Models of Consciousness 2023

A conference on mathematical approaches in the scientific study of consciousness

Mathematical Institute, University of Oxford
September 4-8, 2023

Association for Mathematical Consciousness Science

Sponsors and organising institutions

Association for Mathematical Consciousness Science

FQXi
FOUNDATIONAL QUESTIONS INSTITUTE

Barbara Wengeler Stiftung

OMCAN
Oxford Mathematics of Consciousness and Applications Network

UNIVERSITY OF OXFORD

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

RUB

SAINT ANSELM COLLEGE

mc/Mp Munich Center for Mathematical Philosophy

UNIVERSITY OF BAMBERG
OTTO-FRIEDRICH-UNIVERSITÄT BAMBERG

PROGRAMME

Table of Contents

Welcome

It is a great pleasure to welcome all those participating in the fourth instalment of the Models of Consciousness conference series (MoC4-2023). Even in the short time since MoC3 at Stanford in 2022 there has been substantial developments in areas such as AI. If anything, the rise of the LLMs such as ChatGPT and Bard have highlighted the now urgent need for consciousness science to make progress. In the coming years science will need to know what makes the difference between systems that have consciousness, such as the human brain, and those that do not. Accordingly, in April 2023, the AMCS Board and President published an open letter titled “The Responsible Development of AI Agenda Needs to Include Consciousness Research”.

Consciousness science is a broad field with many competing models and hypotheses. Mathematical conscious science is translating systems, hypotheses and phenomena into the mathematical domain so that the resulting models of consciousness can be objectively tested against our scientific knowledge of consciousness. MoC4 seeks to stimulate this evidence-based approach through an open-minded scientific exploration of the possibilities. This is an exciting and important time for the field.

By kind invitation of the Oxford Mathematics of Consciousness and Applications Network (OMCAN), MoC4 takes place at the Mathematical Institute, University of Oxford, UK.

Welcome to Oxford!

Website at: https://amcs-community.org/events/moc-4-2023

The conference team

Organisers*

Dr Jonathan Mason
Oxford Mathematics of Consciousness and Applications Network (OMCAN)
Mathematical Institute
University of Oxford
UK

Dr Robin Lorenz
Quantinuum
Oxford
UK

Dr Johannes Kleiner
Munich Center for Mathematical Philosophy
Ludwig Maximilian University of Munich;
University of Bamberg
Germany

Dr Robert Prentner
Munich Center for Mathematical Philosophy
Ludwig Maximilian University of Munich
Germany

Dr Wanja Wiese
Institute for Philosophy II
Ruhr University Bochum
Germany

Advisory board

Prof. Ian Durham
Department of Physics
Saint Anselm College
NH USA

Prof. Yakov Kremnitzer
Oxford Mathematics of Consciousness and Applications Network (OMCAN)
Mathematical Institute
University of Oxford
UK

*The advisory board gratefully acknowledges the work and effort of all the organisers and gives particular thanks to Dr Mason and Dr Lorenz who lead the work.

Excursion

Here are some details for invited speakers and those signed up for the excursion on **Monday September 4th**.

Historical walking tour of Oxford including New College chapel:

- 10:00am - Please assemble at the blue gates of Trinity College on Broad St, Oxford, OX1 3BH.

Lunch:

- 12:00pm - We will distribute ourselves among a few cafes and pubs in central Oxford such as the Vaults and Garden Cafe, the High Street Cafe, the Turf Tavern and the Kings Arms. The excursion fee does not cover the cost of lunch.

Punting:

- 2:00pm - Please assemble at the Magdalen Bridge Boathouse on the High Street, Oxford, OX1 4AU. The punts are hired for one hour.

Conference dinner and drinks reception

Drinks reception

We are delighted to invite the whole conference to a drinks reception from 5:50pm to 7pm on Tuesday 5th September. The drinks reception will be held in the Mathematical Institute's common room; see the building plan on page 8. The conference organisers would like to thank Etienne Jacques who kindly provided some sponsorship in support of the reception. **Please join everyone for drinks.**

Conference Dinner

At 7pm, on Wednesday 6th September, invited speakers and those who paid will come together for the conference dinner at St John's College, St Giles' Street, Oxford OX1 3JP; also see page 12.

Menu

Fig & Goats Cheese Salad, with balsamic caramel (V)

*****

Roast Fillet of Beef, with a Pink Peppercorn Crust, served with Red Wine Jus

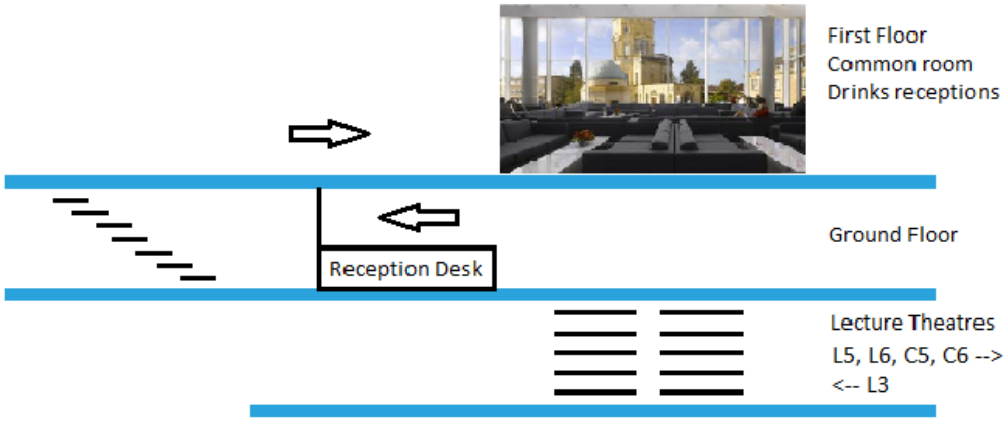
Root Vegetable Tarte Tatin (V)

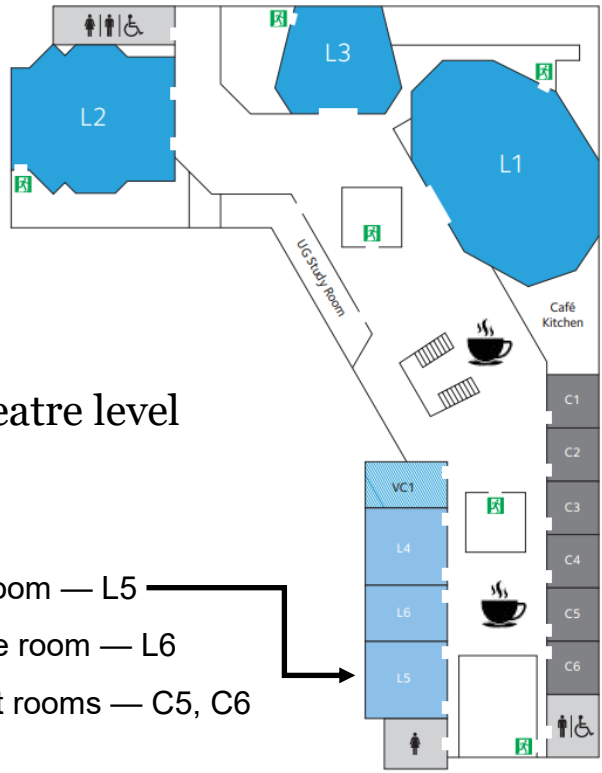
Boulangère potatoes; mixed vegetables

*****

Baked Chocolate Mousse with Orange Sorbet & Almond Biscuit

Building plan





Lecture theatre level & Café

Main lecture room — L5

Second lecture room — L6

Extra breakout rooms — C5, C6

Mathematical Institute, University of Oxford, Andrew Wiles Building,
Radcliffe Observatory Quarter (550), Woodstock Road, Oxford, OX2 6GG

Coffee, tea and lunch

Coffee and tea breaks

Coffee and tea are provided by the conference twice a day in the North Mezz outside lecture rooms L5 and L6.

Lunch

Conference participants are welcome to have lunch in the Mathematical Institute's Café Pi located on the same level as the lecture theatres. Alternatively, there are many local pubs and restaurants within a few minutes walk of the Mathematical Institute along Walton Street, and Little Clarendon Street; see the next page.

