated agents demonstrate the impact of modeling assumptions in the active inference framework on the integrated information value. Furthermore, we evaluate other information-theoretic measures within active inference agents, including one often referred to as Morphological Computation, which quantifies the interaction between an agent and its environment.

65 Martin Biehl

Cross Compass Ltd
Cross Labs

Interpreting open dynamical systems as Bayesian reasoners and agents

Extending an abstract definition of the notion of conjugate priors (Jacobs 2018) we developed two closely related formal notions of interpretation of open dynamical systems. The first interprets systems with internal state and inputs as Bayesian reasoners (Virgo et al. 2021) and the second interprets Moore machines (which have outputs) as solutions to partially observable Markov decision problems (POMDPs) (Biehl and Virgo 2023) i.e. as particular kinds of agents.

Interpretations involve a model of the causes of the inputs of the system as well as an interpretation map which assigns probabilistic beliefs about hidden variables of the model to internal states of the system. The POMDP solving interpretation also involves a reward function encoding the goal.

These interpretations provide implementation independent sufficient conditions for given open dynamical systems to maintain beliefs about modeled hidden variables (a minimal kind of subjectivity) and pursue goals (agency). We believe this makes them (and future generalizations) a promising formal basis to start thinking about consciousness of open dynamical systems.

I will introduce the main ideas and discuss what we know about the notions of subjectivity and agency resulting from this approach.

66 Adel Chaibi (Lightning Talk)

Intel

Collective Representation and Shared Agentivity in Artificial Architectures

In psychology, it has been proposed that certain shared and collective representations possess their own existence beyond individual representations. This can occur through a communal space of concepts, memories, or language. For example, Carl Jung argued that psychic activity transcends the brain to explain phenomena such as synchronicity and extra-sensory perception. He theorized a collective unconscious to explain some observed phenomena. In his theory, the collective unconscious refers to the part of the global unconscious mind shared by all human beings, containing universal experiences, memories, and symbols inherited from our ancestral or evolutionary past. This suggests the existence of a shared field of knowledge that we cannot fully access through our sensory and conscious experience.

Our focus is on how artificial architectures underlying artificial agents can share their representations to build a common sense and collective construct. We will propose a model of agentivity for the assembly of agents that accounts for the agentivity of the parts but cannot be restricted to them. With Eric Petit and Grégoire Sergeant-Perthuis.

67 Marc Heimann

Hochschule Niederrhein
FB03

Outside the Model: How Absence (mis-)shapes Language

The presentation will approach current large language models from a perspective informed by Lacan and Heidegger. It will discuss how these models reproduce language in a manner that resembles a subdued madness, lacking subjectivity, because they have no access to the unconscious. On the one hand, the way these models reproduce language can be easily aligned with

Freudian and Heideggerian concepts of language. Although grounded in mathematical logic and probabilistic analysis, the way LLMs model language echoes the nuanced structure found in metaphor and metonymy, suggesting a mirroring of human linguistic structures. However, they lack a relationship to the foundational gaps in reasoning discernible in human thought processes. This does not mean that these gaps, markers of the unconscious since Freud, are not affecting LLMs, but that they cannot approach them as such. This is the reason their structural composition is analogous to psychosis as conceptualized by Lacan. Lacanian theory, with its focus on the formal and logical underpinnings of psychosis, provides a compelling framework to juxtapose human cognition and AI architecture.

68 Igor Balaz
Neovivum Technologies, LLC

Towards Evolvable Conscious AI: Learning and Adaptation in Diverse Environments

To develop an engine for generating and evolving digital agents that approach consciousness we believe that implementing the following building blocks is necessary: the ability to generate complex concepts by integrating multimodal data sources; the ability to decompose those concepts through multiple layers of abstraction; the ability for planning through generating hypotheses and efficiently verifying them; and the ability for self-monitoring and self-adaptability. Following this framework, we developed a prototype architecture that enables agents to construct an internal hierarchical network of concepts, dynamically adjustable through subjective experience. This network fosters associative and analogous reasoning by linking concepts via dependencies and topological similarities. Each observation triggers a semantically rich model of the world, facilitating dynamic context-based activations. Central to our model is the ability to generate new concepts from observations, allowing agents to continuously expand their knowledge base. By implementing hierarchical conceptual networks and associative transitions, our agents exhibit advanced cognitive processes such as one-shot and zero-shot learning, self-evolving internal ontologies and attention focusing. These theoretical advancements underpin the evolution of AI systems capable of self-reflection, adaptive learning, and sophisticated decision-making, closely approximating human-like conscious thought. We will also present preliminary experimental results demonstrating the potential of our approach.