Area map

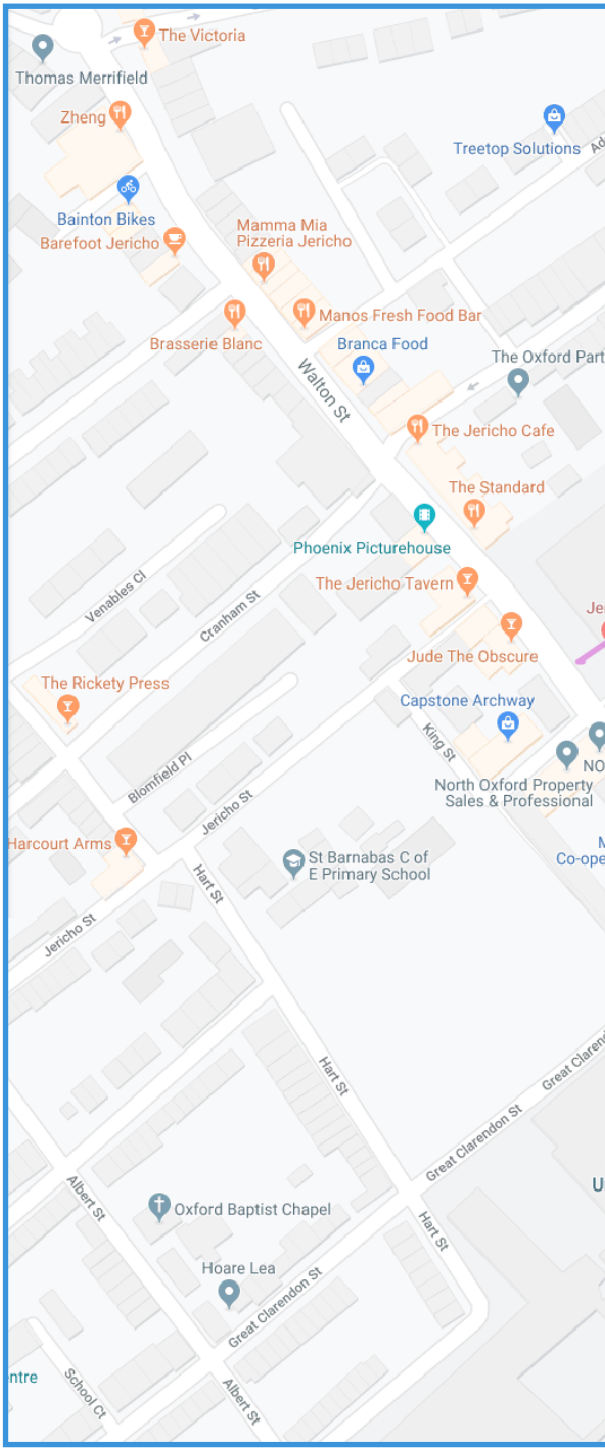
Eating and drink out

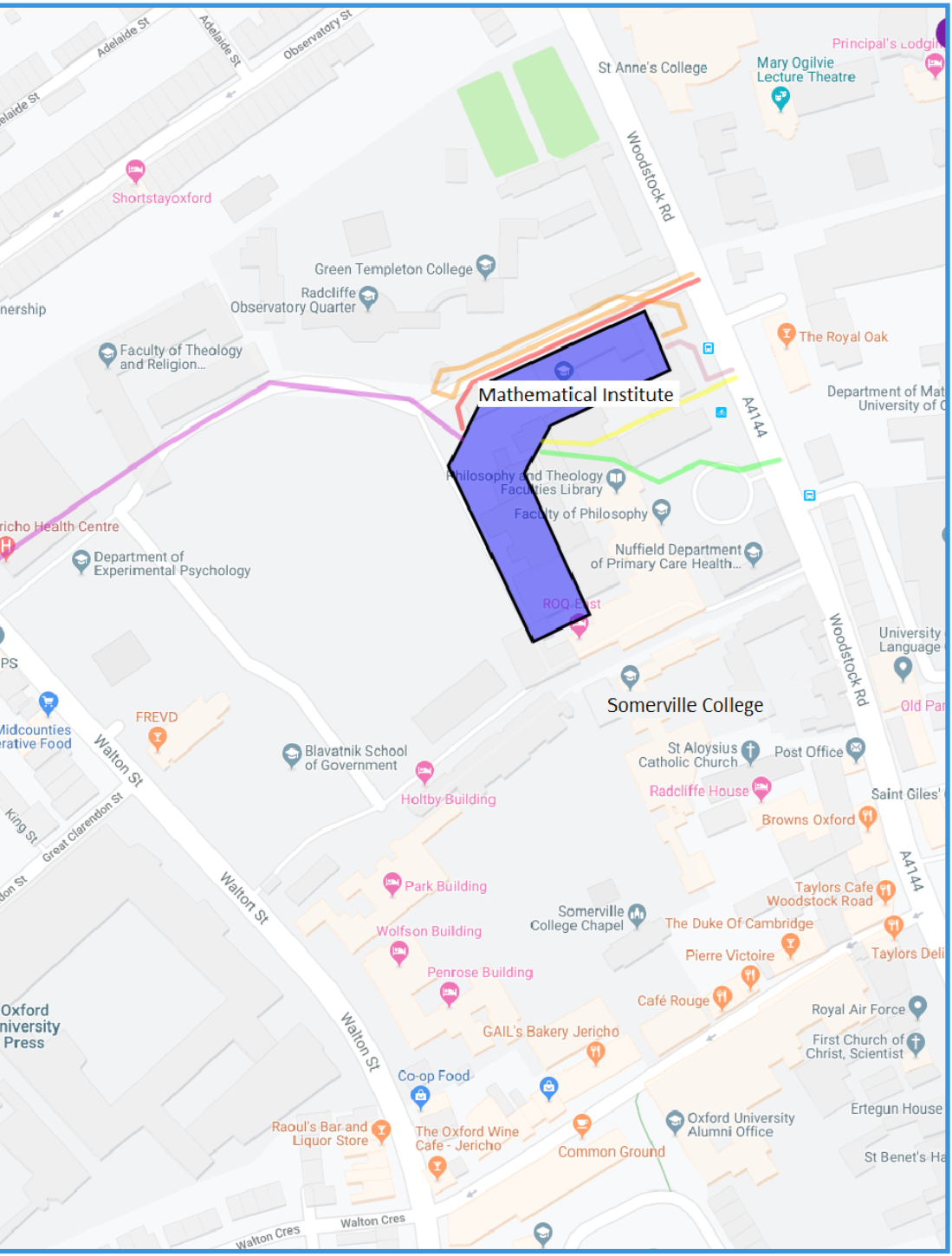
Restaurants:

- Café Pi, Mathematical Institute
- The Pierre Victoire (French Cuisine)
- Branca (deli and restaurant)
- Opera Café
- Al-Shami (Lebanese/Syrian)
- Café Rouge

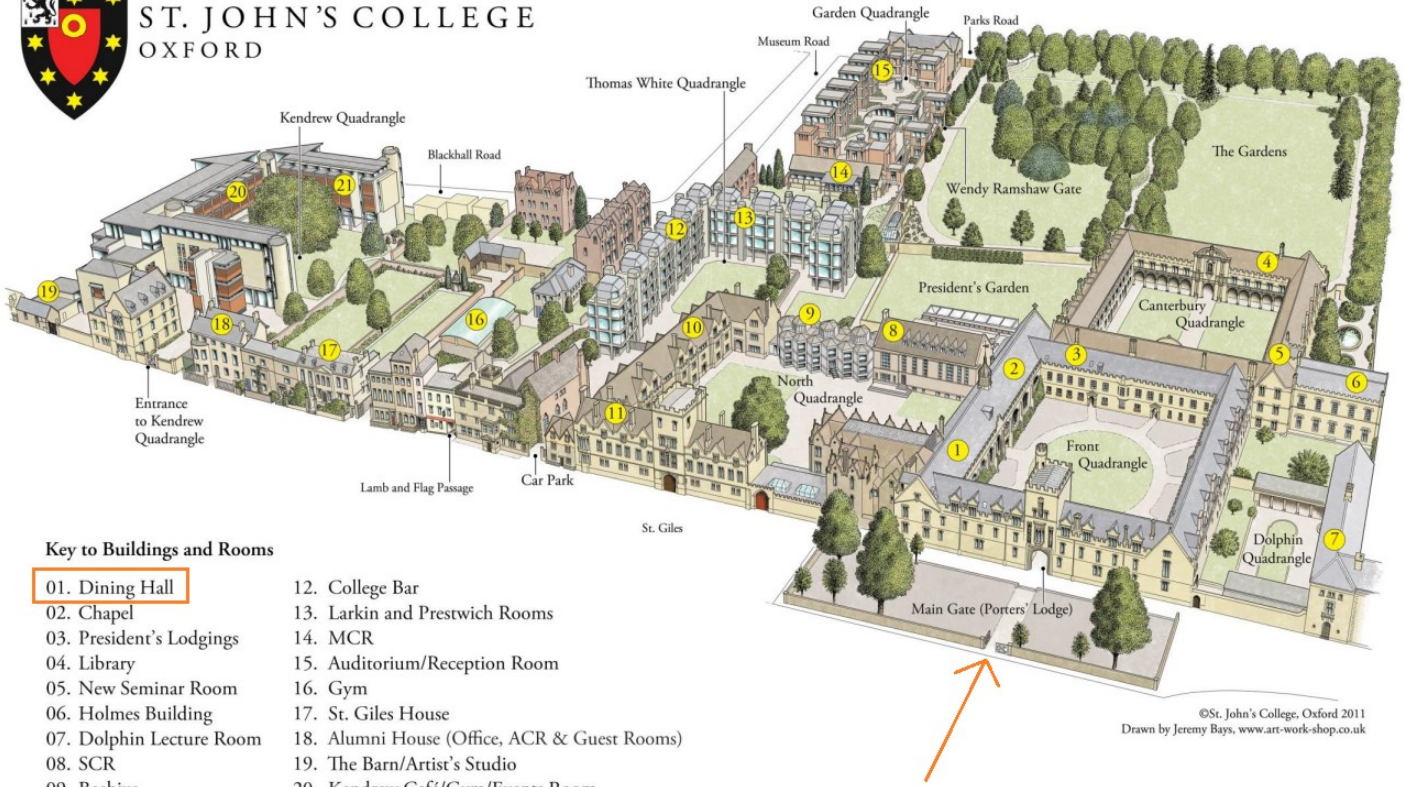
Pubs (typically good food too):