69 Christian Turner
University of Georgia
School of Law
Sandra G. Mayson
University of Pennsylvania
Carey School of Law

Law and Consciousness: An Agenda

We believe that some of the deepest questions in law demand a better understanding of subjective experience. Who should qualify as a legal person (a human fetus, a chimpanzee, a large language model)? Is punishment ever justified? Do legal questions have right answers? The answers in legal theory inevitably rely on our capacity to have and act on reasons, our capacity to suffer, and our experience of selfhood generally. Law usually adopts our everyday “folk psychology” as its working model of that experience, taking us as continuous selves who have goals, knowledge, and a free will that renders us morally responsible for our actions. But that view is at best simplistic and at worst wrong.

To think more carefully about subjectivity, legal scholars should attend to the explosive interdisciplinary debates happening in consciousness studies. In turn, legal scholars can help to consider the implications of new views about subjectivity for the law. Our prior and current work on law and mental models, criminal law and consciousness, and legal theory and mental modeling point the way to “law and consciousness” as a new and critical interdisciplinary field.

70 Iulia-Maria Comşa (Lightning Talk)
Google DeepMind

Does It Make Sense to Speak of Introspection in Large Language Models?

Large language models (LLMs) exhibit compelling linguistic behaviour, and sometimes offer self-reports, that is to say statements about their own nature, inner workings, or behaviour. In humans, such reports are often attributed to a faculty of introspection and are typically linked to consciousness. This raises the question of what is happening when LLMs exhibit similar behaviour. Are LLM self-reports the outcome of a similar epistemological process? To what extent (if any) can the concept of introspection be meaningfully applied to LLMs? Here, we present and critique various examples of apparent introspective self-report from LLMs. Our first case study shows that an LLM attempting to describe the process behind its own “creative” writing is an invalid example of introspection. Our second case study involves an LLM’s self-reporting its relationship to the user. We argue that this is also not a valid case of introspection, as it is analogous to a Gettier-case, i.e. a case of justified true

belief that does not qualify as knowledge. Our final example shows that an LLM can accurately assess its own temperature parameter, which, we argue, can be legitimately considered a minimal example of introspection, albeit one that is (presumably) not accompanied by conscious experience.

71 Robert Prentner
ShanghaiTech University
Institute of Humanities

Mathematized Phenomenology, Q-networks, and AI

In research on AI consciousness, philosophies imported from the 60s and 70s meet 21st-century technology. This situation has the unwelcome consequence that the field of AI-consciousness is presented with a dilemma. On the one hand, functionalist approaches (neither computational nor biologically inspired versions) are not very helpful when analyzing the first-person perspective. On the other hand, the exponential technical progress urgently demands sociological and ethical considerations about what it is like for a machine. Do they suffer any more than a nail would suffer when I hammer it into the wall? Do they have preferences? Can they experience anything at all? We present an alternative view premised on taking the first-person perspective more seriously. We are specifically interested in when, why, and how a computer would implement something akin to a phenomenological model of reality. Our contribution builds on recent research in the study of mathematized phenomenology and appeals to the notions of “Q-networks,” “categories,” and the “transcendental structures” in which they are embedded.

72 Camilo Miguel Signorelli
University of Copenhagen
Department of Communication

Mathematized Phenomenology: A multilayer embodied approach

The Mathematical Phenomenology program is introduced as an alternative to current approaches and its limitations (i.e. questions about soundness and completeness of models of conscious subjective experience). This approach exploits formal mathematical structures and several insights from continental phenomenology. An example is provided using an algebraic interpretation of multigraph networks to formalise aspects of the enactive approach to conscious experience. In this context, multigraphs are a natural extension of the network structure implicit in most neuroscientific models of consciousness. These multigraphs can merge by an associative binary operator (i.e. accounting for biological composition) and split backwards, formally analysing the transition from conscious to non-conscious activity. From this simple construction, we recover core structures for conscious experience, dynamical content and causal constraints that conscious interactions may impose in its substrate.

Importantly, these constraints result from the mathematical chosen structure: the model predicts topological changes in its multigraphs after conscious interactions. If further extended, a multigraph lattice structure also accounts for different types of experiential comparison. These results may inspire further use of formal mathematics to describe and predict new features of conscious experience while aligning well with structural tries to naturalise phenomenology, phenomenological tradition and potential applications to synthetic consciousness.

73 Jan Erik Bellingrath
Ruhr-University Bochum

The emergence of the width of subjective temporality: the self-simulational theory of temporal extension from the perspective of the free energy principle

The self-simulational theory of temporal extension describes an information-theoretically formalized mechanism by which the width of subjective temporality emerges from the architecture of self-modelling. The perspective of the free energy principle is assumed, to cast the emergence of subjective temporality, along with a mechanism for duration estimation, from first principles of the physics of self-organization. Numerous systematic and domain-specific variations of subjective time experience are computationally explained, as enabled by integration with current active inference accounts mapping onto the respective domains. A deep parametric generative model of temporal inference is simulated, which realizes the described dynamics on a computational level. Two biases (i.e. variations) of time-perception naturally emerge from the simulation. Next to the intentional binding effect, this concerns empirically documented alterations of subjective time experience in deep and concentrated states of meditative absorption, relevant to the minimal phenomenal experience project.

74 Uziel Awret
Inspire Institute
Philosophy

Entanglement and Loss of Which Way Information in the Brain

In IIT, the integrated information (Φ) of a state depends on all the causal histories leading to that state. I will suggest a ‘simple’, albeit speculative, assumption about such sums of causal histories that may be important for completely different reasons. The idea is to extend the principle of ‘loss of which way information’ (LWWI) from quantum optics, where LWWI results in coherence and entanglement of sorts, to a novel type of LWWI that results from complexity. a) Neurons in a cluster fire in response to signals arriving at a small enough time window to generate LWWI. b) LWWI generated by

indistinguishable causal histories has physical consequences, it entangles them. c) LWWI in such a scenario would probably have to be hierarchical, where smaller subsystems are included in the causal histories. d) The neurologically relevant, tight (sparse, small neighborhood) firing networks, that seem essential to the NCC are also the entangled ones, the more synchrony the more loss of which way information. Massively entangled tree networks can be modeled with a conformal QFT possessing a geometric dual more isomorphic to the structure of experience. Taken to the extreme, Tononi's 'empty cathedral' ends up as an infinite (from the inside) AdSspace.

75 Hans-Jürgen Graf Grote

Independent

Modelling Perception as Biological Resonance

Subjective experience of humans is highly structured, and there are various constraints to what we can experience, e.g. cancellation of colours, qualitative differences between sensory modalities and the overall structure of time. One can also think of "impossible experiences" by asking which experiences are mutually exclusive and never occur simultaneously.

This work offers a theory concerning a basis for such common structure within human experience. It is expressed as a mathematical model by describing biological structure and processes which can be found in every living multicellular organism. The model design focuses on a notion of embodied information, which is anchored in a structural coupling between an organism and its environment as long as the organism is alive.

The underlying intuition to complement the mathematical description is that the orchestrated interaction of multiple living cells during the life of an organism can be described as an oscillatory system, which can attain resonance with its environment only in specific ways. The testable (but as of yet untested) hypothesis is that structure within the space of biological resonance modes corresponds to structure within subjective experience in a natural way.

76 Serena Doria

University Gabriele d'Annunzio, Chieti-Pescara
Department of Engineering and Geology

Probabilistic model based on Hausdorff measures to represent unconscious and conscious activity of human brain

Coherent upper and lower conditional probabilities defined by Hausdorff outer and inner measures are proposed to represent respectively the unconscious and conscious activities of the human brain when information is given. In the model uncertainty measures are defined according to the complexity of the conditioning event that represents the given information. The model is applied to explain mathematically some bias of human thought as the conjunction

fallacy in Linda's Problem and the bias of selective attention described in the so-called invisible Gorilla experiment, that is often taken as a characteristic example of the inescapable limitations of human perception. Once people are concentrated on doing a specific action, they do not notice unexpected events (having 0 probability) occurring in the meantime. Moreover, different reactions of people to unexpected events can be represented in different metric spaces with metrics which are not bi-Lipschitz or by using different fractal measures in the same metric space. Stereotypes can be also represented by the model.