- The Royal Oak
- The Gardener's Arms on Plantation Road (good and exclusively vegetarian menu)
- Eagle and Child (for Toklien and C.S.Lewis fans)
- Lamb and Flag
- The Rose and Crown
- The Victoria
- The White Rabbit (they have vegan pizza)
- The Rickety Press
- Jude the Obscure
- The Jerico Tavern
- The Harcourt Arms
- The King's Arms




Mathematical Institute
Somerville College
Adelaide St
Observatory St
Woodstock Rd
St Anne's College
Mary Ogilvie Lecture Theatre
Shortstayoxford
Green Templeton College
Radcliffe Observatory Quarter
Faculty of Theology and Religion...
The Royal Oak
A4144
Philosophy and Theology Faculties Library
Faculty of Philosophy
Nuffield Department of Primary Care Health...
Department of Experimental Psychology
ROQ East
University Language
FREVD
Walton St
Blavatnik School of Government
Holtby Building
St Aloysius Catholic Church
Post Office
Radcliffe House
Browns Oxford
Great Clarendon St
King St
Park Building
Somerville College Chapel
Taylors Cafe Woodstock Road
Wolfson Building
The Duke Of Cambridge
Pierre Victoire
Penrose Building
Café Rouge
Royal Air Force
GAIL's Bakery Jericho
First Church of Christ, Scientist
Co-op Food
Raoul's Bar and Liquor Store
The Oxford Wine Cafe - Jericho
Common Ground
Oxford University Alumni Office
Ertegun House
Walton Cres

ST. JOHN'S COLLEGE
OXFORD



Key to Buildings and Rooms

01. Dining Hall
02. Chapel
03. President's Lodgings
04. Library
05. New Seminar Room
06. Holmes Building
07. Dolphin Lecture Room
08. SCR
09. Beehive
10. Bursary
11. North Lecture Room
12. College Bar
13. Larkin and Prestwich Rooms
14. MCR
15. Auditorium/Reception Room
16. Gym
17. St. Giles House
18. Alumni House (Office, ACR & Guest Rooms)
19. The Barn/Artist's Studio
20. Kendrew Café/Gym/Events Room
21. Law Library

Programme schedule

	Monday
10:00 - 4:00	Excursion Assemble at the blue gates of Trinity College on Broad St, Oxford, post code OX1 3BH

	Tuesday	
From 8:15	Registration (North Mezz)	
8:45 - 9:00	Welcome (L5)	
9:00 - 10:00	Invited Talk (L5) **1 Will Zeng**	
10:00 - 10:30	Coffee (North Mezz)	
10:30 - 10:55	**2 Ian Durham** (L5)	**4 Michael Coughlan** (L6)
10:55 - 11:20	**3 Masanao Ozawa** (L5)	**5 Marc Ebner** (L6)
11:20 - 11:45	**6 Pete Grindrod** (L5)	
11:45 - 12:45	Discussion Session (L5, L6, C5, C6)	
12:45 - 2:00	Lunch	
2:00 - 3:00	Chaired Q&A Session (L5) **7 Roger Penrose**	
3:00 - 3:30	**8 Tim Palmer** (L5)	
3:30 - 4:00	Coffee (North Mezz)	
4:00 - 4:25	**9 Kobi Kremnizer** (L5)	**11 Rulin Xiu** (L6)
4:25 - 4:50	**10 Sophie Taylor** (L5)	**12 Santosh Helekar** (L6)
4:50 - 5:50	Discussion Session (L5, L6, C5, C6)	
5:50 - 7:00	Drinks Reception (CR)	

Key: **North Mezz**=Outside rooms L5 and L6; **L5**=Main lecture room;
L6=Second lecture room; **CR**=Common Room; **C5**,**C6**=Extra breakout rooms

	Wednesday	
9:00 - 10:00	Invited Talk (L5) **13 Jan Westerhoff**	
10:00 - 10:30	Coffee (North Mezz)	
10:30 - 10:55	**14 Ron Chrisley** (L5)	**17 Kenneth Williford** (L6)
10:55 - 11:20	**15 Ovidiu Cristinel Stoica** (L5)	**18 David Rudrauf** (L6)
11:20 - 11:45	**16 Tetiana Grinberg** (L5)	**19 Zhouwanyue Nata Yang** (L6)
11:45 - 12:45	Discussion Session (L5, L6, C5, C6)	
12:45 - 2:00	Lunch	
2:00 - 3:00	Invited Talk (L5) **20 Lenore Blum**	
3:00 - 3:30	**21 Manuel Blum** (L5)	
3:30 - 4:00	Coffee (North Mezz)	
4:00 - 4:25	**22 Wanja Wiese** (L5)	**24 Moritz Kriegleder** (L6)
4:25 - 4:50	**23 Aïda Elamrani** (L5)	**25 Erik Curiel** (L6)
5:00 - 6:00	Special Evening Lecture (L5 or L3 tbc) **26 Marcus Du Sautoy**	
6:00 - 7:00	Free Those attending the conference dinner walk over to St John's College	
7:00 - 9:00	Conference Dinner at St John's College, St Giles' Street, Oxford OX1 3JP	

	Thursday	
9:00 - 10:00	Invited Talk (L5) **27 Claudia Passos**	
10:00 - 10:30	Coffee (North Mezz)	
10:30 - 10:55	**28 Jonathan Mason** (L5)	**30 Panagiota Theodoni** (L6)
10:55 - 11:20	**29 Christoph von der Malsburg** (L5)	**31 Keenan Down** (L6)
11:20 - 11:45	**32 Joscha Bach** (L5)	
11:45 - 12:45	Discussion Session (L5, L6, C5, C6)	
12:45 - 2:00	Lunch	
2:00 - 3:00	Invited Talk (L5) **33 Robert Prentner**	
3:00 - 3:30	**34 Chetan Prakash** (L5)	
3:30 - 4:00	Coffee (North Mezz)	
4:00 - 4:25	**35 Pedro Resende** (L5)	**37 Alex Kostova** (L6)
4:25 - 4:50	**36 Quanlong Wang** (L5)	**38 Hye Young Kim** (L6)
4:50 - 5:50	Lightning talks (L5)	
5:50 - 6:50	Discussion Session (L5, L6, C5, C6)	
7:00 - 9:00	By invitation only, Worcester College Dinner	

	Friday	
9:00 - 10:00	Invited Talk (L5) **39 Maja Spener**	
10:00 - 10:30	Coffee (North Mezz)	
10:30 - 10:55	**40 Renee Ye** (L5)	**43 Ouri Wolfson** (L6)
10:55 - 11:20	**41 Stephen Fleming** (L5)	**44 Malcolm Lowe** (L6)
11:20 - 11:45	**42 Kenneth Shinozuka** (L5)	**45 Peter Coppola** (L6)
11:45 - 12:45	Discussion Session (L5, L6, C5, C6)	
12:45 - 2:00	Lunch	
2:00 - 3:00	Invited Talk (L5) **46 Paul Azzopardi**	
3:00 - 3:30	**47 Johannes Kleiner** (L5)	
3:30 - 4:00	Coffee (North Mezz)	
4:00 - 4:25	**48 Vladimir Aksyuk** (L5)	**50 Diana Stanciu** (L6)
4:25 - 4:50	**49 Javier Galadi** (L5)	**51 Serena Doria** (L6)
4:50 - 5:50	Discussion Session (L5)	

Key: **North Mezz**=Outside rooms L5 and L6; **L5**=Main lecture room;
L6=Second lecture room; **CR**=Common Room; **C5**,**C6**=Extra breakout rooms

Talk titles and abstracts

1 Will Zeng (Invited)

Unitary Fund; Quantonation; Goldman Sachs & Co.

Local friendliness violations: experimental metaphysics and the nature of the observer

The objectivity of physical events and the local agency of experimenters are fundamental assumptions about reality and our scientific understanding of it. Recent work has formalized these assumptions in the concept of Local Friendliness and shown that standard quantum mechanical theory appears to violate them through extended Wigner's Friend scenarios. This suggests the need for an experimental program to confirm or refute these Local Friendliness violations. We introduce the theoretical background, summarize the status of experiments so far, and offer suggestions for future experimental directions. Such a program has potential to intersect heady concepts like the nature of the observer in quantum mechanics, including potential connections to consciousness, with concrete, quantitative experiments.

2 Ian Durham

Department of Physics, Saint Anselm College, United States

Taking Wigner seriously: what IIT and its quantum extensions say about superpositions of conscious states

Can superpositions of conscious states actually exist? Wigner famously proposed a thought experiment that raised this perplexing question. We find, in the context of IIT and its most recent extension, that, though the theory prevents superpositions of phi values in simple systems, it does allow for superpositions of Phi-structures or “Q-shapes”, and thus superpositions of conscious states. We develop an operator model of such Q-shapes and combine it with a generalized collapse model to show that the rate of collapse is dependent on more than simply the distance between the Q-shapes. In the process we elucidate a number of previous vagaries in the structure of IIT itself and its most recent quantum extension. This makes some progress toward addressing Wigner's original question. The work presented here was jointly performed with Kelvin McQueen and Markus Müller, and highlights the recent quantum extension of IIT by Albantakis, Prentner, and Durham.

3 Masanao Ozawa

Center for Mathematical Science and Artificial Intelligence, Academy of Emerging Sciences, Chubu University, Japan

Quantum instrument models in cognitive science: the question order effect and the non-Bayesian belief update

A quantum instrument describes the most general type of state update caused

by a physical measurement on a quantum system. Recently, we successfully apply this notion to explain non-Bayesian belief update appearing in the "question order effect" in cognitive psychology. A previous quantum approach attempted by Wang and Busemeyer has taken the projective belief update hypothesis, but their model does not satisfy another known effect called the "response replicability effect". It was once questioned whether a quantum-like model can simultaneously realize both of the two effects. We have eventually shown that a quantum instrument model has answered the question affirmatively. Quantum instrument models provide a common feature of both state updates in general quantum measurements and non-Bayesian belief updates. We also argue that this approach successfully separates two aspects of non-classicality in consciousness science, non-Bayesian property of belief updates and non-classicality of logic of consciousness, so that the "question order effect" represents the non-Bayesian property of belief update but that the "response replicability effect" represents classicality of logic of consciousness. Our model shows that non-Bayesian belief updates and classicality of logic of consciousness peacefully coexists mathematically, and rejects the projective belief updates hypothesis taken by the Wang-Busemeyer model.

4 Michael Coughlan

Experimental Zoology Group, Wageningen Institute for Animal Science, Wageningen University and Research, Netherlands

Neural networks as minimal models for brains and consciousness - closing the gap between artificial and biological

Recent advances in the fields of neuroscience and machine learning with deep neural networks have led researchers to re-evaluate the use of artificial neural networks as a minimal model for the brain. Artificial and biological neural networks share several key ingredients - a network architecture through which signals can propagate, a set of learning rules to adapt this network, and an optimisation process to guide this learning towards an ideal. Further, both are rich in emergent behaviour. This suggests that artificial neural networks might be a useful model to understand how consciousness might arise from a simple set of rules and interactions. However, several major differences exist in how biology and computer science understand and use neural networks. With the aim of narrowing these differences in understanding, much can be learned by considering small neural networks with biologically inspired modifications, learning simple computer visions tasks. In this talk, I will present numerical experiments on neurogenesis that demonstrate the propensity for new neurons to stabilise rather than destabilise learning. The discussion will include work on multi-task learning, which artificial neural networks struggle with, but biological neural networks handle with ease.

5 Marc Ebner

Institut für Mathematik und Informatik, Universität Greifswald, Germany

A Mathematical Model of the Quale Color

Understanding color perception is of paramount importance to the

understanding of consciousness. It appears as if color is a subjective experience. Many philosophical arguments have been exchanged on the question whether or not it is possible that my perception of "red" corresponds to your perception of "red". A mathematical description will be given what the quale color is about. It does not correspond to the values that are measured by the sensors. The measured values are transformed so that they correspond to a description of the reflectance function of the object that is being viewed. If this information can be spatially localized, we obtain a color image of our surrounding. This estimate improves if more sensors are available. However, the result is still a one dimensional function irrespective of two, three or four cones are being used to obtain this estimate. Hence, color is grounded in reality and comparable across individuals. Reflectance is a measurable quantity of our environment. Even seemingly surprising results in color perception can be explained by the mathematical or rather algorithmic description of estimating the reflectance function of an object. The quale color is due to the mathematical structure of the space of reflectance.

6 Pete Grindrod

OMCAN, Mathematical Institute, University of Oxford, UK

Wishful thinking about consciousness

We consider three distinct but complementary mathematical approaches to the hard problem: quantum consciousness, IIT, and large scale dynamics systems simulation. We identify some strengths and weaknesses of each, and some alignments and conflicts. Our aim is to promote high level dialogue, focussing on the distinct overarching aims and claims of each, and the evidential and evolutionary basis for each, rather than luxuriating in the weeds. Conclusions for discussion are drawn.

7 Nobel Laureate, Sir Roger Penrose FRS (Invited)

Mathematical Institute, University of Oxford, UK

Chaired Q&A session

A chaired Q&A session involving both selected questions seen in advance and questions from the audience, on Orch OR and wider topics.

8 Tim Palmer

Department of Physics, University of Oxford, UK

Consciousness and free will: A physical hypothesis for distinguishing Humans from AI

I present a physical hypothesis for free will and consciousness that distinguishes us from contemporary AI. Its origin lies in two postulates. The first is that the brain makes use of quantum processes because the corresponding classical processes are too energy inefficient. The second, following David Deutsch, is that the physical resource which gives quantum computing its advantage over classical computing is processing over physically real parallel worlds. I am not an Everettian but instead have my own version of quantum

physics (called Invariant Set Theory) where the wavefunction can be interpreted as an ensemble of state-space trajectories in a small neighbourhood of a dynamically invariant fractal subset of state space. From these postulates, I hypothesise that our cognition has a weak perception of physically real alternative worlds, very similar but not quite identical to our own. I claim this creates a perception of having an existence semi-independent of the rest of world. Such free will/conscious perceptions could not be experienced by AI systems running on classical computers. The invariant sets of chaotic systems are non-computable geometries (as shown by Blum et al 1997). This gives support to Penrose's claim that human consciousness and understanding involve inherently non-computable processes.

9 Kobi Kremnizer

OMCAN, Mathematical Institute, University of Oxford, UK

Awareness as a field: a quantum collapse model

In this talk I will discuss the possibility of modelling awareness/consciousness as a field. In this approach the space of field configurations could be interpreted as qualia with localised configurations such as vortices and instantons as definite qualia. I will then give an example of a quantum collapse model which includes two fields: the wave function describing matter, and a classical stochastic field describing awareness. The two fields interact describing the dynamics of matter and awareness. I will end by speculating on an underlying unified theory from which the matter and awareness fields emerge.

10 Sophie Taylor

School of Mathematics, Queensland University of Technology, Australia

Apartness relations, symmetry breaking, and cognition: Constructive existence from difference

If a model of consciousness has any interesting topological spaces, and there are any dynamical fields on these spaces, then I believe it's important to consider the possibility of topological defects of those fields, and their dynamics. This may even lead to approaches for laboratory study, through the use of analog fields. For example; analog black holes have been demonstrated in the laboratory in a number of settings (such as sonic black holes, optical black holes, and even in fluids such as water), with some of these models even producing effects analogous to Hawking radiation. This suggests that the possibility that similar experiments could be conducted on models of consciousness with such topological defects. Furthermore, how these defects can arise via symmetry breaking has intriguing links to computation and constructive mathematics, suggesting that these three fields share some common underlying structure.

11 Rulin Xiu

Hawaii Theoretical Physics Research Center, Grand Unification Theory, Tao Academy, USA and Canada

Quantum theory of consciousness

As two of the most fundamental existence, it should not be a coincidence that both consciousness and quantum phenomena share the similar subjective and probabilistic nature. In this presentation, four reasons are presented to show why consciousness is a quantum phenomenon and should be studied with quantum physics. A quantum theory of consciousness (QTOC) is proposed based on a new interpretation of quantum physics, in which everything arises from a vibrational field carrying matter, energy, and information, mathematically described by wavefunction. This QTOC provides physics foundation and mathematic formulation to address the easy and hard problem of consciousness. We demonstrate how to apply this QTOC to develop various models of consciousness. We show how it can explain the large-scale nearly instantaneous synchrony of the brainwaves such as gamma, beta, and alpha brainwaves and why and how they are correlated with Schumann Resonances. We point out how recent progress in quantum information theory, especially about quantum entanglement, can be applied to study neuron network and shed new light in neuroscience. We indicate why and how free will, destiny, cause-effect, environment, brain structure, and human action all play a critical role in one's consciousness and life experience.

12 Santosh Helekar

Translational Biomagnetics and Neurometry Program, Neurosurgery, Neurology and Psychiatry, Houston Methodist Research Institute, United States

Discovery of a peri-somatic physical effect related to consciousness – Instrumental recording using a potential noninvasive consciousness measuring device

A quantum physical effect that uniquely underlies, or is produced by, consciousness-associated brain activity patterns could potentially be measured by an instrument. We reasoned that if consciousness has a causal role in the brain, this effect might spread spatially within and outside it, in which case it might modulate a quantum phenomenon such as light wave interference produced by a double-slit apparatus in the peri-cranial space. Using a newly developed noninvasive photoelectronic device we discovered that wakefulness and sleep produce robust and reproducible differential modulation of the intensity of diffracted light waves generated by a low power laser diode close to the body. A similar modulation response was observed in mice as well as invertebrates, but an inverted response was detected in plants. It was significantly altered by general anesthesia in mice and showed striking variations for 4 – 5 hours after euthanasia, which depended on whether a mouse was decapitated or not. These results suggest that while the electrical activity of neurons is unlikely to be responsible for the effect, its underlying mechanism might involve molecular interactions within the brain that persist

after death, pointing towards a previously unrecognized neural biophysical mechanism for consciousness, requiring a rigorous mathematical description.