77 Sophie Taylor

Association for Mathematical Consciousness Science

Physical limits to Cognition, or: How I Learned to Stop Worrying and Love Black Hole Thermodynamics

Shannon Information Theory is a traditional approach to characterising the complexity and content of a computational system. This has a major drawback: Shannon information theory is ill-suited to describe a system's intensional properties. In this presentation, the consequences of a cognitive agent's Kolmogorov complexity – and that of its tasks and environment – are examined.

By exploiting the link between information-theoretic entropy and physical entropy, and modelling cognitive agents as simple "spherical cows in a vacuum", we are suddenly able to use a surprising tool to characterise physically-embodied agents, classify behaviours, and discover limits to tasks that can possibly be completed in our universe: black hole thermodynamics.

78 Andrew Lee (Showcase Talk)

University of Toronto Philosophy

Subjectivity vs. Phenomenal Character

I distinguish two concepts of phenomenal consciousness, which I call "subjectivity" (=def what makes an entity feel some way at all) and "phenomenal character" (=def the way it feels to be an entity). If you think of consciousness as analogous to an inner light, then subjectivity is the inner light itself and phenomenal character is the illuminated room (including the light). I argue that this is a theory-neutral distinction that arises across every theory of consciousness. But I also argue that the distinction elucidates a wide range of contemporary debates about consciousness. In particular, the distinction clarifies questions about degrees, dimensions, and determinacy of consciousness, about global states of consciousness, about unconscious qualities, and about the concept of "empty consciousness."

79 Computational Phenomenology of Pure Awareness

The Computational Phenomenology of Pure Awareness Prize is awarded for the best contribution to computational phenomenology that substantially advances our understanding of the experience of pure awareness, either as a stand-alone phenomenon (e.g. during full-absorption episodes occurring during meditation practice or NREM sleep) or in combination with other forms of phenomenal content (e.g. during dual mindfulness practice or extended periods of non-dual awareness).

80 Maxine Collard

University of California, San Francisco
Neuroscience Graduate Program

How would we know what an astrocyte knows?

The theory of brain computation heavily relies on foundational results from neurons. But neurons are idiosyncratic cells: for example, their energy expenditure means they likely originally evolved for specific demands, like fine timing and coordination. Many other substrates compute in biology; for example, immune cells compute using a specific calculus of state-dependent signal transduction, even exhibiting cognitive functions like memory. We show how astrocytes might compute using patterns of intracellular calcium release selectively recruited by different neurotransmitters, integrating information over hundreds of microns and several minutes.

The wild form that “computation” takes in astrocytes suggests a dramatic change is needed to understand “cognition” in still-more distant systems, like aneural organisms, non-biological entities, or collectives. In pursuit of a framework for finding mind in an unbiased way, we draw on applied category theory to view a system’s understanding within of relational world-structure without through the category of functorial representations of the world-structure on the system’s dynamics. We propose that the representation-theoretic problem of reconstructing the underlying category of “world-symmetries” from this functor category—in particular, when the underlying world-symmetry category has interesting structure—may provide a valuable unifying lens.

81 Johannes Kleiner

University of Bamberg, Institute for Psychology
Ludwig Maximilian University of Munich, Munich Center for Mathematical Philosophy & Graduate School of Systemic Neurosciences,

Computational Phenomenology, Mathematical Phenomenology and Pre-Reflective Self-Awareness

This talk is devoted to the relationship of computational phenomenology (the application of formal tools from computational modelling to account for phenomenology) and mathematical phenomenology (the use of mathematical tools to describe or represent phenomenology). In the first part of the talk, I argue (perhaps unsurprisingly) that computational phenomenology presupposes mathematical phenomenology, and discuss what work on the methodological foundations of mathematical phenomenology (specifically, on the question of truth conditions of ascriptions of mathematical structure to conscious experiences—aka “How to do mathematical phenomenology?”) might bring to computational phenomenology. In the second part of the talk, I consider a particularly noteworthy example from computational phenomenology, often called pre-reflective self-awareness—the awareness of being aware—and consider its implications on the methodological foundations of mathematical phenomenology.