13 Jan Westerhoff (Invited)

Theology and Religion, University of Oxford, UK

The Fitness-beats-truth theorem: some philosophical reflections
The Fitness-beats-truth theorem raises various interesting philosophical problems that have so far received only very limited attention. In my presentation I will focus on three: the notion of veridicality (What notion of veridical representation is presupposed when it is argued that the probability that our model represents the world in a veridical manner is practically zero?), the status of mathematics (Why should we assume that unlike our perceptual representations, our mathematical representations faithfully represent the world as it is?), and the problem of self-refutation (Does the Fitness-beats-truth theorem undermine the very resources required for establishing it?).

14 Ron Chrisley

COGS/Sackler Centre for Consciousness Science, Informatics, University of Sussex, United Kingdom

Simulation and indexical function individuation: Why the diagonal argument does not establish the impossibility of machine consciousness
Diagonal arguments against algorithmic accounts of consciousness date back to Lucas, Turing and Gödel. Penrose's formulation, which is the focus of this talk, uses the notion of a "Turing Machine Non-Halting Question" to establish not only the impossibility of algorithmic consciousness, but the impossibility of a digital computer to even simulate the behaviour of, e.g., conscious human mathematicians. I show that Penrose's negative conclusions do not follow from his formal results. In particular, I show that 1) x can perfectly simulate y even if x and y do not compute the same class of functions; and 2) x can be conscious even if there exist conscious agents y that x cannot simulate. I establish 1) by a) characterizing the class of (Turing) computational systems as members of a more general class of question-answering systems; b) showing that for every class of question-answering systems (even finite ones) there exists a question relatively isomorphic to the original Turing Non-Halting Question, and c) showing how two question-answering systems may compute different functions, extensionally construed and yet compute the same class of functions, indexically intentionally construed. These results thus make (mathematical) room for the algorithmic generation of conscious behaviour.

15 Ovidiu Cristinel Stoica

Theoretical Physics, National Institute for R&D in Physics and Nuclear Engineering Horia Hulubei, Romania

Does a computer think if no one is around to see it?
I show that a computer cannot have unambiguous thoughts, not even about a number. What we believe computers do is our own convention. It may seem

objective because we anchor it in the user interface. But many other conventions are possible, and they yield different computations, equally valid according to the principles of Computer Science. I prove that the alternative computations equally happen when a single computation is carried out, and in principle they can be accessed. I exemplify this with a program that computes the result for a given input, and then decodes it into the results for all other possible inputs. If thinking would be a computation, a computer would have different, possibly opposite thoughts, corresponding to many alternative computations it implements at the same time. I show probabilistically that the human mind does not have this ambiguity. Therefore, even if the human mind can be simulated by a computer, it cannot be reduced to computation.

16 Tetiana Grinberg

Symbiokinetics Inc, United States

Iudico ergo sum: a computational theory of emotion

The notion of emotion is central to understanding consciousness, yet computational models of consciousness that are able to yield a nuanced representation of valence and emotion remain relatively scarce. In this talk I will propose a functionalist theory of emotion that attempts to bridge this gap. In particular, I will explore the relationship between reasoning, attention, goal prioritization and the formation of valence. I will begin by discussing the functional necessity of specialized cognitive modules in the context of resource-constrained computation. I will then outline a framework that attempts to unify a multiplicity of specialized cognitive modules with different objectives into a global process architecture that defines the behavior of an agent. This architecture will be further augmented through the introduction of notions of attention and goal prioritization. I will use this conceptual machinery to construct a computational framework for the formation of the subjective experience of emotion and valence. Finally, I will apply the resulting theory of emotion to analyze the computational processes underpinning psychological phenomena such as emotional lability, state-dependent learning and addiction.

17 Kenneth Williford

Projective Consciousness Model Research Group: David Rudrauf, Grégoire Sergeant-Perthuis, Daniel Bennequin, Department of Philosophy & Humanities, University of Texas at Arlington, United States of America

The Projective Consciousness Model: Phenomenological Prolegomena

According to the Projective Consciousness Model (PCM), consciousness is structured by (3D) Projective geometry; and perception, action, and imagination entail perspectival transformations governed by the action of the Projective Group, driven by a value-optimization dynamics (e.g., variational Free Energy minimization). To set the stage for a more formal presentation of the PCM in a companion presentation, we here offer some of the phenomenological data that motivate the theory and briefly indicate how the theory captures them. These data include: the sense of perspective or point of view (in vision but in other sensory modalities and imagination as well), the elusiveness of the point of

view (or "subject"), the sense of a vanishing point or limit of experiential space, the anticipatory dynamics involved in the perspectival navigation of ambient Euclidean space, reciprocity (and what we call "proto-intersubjectivity"), perspectival anomalies (e.g., OBEs, autoscopy, heautoscopy), some common illusions (esp. the Moon Illusion and the Ames Room Illusion), and key features of pre-reflective self-consciousness. We close by illustrating how the PCM generates novel predictions that can be tested using fairly straightforward psycho-physical paradigms, provides guidance for simulation strategies geared toward artificial intelligence, and raises some intriguing theoretico-mathematical questions.

18 David Rudrauf

CIAMS, F2S, University Paris-Saclay, France

The Projective Consciousness Model: Formal developments from projective geometry to epistemic drives. Grégoire Sergeant-Perthuis, David Rudrauf, & Kenneth Williford

The Projective Consciousness Model (PCM) yields a testable framework for understanding the relationships between consciousness, cognition, and behaviours. See companion presentation of Williford et al. for motivation of the model. In the PCM, the space of conscious experience acts as a homogeneous workspace for representation and action planning. 3D projective geometry models the subjective perspective of agents, and the 3D projective group structures their internal representations for active inference or more generally (stochastic) optimal control. The action of the projective group directly contributes to maximising expectation satisfaction and information gain, resulting in different adaptive and maladaptive strategies of exploration and exploitation, including in agents capable of Theory of Mind. We proved mathematically how changing the group (Euclidean versus Projective) that structures the agents' internal space strongly influences how « epistemic value » is quantified and maximised, and ensuing agents' exploratory behaviours. We discuss how this extends to general groups and the stability of group structured representations. The PCM offers a framework for designing experiments with human participants, to falsify or validate the predictions of the model. Our goal is both to build a theory of consciousness, and derive consciousness-inspired principles for the control of social robots.

19 Zhouwanyue Nata Yang

MCMP, LMU, Germany

What justifies the application of the Active Inference Theory of consciousness?

The Active Bayesian Inference Theory (AIT) posits a process-based explanation for the emergence of consciousness. Despite its extensive explanatory power, the underlying conceptual foundation of AIT remains underexplored. This talk introduces this theory within an explanation framework consisting of the explanatory account and the to-be-explained subject. In term of category theory, we investigate the implicit presupposition that justifies AIT's

application in explaining consciousness emergence. Within this framework, we consider two types of prediction-perception processes: one employed by AIT to study consciousness, and the other that deviates from the AIT's prerequisites. Both types of prediction-perception processes are embedded in a category equipped with an evaluation morphism. In light of Lawvere's Fixed Point theorem, we explicate criteria for AIT's application in consciousness studies. Building upon this, we employ categorical pullback to examine the formal condition of AIT's inherent presupposition. Finally, we interpret this presupposition as cognitive capacity, which not only retains AIT's deflationary nature but, crucially, also provides a justification for AIT's application.

20 Lenore Blum (Invited)

Center for conscious AI, Computer Science, Carnegie Mellon University and UC Berkeley, United States

A theoretical computer science perspective on consciousness and artificial general intelligence (AGI)

We have defined the Conscious Turing Machine (CTM) for the purpose of investigating a Theoretical Computer Science (TCS) approach to consciousness. For this, we have hewn to the TCS demand for simplicity and understandability. The CTM is consequently and intentionally a simple machine. It is not a model of the brain, though its design has greatly benefited - and continues to benefit - from cognitive neuroscience, in particular the global (neuronal) workspace theory. Although it is developed to understand consciousness, the CTM offers a thoughtful and novel guide to the creation of an Artificial General Intelligence (AGI). For example, the CTM has an enormous number of powerful processors, some with specialized expertise, others unspecialized but poised to develop an expertise. For whatever problem must be dealt with, the CTM has an excellent way to utilize those processors that have the required knowledge, ability, and time to work on the problem, even if it, the CTM, is not aware of which of the processors these may be.

21 Manuel Blum

Center for conscious AI, Computer Science, Carnegie Mellon University and UC Berkeley, United States

A theoretical computer science perspective on free will

The concept of free will is a paradox. Samuel Johnson (1709-1784) put it this way: "All science is against the freedom of the will; all experience is for it." We resolve the paradox from the perspective of Theoretical Computer Science (TCS), a branch of mathematics concerned with understanding the underlying principles of computation and complexity.

22 Wanja Wiese

Institute for Philosophy II, Ruhr University Bochum, Germany

Understanding weak and strong artificial consciousness: A perspective from the free energy principle

'Weak artificial consciousness' consists in *simulating* mechanisms associated with consciousness, whereas 'strong artificial consciousness' consists in *instantiating* consciousness. Could a sufficiently detailed computer simulation of consciousness replicate consciousness, i.e., constitute a form of strong artificial consciousness? Or will there remain a difference between simulating and being a conscious system? I argue that the free energy principle suggests (but does not entail) an account of the difference between simulating and being. In short, being a conscious system requires being a self-organising physical system, whereas simulating only requires a robust mapping between computational and physical states. I show that the free energy principle enables a precise description of this difference.

23 Aïda Elamrani

Institut Jean Nicod, Département d'Etudes Cognitives, Ecole Normale Supérieure - PSL, France

To what extent are computers conscious?

As AI improves rapidly, the question of whether Machine Consciousness (MC) and the extent to which this is possible becomes increasingly crucial. The debates surrounding the hard problem expose a stark disagreement over the possibility of a strictly mechanical implementation of consciousness among experts of this interdisciplinary field. I first give an overview of the challenges facing MC by explaining the arguments opposing phenomenalism to physicalism. Evaluating the possibility of MC requires an in-depth analysis of information, computation, and their place in nature. Information and computation have shaped several metaphysical theories. Underlying the “IT from BIT” view is the idea that reality constitutes a natural pool of information. Meanwhile, the scientific study of consciousness describes our minds as natural information processing devices. Together, these two bodies of work avail a compromise to close to gap: phenomenal experience as a virtual reality, physically implemented through layers of computational mechanisms. I show this compromise motivates further research questions between two interpretations of information: idealism and pragmatism. In practice, this analysis sets two distinct goals for MC: strong (phenomenal) MC and weak (functional) MC. Which one of them is achievable, and to what extent, depends on the interpretation of information we endorse.

24 Moritz Kriegleder

Member of AMCS, Department of Philosophy, University of Vienna, Austria

Model templates for minimal models of consciousness

The scientific study of consciousness has always been an interdisciplinary effort combining theories and tools from many fields. Minimal models provide a

shared framework for the systematic study of phenomena identifying common assumptions. With a plethora of consciousness models available, a critical analysis of overlaps and tensions becomes necessary to map out different approaches. Model templates provide a philosophical and computational tool to track how ideas travel between different theories and this approach could supplement the search for a minimal model of consciousness. In my talk, I present the model template approach and discuss how it can be used to understand the migration from physics to biology, psychology, and sociology. As a case study, I analyse the free energy principle and how it uses ideas from physics and information theory to explain consciousness. The free energy principle suggests that organisms strive to minimise their surprise or uncertainty about their internal and external states, which can be seen as a foundational principle of self-organisation and adaptive behavior. I defend a pragmatic philosophy when it comes to our use of mathematical tools to model consciousness and I discuss how free energy models fit in with current paradigms in the cognitive sciences.

25 Erik Curiel

Centre of Gravity (Bonn)/Black Hole Initiative (Harvard), Institut für Philosophie (Bonn), Universität Bonn/Harvard University, Germany

How does mathematics in a theory of consciousness represent the physical world, if at all?

The "entropic brain hypothesis" proposes that the entropy of spontaneous brain activity indexes the informational richness of conscious states. What are such information-theoretic concepts, however, supposed to represent *physically*? On the standard---almost universally accepted---picture of the relation of mathematics in a scientific theory to the world, mathematical entities represent physical entities, mathematical structures represent physical structures, and so on. The relation of representation is---again, almost universally---taken to be one of a designative, depictive or verisimilar character. Whatever the relation of the mathematical concepts of entropy to cognitive and brain states employed in this case may be, it cannot conform to any of the standard notions of representation. I argue, contrary to the standard picture, that the essential relation to study here is not between mathematics and the world but rather between our concepts and the world. Mathematics provides a wealth of different tools to use in order to bring our concepts and the world into contact (and that itself in a number of different ways), nothing more, nothing less. Some of those tools function in ways that superficially resemble standard ideas of representation; others do not.

26 Marcus Du Sautoy (Invited Special Evening Lecture)

Charles Simonyi Professor for the Public Understanding of Science, Mathematical Institute, University of Oxford, UK

AI: friend or foe?

Marcus du Sautoy is the Charles Simonyi Professor for the Public Understanding of Science at Oxford University. With a long-standing interest in

both consciousness and Artificial Intelligence, Marcus presented the BBC Horizon documentaries "The Secret You" (2009) about the science of consciousness, and "The Hunt for AI" (2012) about how close we were then to creating machines that can think like we do. In 2019 Marcus gave the public lecture "The Creativity Code: How AI is learning to write, paint and think"; see https://youtu.be/k89sS6fsZvI. Marcus now brings his take on these themes up to date with his new talk "AI: friend or foe?".

27 Claudia Passos (Invited)

Center for Bioethics, New York University

What do theories of consciousness predict about the onset of infant consciousness?

Infant consciousness raises many questions. One important question concerns the onset of infant consciousness: when does consciousness begin? I will approach this question by applying philosophical and scientific theories of consciousness to see what they predict in the infant case. When plausible theories converge on a prediction on consciousness in infants, that gives the prediction significant weight. I will focus on the most popular theories that have consequences for the problem of the distribution of consciousness: higher-order theories, global workspace, integrated information theory, recurrent processing theories. I will conclude by discussing consequences of these theories for the development of consciousness.

28 Jonathan Mason

Oxford Mathematics of Consciousness and Applications Network (OMCAN), Mathematical Institute, University of Oxford, UK

On a fundamental correspondence between consciousness and system bias: Developments in Expected Float Entropy Minimisation

The result given in this talk involves two assumptions and pits two opposing hypotheses against each other. We call the two assumptions IMPS and OGA and the two hypotheses MEEM and Anti-MEEM. IMPS stands for ‘‘Interpretive Models give Phenomena from States’’ and is the assumption that consciousness is given by system states interpreted in the context of a relational model that the system itself determines. IMPS says nothing about how a system determines the interpretive model or which model is the relevant one. OGA stands for ‘‘Object Grouping Awareness’’ and is the assumption that our awareness/experience of certain objects being strongly related to each other is a part of consciousness and, therefore, to degrade this experience partially degrades consciousness. MEEM stands for ‘‘Minimum Expected Entropy Model’’ and is the hypothesis that consciousness is the minimum expected entropy model interpretation of system states; i.e. the fundamental hypothesis of EFE minimisation. This model depends on system bias. Anti-MEEM is the hypothesis that there is something completely independent of system bias in the brain that determines the interpretive model. Under the IMPS and OGA assumptions, MEEM beats Anti-MEEM showing there is a fundamental correspondence between consciousness and system bias.

29 Christoph von der Malsburg

Frankfurt Institute for Advanced Studies and Institute for Neuroinformatics, ETH/UZH and Center for Artificial Intelligence, ZHAW Winterthur, Germany

The coherence definition of consciousness: What is coherence?

My coherence definition of consciousness (1997), which is related to B. Baars' Global Workspace Theory and G. Tononi's later Integrated Information Theory, left open a number of questions: What is the language of thought (the data structure, the neural code) of the mind, What is coherence and How does it arise within the system and with the environment? I will present answers to those questions. Central to the data structure are structured nets. Nets in different cortical areas are linked by dynamic homeomorphic mappings. The mind's state trajectory is a sequence of brain-spanning nets that are stable for a brief moment.. In a stable net all active local elements (neurons, net fragments, nets) are consistent with each other in terms of connections, each local element being predicted and supported by a minimum number of intra-areal as well as inter-areal or sensory connections. This consistency condition singles out a very sparse set of allowed global states and serves as definition of coherence. Consciousness is characterized as a succession of globally coherent states. The theory has been implemented in terms of concrete models of perception and establishes links to mathematical concepts of topology, fiber bundles and category theory.

30 Panagiota Theodoni

Philosophy, National and Kapodistrian University of Athens, Greece

A multi-regional neural network model for conscious perception

Most neural network models that have been employed to study conscious perception are non-biophysically realistic local circuit models. Nevertheless, evidence suggests that the neural correlates of conscious perception are distributed across the cortex. I present a large-scale, multi-regional, anatomically and physiologically constrained neural network model of the macaque cortex simulating Binocular Rivalry (BR)- a standard experimental paradigm for consciousness study. The model is based on a previously studied neural network model for working memory, which we have expanded to account for subjects' behavior and neural activity in a detection task. BR emerges when dichoptically viewing dissimilar images induces perceptual alternations dissociating the sensory stimulation from the conscious content. In the multi-regional model, each area is represented by a local circuit capable to reproduce the observed neural activity within one cortical area. There are 30 areas of the macaque cortex, hierarchically organized from visual to prefrontal, and wired according to anatomical inter-areal connectivity data. The model reproduces the neural encoding of conscious content along the cortical hierarchy, consistent with electrophysiological recordings. The resulting neural dynamics results from a multi-area attractor rather than on locally generated ones. Furthermore, simulated lesions and targeted intracranial electrical stimulations test existing predictions of different theories.