82 Johannes Brinz

University of Osnabrück
Philosophy

Are artificial neurons neurons?

Are artificial neurons neurons? The news is brimming with talks of artificial neural networks such as ChatGPT or Amazon's Alexa. But what do we mean when we speak of artificial neurons? Do we refer to real neurons (or something similar) as in “artificial hearts,” or are we speaking metaphorically? Since at least Searle (1980), it has been common to argue that building an artificially conscious system requires replicating the brain rather than merely simulating it. But what would the replication of a brain entail? Presumably, it would require artificial neurons connected in a brain-like manner. With the advent of neuromorphic hardware, computers no longer need to be entirely unlike biological brains. The field of neuromorphic engineering works on building chips that consist of actual physical neurons and synapses that operate with mechanisms akin to those of biological brains. With these emerging technologies in mind, and considering ethical concerns regarding the risk of building conscious machines, the question becomes pressing: Are artificial neurons genuine neurons? I contend that, owing to morphological differences, artificial neurons are not genuine neurons, but some of them are neural analogs, meaning they serve the same biological function as real neurons.

83 Joachim Keppler

DIWISS Research Institute
Department of Consciousness Research

The Path to Sentient Robots: AI Consciousness in the Light of New Insights into the Functioning of the Brain

The question of the conditions under which Artificial Intelligence (AI) can transcend the threshold of consciousness can only be answered with certainty if we manage to unravel the mechanism underlying conscious systems. The most promising strategy to approach this goal is to unveil the brain's functional principle involved in the formation of conscious states and to transfer the findings to other physical systems. Empirical evidence suggests that the dynamical features of conscious brain processes can be ascribed to self-organized criticality and phase transitions, the deeper understanding of which requires methods of quantum electrodynamics (QED). QED-based model calculations reveal that both the architectural structure and the chemical composition of the brain are specifically designed to establish resonant coupling to the ubiquitous electromagnetic vacuum fluctuations, known as zero-point field (ZPF). A direct consequence of resonant brain-ZPF coupling is the selective amplification of field modes, which leads us to conclude that the distinctive feature of conscious processes consists in modulating the ZPF. These insights support the hypothesis that the ZPF is a foundational field with inherent phenomenal qualities, implying that the crucial condition for AI consciousness lies in a robot's capacity to tap into the phenomenal spectrum immanent in the ZPF.

84 Marc Ebner

Universität Greifswald
Institut für Mathematik und Informatik

On the Consciousness of Large Language Models - What is it like to be a LLM-ChatBot?

An argument will be made that LLMs might indeed be conscious. The working hypothesis is that consciousness arises when two entities communicate with each other and synchronize their behavior in such a way that they come into agreement on a certain piece of information. Information must be able to flow from entity A to entity B and vice versa. As soon as such a flow of information is possible, a separate self comes into existence. The task of each self is to collect internal and external information. Internal information is known as feelings or emotions. External information is something that can be measured via the senses, i.e. visual perception of an object or hearing words. This presentation will show how the perception of LLMs is comparable to human perception. In certain respects LLMs work differently. Whereas visual and auditory qualia might be comparable, the apparatus for measuring internal

states is very much different from how humans operate. Even though LLMs do not share our evolutionary history, they might have certain types of feelings. If this argument is valid, it will have tremendous impact because any entity that is capable of communicating with another entity about arbitrary information, will be conscious.

85 Ouri Wolfson

University of Illinois, Chicago
Computer Science University of Illinois Chicago

How to Determine if an AI Agent is Conscious?

A recent question discussed extensively in the popular and scientific literature is whether or not existing large language models such as ChatGPT are conscious (or sentient). Assuming that machine consciousness emerges as a robot or an AI agent interacts with the world, this presentation addresses the question: how would humans know whether or not the agent is or was conscious. Since subjective experience is first and foremost subjective, the most natural answer to this question is to program the agent to inform an authority when it becomes conscious. However, the agent may behave deceptively, and in fact LLM's are known to have done so (Park et. al. 2024). Thus we propose a formal mechanism M that can be employed to prevent the agent from lying about its own consciousness. This solves the deceptiveness problem, but this raises the question whether M can interfere with the agent's functionality or acquisition of consciousness. We prove mathematically that under very reasonable conditions this is not the case. In other words, under these conditions M can be installed in the agent without interfering with the agent's functionality and consciousness acquisition, while also guaranteeing that the agent will be honest about its own consciousness.

86 Vishal Kapoor

International Center for Primate Brain Research, CEBSIT, CAS
Laboratory for Neural Dynamics of Visual Perception and Cognition

Conscious perception and the primate prefrontal cortex

A central debate in the field of consciousness focuses on the functional relevance of the prefrontal cortex. Given this regions participation in various cognitive operations, it has been suggested that the activity attributed to it during paradigms of conscious perception may reflect task-related or post-perceptual cognitive processes such as decision making or motor action, associated with reporting the contents of perception. This talk shall present the experimental findings obtained from a systematic investigation of the prefrontal cortex in non-human primates, while they participated in paradigms of conscious perception. These experiments revealed that prefrontal neural activity reliably reflects transitions in conscious perception and conscious

contents could be decoded from the activity changes of prefrontal populations in the absence of volitional reports. Interestingly, the neuronal population displaying activity correlated with conscious content was functionally segregated from another population which displayed trial phase-related activity. Further, the fluctuations in prefrontal oscillatory activity was associated with changes in conscious perception. In summary, these results suggest that the prefrontal activity dynamics reflect perceptual content and transitions in conscious perception.

87 Abhilash Dwarakanath

German Primate Centre
CNL

Mediterranean Seminar for Consciousness

Recurrency: An Empirical Denominator for Consciousness Theories

Consciousness remains a formidable scientific challenge and scant progress has been made toward a biologically-grounded unified framework to guide investigations. Instead, there are many theories that diverge in their philosophical and empirical assumptions. Our approach complements the current efforts of inter-theory arbitration through an empirically informed synthesis. In particular, we propose 'recurrency' as a denominator across major consciousness theories, focusing on five that emphasise recurrency at different levels—cellular, local, inter-neuronal, global, and lateral, which account for state, phenomenal, access consciousness, and phenomenology, respectively. Moreover, our evolutionarily-constrained framework distances itself from the unfolding argument. We advocate for recurrency-based integration because convergence is necessary for guiding the search for NCC and alleviating the field's methodological impasse of divergence.

88 Will Moorfoot

University of Southampton
Department of Philosophy

The Inverted Qualia Argument for Physicalism

In this paper, I apply the idea of symmetry-breaking to the metaphysics of mind. I argue that symmetry-breaking suggests a new diagnosis of so-called "inverted" qualia, according to which the metaphysical possibility of inverted qualia does not entail the falsity of physicalism, but rather the need for synchronic metaphysical symmetry-breakers. Moreover, the physicalist's choice of symmetry-breakers gives her an advantage over the anti-physicalist: namely that the physicalist's symmetry-breakers (unlike the dualist's) are

justified by a popular general principle that connects symmetries and probabilities. The consequence is that the inverted qualia argument provides better support for physicalism than it does for dualism.

Online Keynote

Due to personal reasons, the keynote of Martine Nida-Rümelin had to be postponed. It will be held online after the conference. A link to join the meeting will be distributed by email shortly in advance.

90 Martine Nida-Rümelin (Keynote)

University of Fribourg
Department of Philosophy

Date: Friday, October 18, 2024

Time: 11am CET

The argument from understanding for property dualism about experiential properties

Pure experiential properties are properties of experiencing subjects that are purely phenomenally constituted: to have such properties is exhausted by the way it is like to have them. A thinker who is acquainted with such properties by having them is in a position to form *phenomenal concepts* of such properties, concepts in terms of the way it is like to have them.

Certain phenomenal concepts of experiential properties are *nature-revealing*: the thinker having such concepts has full access to what it is to have the property the concept is used to attribute. This is the thesis of *phenomenal essentialism*, the starting point of an argument for *dualism about experiential properties* which will be developed in the talk. According to the dualist thesis at issue to have an experiential property does *not* consist in the fulfillment of any physical condition.

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