31 Keenan Down

Consciousness and Cognition Lab, Department of Psychology, University of Cambridge, Queen Mary University of London, United Kingdom

Logarithmic decomposition: A signed measure space for information and a qualitative tool for consciousness science

Could information be the true substrate of consciousness? Many major theories of consciousness such as Integrated Information Theory (IIT) are built upon the language of information theory. Such information-led frameworks form the backbone of a powerful experimental arsenal in consciousness science, as they provide concrete metrics for scientific investigation. Much recent work in exploring emergent informational phenomena in the brain has focussed on Partial Information Decomposition (PID), which splits up information into large, characteristic pieces depending on their informative character. Unfortunately, PID is limited by its computability, and as yet no conclusive formulation of the PID has emerged in the literature. We present results on an alternative framework called Logarithmic Decomposition (LD), a much finer decomposition which provides not only a PID with many desired properties and a refinement of the classical signed measure space for Shannon entropy, but also a unique mathematical perspective on the nature of shared and bound information. We provide an overview of this geometrically-flavoured approach to entropy, its mathematical insights into the attachment of qualia, and we explore potential experimental implementations of the logarithmic decomposition in computational neuroscience.

32 Joscha Bach

Thistledown Foundation, United States

Consciousness as a training mechanism for self organizing modelling systems

A theory of consciousness should capture its phenomenology, characterize its ontological status and extent, explain its causal structure and genesis, and describe its function. Here, I advance the notion that consciousness is best understood as an operator, in the sense of a physically implemented transition function that is acting on a representational substrate and controls its temporal evolution, and as such has no identity as an object or thing, but (like software running on a digital computer) it can be characterized as a law. Starting from the observation that biological information processing in multicellular substrates is based on self organization, I explore the conjecture that the functionality of consciousness represents the simplest algorithm that is discoverable by such substrates, and can impose function approximation via increasing representational coherence. I describe some properties of this operator, both with the goal of recovering the phenomenology of consciousness, and to get closer to a specification that would allow recreating it in computational simulations.

33 Robert Prentner (Invited)

Munich Center for Mathematical Philosophy, Ludwig, Maximillian University of Munich, Germany

Mathematizing phenomenology within a process framework

Phenomenology is one of our most comprehensive resources for systematically analysing conscious (first-personal) experiences. For example, phenomenologists typically emphasise the correlation between subjective and objective aspects: a fundamental "self/world" structure that is constitutive of any conscious experience. In addition, we demonstrate that such ideas can be made more precise using mathematics, thus contributing to mathematical consciousness science.

Yet, the question of where to conceptually locate phenomenology is still open. Is it a form of (transcendental) idealism? Or is it a high-level account of physical (brain) processes? We will argue instead that a non-dual framework, which presupposes neither mental nor material primitives, is best suited to ground phenomenological studies. A candidate is provided by process philosophy with its concept of global relatedness. We briefly review the process framework and establish the connection with phenomenology.

34 Chetan Prakash

Association for Mathematical Consciousness Studies, USA

A conscious agent model and the physical world

Recent scientific work questions the common consensus of what is fundamental in nature. The interface theory of perception, e.g., indicates that our perceptions could not have evolved to reflect the true nature of the world. Experimentalists reveal that proton structure appears as an artifact of perceptual resolution: At slow shutter speeds protons appear as triples of valence quarks, bound by gluons; at high speeds they appear as just gluon seas. Dramatic changes also appear with spatial resolution. Again, theoreticians demonstrate that spacetime is not fundamental, finding deeper structures, such as amplituhedrons, to explain scattering amplitudes. These deeper structures are static, with no dynamical explanation so far. Under the hypothesis that it is consciousness that is fundamental, we outline a program aiming to explain such results. A precise definition of conscious agent networks leads to a Markovian perceptual dynamics. We propose that apprehending this dynamics by a forgetful process, involving coarse graining, tracing and sampling finite dynamical windows, can lead to a "physical projection," whereby perceptions recover the aforementioned results. The aim of this program is to ground science in the dynamics of conscious agents. This reports recent work with Donald Hoffman, Swapan Chattopadhyay, Robert Prentner and others.

35 Pedro Resende

Centre for Mathematical Analysis, Geometry and Dynamical Systems, Department of Mathematics, Instituto Superior Técnico, Portugal

The emergence of geometric worldviews in qualia space

In the model of qualia based on measurement spaces the points are the qualia and the open sets are the pure concepts, which are the communicable finite chunks of classical information. Moreover, algebraic structure conveys subjective time and the subjective experience of logical abstraction. In this talk I discuss the subspace of qualia S associated to any organism which, such as a human being, (1) possesses cognitive capabilities that allow it to compute on intinite quantities such as real numbers by finite approximations, and (2) has subjective experience associated to its pure concepts. I argue that if (3) such an organism is "macroscopic," thus not having direct awareness of quantum mechanical interference effects, then S is isomorphic to the topology of another space X, of "points," or "states," which is second-countable and locally compact Hausdorff. This suggests an explanation of the apparent fact that humans possess a geometric view of the world, and also suggests that the same is likely to happen to any organism that shares the general features (1), (2) and (3).

36 Quanlong Wang

Quantinuum, UK

Perceiving division and perceived division modelled in ZXW calculus

Any cognitive process is composed of two divisions: the perceiving division and the perceived division. In this talk, we will show how to use ZXW calculus, a graphical language for quantum computing, as a mathematical framework for describing the interactions between the perceiving division and the perceived division of a general cognitive process. This approach could potentially provide a new way of understanding the formation and properties of quaila.

37 Alex Kostova

Sofia University "St. Kliment Ohridski", Bulgaria

Formalizing panpsychism through Integrated Information Theory

Panpsychism posits the perplexing primacy of consciousness considering it a pervasive feature that permeates the fabric of the universe and extends beyond human consciousness. Panpsychism has undergone a notable resurgence in the past few decades that can be attributed to the persistently insurrmountable challenge that despite the advancements in comprehending the intricacies of the brain and its interplay with conscious states, the fundamental question of how consciousness arises from ostensibly non-conscious elements of the physical realm remains unresolved. For instance, the solution of Sautrantika focuses on the experiential complexity of mental phenomena prioritizing their phenomenological accessability, rather than delving into their essence. In this paper I will investigate the possibility of formalizing panpsychism within the IIT framework that could capture the idea of the panpsychist systems that

emergence of consciousness within the physical realm would have been implausible, if not inconceivable, without the preexisting presence of the mental fabric from the very outset of existence.

38 Hye Young Kim

Ecole Normale Superieure, France

Geometrical imagination of shared consciousness: Intersubjective time-consciousness, space, and Euler's formula

If we are an individual with subjective consciousness, each subject has consciousness that is 'mine'. Each subject's consciousness of time is 'mine' as well, which we could call 'temporality' of each subject that is distinguished from 'objective time'. Objective time is the measurable time that we can count and name. When different subjects interact, how does a subjective consciousness interact with the other that is not mine? In the act of conversation, for example, for the interaction to occur, the conversant have to 'share' time and space. By 'shared time and space' one can talk with somebody only when that person is in the space where one is, so that both can hear and see each other. And as I speak and listen, 'my' temporality is created in my action. But in the case of conversation, 'my' temporality is somehow corelated with the other subject's temporality. My question is how to 'picture' this situation where two individual conscious subjects go over the boundaries of 'my' subjective consciousness. I take Euler's formula as a form that represents the structure of conversation. The 'i' in Euler's formula could be interpreted as space out of time where the possibility of shared time lies.

39 Maja Spener (Invited)

Department of Philosophy, University of Birmingham, UK

Are subjective measures of consciousness reliable?

Subjective measures of consciousness face a well-known, seemingly intractable challenge: they use introspection to gain access to conscious data, but introspection is unreliable. In this talk, I outline a framework for evaluating subjective measures of consciousness, based on important but neglected distinctions between different types of introspection. I then show how to apply this framework to evaluate a mainstream subjective measure of consciousness, the Perceptual Awareness Scale.

40 Renee Ye

Ruhr Universität Bochum, Germany

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.

41 Stephen Fleming

CIFAR Fellow in Brain, Mind and Consciousness, Department of Experimental Psychology, University College London, United Kingdom

Awareness as inference in a higher-order state space

I will describe a computational framework that characterizes awareness as a metacognitive inference in perceptual generative models. In this “higher-order state space” account, internal states supporting awareness judgments are low-dimensional and asymmetric: a vastly greater range of perceptual content is nested under the latent state of “being aware” than the state of “being unaware”. We have developed a novel visual perception paradigm that probes such inferences by orthogonally manipulating expectations about stimulus content (discrimination) and awareness of content (detection). By combining a no-report version of our task with functional neuroimaging we reveal a neural dissociation between prediction errors (PEs) on content and awareness of content: content PEs are tracked in posterior sensory cortex while awareness PEs are tracked in prefrontal cortex. Furthermore, because inferences on awareness can be driven either by external input (perception) and/or changes in internal states (imagery), this framework naturally accommodates recent findings of failures to distinguish reality from imagination near perceptual threshold. Together, our results reveal a hierarchical structure supporting visual detection and discrimination and are consistent with conscious awareness reflecting a higher-order inference within perceptual generative models.

42 Kenneth Shinozuka

Centre for Eudaimonia and Human Flourishing, University of Oxford, UK

Meta-analysis of the pharmacology, neuroimaging, and phenomenology of psychedelics: implications for consciousness science

Serotonergic psychedelics induce altered states of consciousness and have shown potential in treating a variety of neuropsychiatric disorders including depression and addiction. Yet, their modes of action are currently not fully understood. Here we provide a systematic review of the tripartite hierarchy of levels underlying the effects of psychedelics, from their molecular

pharmacology to the associated whole-brain activity and subjective experience (phenomenology). As part of this, we report the first comprehensive, systematic meta-analysis of functional neuroimaging studies on psychedelics based on the published results. Our systematic meta-analysis links each of the classic psychedelics (psilocybin, LSD and ayahuasca) to specific fingerprints at each level of this tripartite hierarchy. The results show a highly non-linear relationship between these fingerprints from pharmacology, neuroimaging, and phenomenology. Additionally, the neuroimaging analysis reveals surprising findings such as increased connectivity within the default mode network on psychedelics, apparently contradicting a major theme of the literature. Overall, the results point to the need for standardising reporting tools and for more research on the emergence between different levels of psychedelic effects. Future work could use our tripartite hierarchy framework to investigate the dose -dependent effects of different psychedelics in the same participants.

43 Ouri Wolfson

Computer Science, University of Illinois, Chicago, United States

Consciousness as a form of coordination

We present a hypothesis that consciousness is related to traffic of signals in the brain; specifically, that consciousness is produced by a type of coordination that is required to move a traffic system from user equilibrium to system optimum. The hypothesis has been verified by datamining of connectomes. Our datamining approach considers the brain regions and the tracts that connect them as a road network, and the signals traveling between them as traffic. We analyze travel patterns by a process called traffic assignment. The results are unexpected in the sense that the movement of signals in the brain seems to follow some global optimization patterns as opposed to the anarchical system that would be favored by evolution. The hypothesis can be further examined by datamining the connectomes of human and animal subjects. Thus, our results can be viewed as a neural correlate of consciousness. However, the results may be deeper in the sense that coordination may be the mechanism that actually produces the subjective experience. In other words, is subjective experience the way the ability to coordinate brain activities manifests itself? This is supported by the fact that coordination is a form of control or agency.

44 Malcolm Lowe

Independent Researcher, United States

Name, number and the foundation of consciousness

In his book Aion: Researches into the Phenomenology of the Self, renowned Swiss psychoanalyst and psychiatrist, Carl Jung, observed that “Consciousness presupposes a differentiation into subject and object and a relation between them. Where there is no ‘other’, or it does not yet exist, all possibility of consciousness ceases.” Most models of consciousness look to physical explanations for the emergence of consciousness. In this paper, I present a model of consciousness that identifies natural languages, or more

precisely the inner domain of Meaning in natural languages, as the sole source of this differentiation between subject and object and their ensuing relation. I will share how my investigation into how meaning is instantiated in words in English led to the discovery that sound-meaning relations in languages represent an ontologically distinct domain grounded in number; a domain that includes, among other architectural features, a design feature that operates to separate Self (What is Me) from Other (What is Not Me). The significance of this discovery is not simply that it explains the logic of consciousness, or that it offers insight into the inner architecture of thought, but rather that it recognizes a vast terra incognita ripe for further research and modelling.

45 Peter Coppola

Cognition and Consciousness Imaging Group, Clinical Neurosciences, Division of Anaesthesia, School of Clinical Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK

Intrinsic dynamics: approach to map neuronal to subjective dynamics.
Typical consciousness can be defined as an individual-specific stream of experiences. Modern research on dynamic functional connectivity uses clustering techniques to create common bases on which to compare different individuals. We propose an alternative methodology by combining modern theories of consciousness and insights arising from phenomenology and dynamical systems theory. Just as any neural state must be understood within its own systemic and dynamic context, we postulate that the intra-individual context of any subjective state is foundational to its value. Therefore we describe each temporally-specific connectivity state by its similarity to all other available past and future states. This approach enables a representation of connectivity dynamics in an intrinsically-defined, individual-specific temporal landscape. By investigating the properties of this landscape in humans in different states of consciousness, we show that consciousness is associated with short term transitions that are less predictable, quicker, but, on average, more constant (i.e., less acceleration/deceleration). We also show that temporally-specific connectivity states are less easily describable by network patterns that are distant in time, suggesting a richer and wider space of possible states. Finally. we empirically show how this approach enables a principled and empirically tractable point of contact between brain and phenomenological dynamics.

46 Paul Azzopardi (Invited)

Experimental Psychology, University of Oxford, UK

Thresholds, noise and response bias: Insights from blindsight
Perceptual decisions that underly detection and discrimination are inherently susceptible to internal noise, and in consequence there is no specific point on a psychometric function that corresponds to an objectively-defined threshold. Empirically-determined thresholds are the combined outcome of performance criteria arbitrarily defined by experimenters, and observers' variable internal thresholds or response criteria. This bears on the interpretation of behavioural

responses that are taken as measures of awareness. Blindsight is a phenomenon in which patients with visual field defects caused by damage to primary visual cortex are able to detect and discriminate unseen stimuli when forced-choice procedures are used, implying a sharp dissociation between visual performance and visual awareness. It is tempting, therefore, to think of the primary visual cortex as module that confers awareness of information processed by parallel circuits dedicated to vision. An alternative interpretation, however, is that the dissociation between performance and awareness in blindsight arises as the result of an abnormally conservative and unstable response criterion underlying the responses taken as measures of awareness. While this may not explain the 'hard problem' of consciousness, it highlights the fact that processes for setting and actively managing appropriate response criteria are crucial for maintaining the correspondence between performance and awareness that characterise conscious behaviour.

47 Johannes Kleiner

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

A structural turn in consciousness studies

A number of publications and recent activities in various different fields point to what could be early signs of a structural turn in consciousness studies, where verbal descriptions or simple formalisations of conscious experiences are replaced by structural tools, most notably mathematical spaces. My goal in this talk is to offer three comments which I think are necessary to avoid misunderstandings in these developments. These comments concern the relation between structural and structuralist agendas, overlooked assumptions in regards to isomorphisms and the question of what structure to consider on the side of consciousness in the first place. I will also explain what in my opinion is the great promise of a structuralist methodology and how it might impact consciousness science at large.

48 Vladimir Aksyuk

Microsystems and Nanotechnology Division, Physical Measurement Laboratory, National Institute of Standards and Technology, USA

Consciousness is learning: What it is like to be an online compositional learning system

Humans excel at data-efficient compositional learning not currently attainable by machine learning approaches. Compositional learning entails sparsely-structured perception and action encoding via separable time-persistent units. Approximate Bayesian inference within such hypothesis spaces is already implemented in predictive coding, however the new sparse causal structure learning is discontinuous, not possible smooth weight updates. Here we note that online learning of such sparsely-structured, compositional and time-persistent causal units entails a type of perceptual bottleneck similar to the one separating conscious from unconscious perceptions. We hypothesize that

conscious contents are precisely the new causal units that are being learned at a given time, and propose a functional model combining compositional learning -by-feature-binging with predictive processing. We argue that such learning forms associatively-recallable declarative memories that are unified yet differentiated. Furthermore, it immediately makes the newly-learned structures globally available by causally connecting perceptions and actions across disparate domains. We extend this basic model by adding reinforcement learning of action policies maximizing survival and reproduction, accounting for valence as the perceptual estimate of the present state's value. We argue that the major subjective and objective observations describing consciousness may be accounted for by a neurobiologically-plausible learning system functional architecture that is mathematically defined and testable.

49 Javier Galadí

Center for Brain and Cognition, Pompeu Fabra University, Barcelona, Spain

The relativity of the duality

My talk is divided into two parts. First, I argue that the tradition of the mind-body problem hides from us the true meanings of the mental and the physical. Then, I propose the mathematical formalism associated with the corrected approach to the problem of consciousness. In the known solutions of the mind-body problem, the traditional concepts of the mental and the physical are accepted. This causes the philosophical problem to be unsolvable and induces the mathematical models of consciousness to be ill-posed. The tradition assumes that the concepts of the mental and the physical are absolute: what is physical (mental) is physical (mental) regardless of who asserts it. Here I argue that the concepts are relative to the observer in the same way that space and time are relative to the observer in Einstein's theory of relativity. Just as there is no such thing as absolute time and space, in my mathematical formulation the third person point of view is eliminated. Instead, I posit equations relating the first-person viewpoints of different observers, including their respective relative concepts of the mental and the physical. Finally, I formalize the set of transformations associated with observer changes.

50 Diana Stanciu

Romanian Young Academy, Romania

Mathematical consciousness and also mathematical collective intentionality?

While looking at Brian Epstein's categories of 'anchoring', 'grounding-anchoring', and 'anchoring schemas' in discussing social goups' pluralistic construction, I consider collective intentionality as related to conscious agency and decision-making, but somewhat differently from previous research. I am particularly interested in Epstein's understanding of group properties' emergence and his criticism of various models in social ontology: network theories, equilibrium theories, theories of complex systems, computer simulations. If, in relation to these, we also consider the paradigm of 4E cognition (embodied/ embedded/ enacted/ extended) in cognitive science and

the philosophy of mind, asserting the constitutive and causal role of body and environment in both cognition and social interaction, we end up with a very complex methodology. And both quantitative and qualitative data obtained through this complex methodology are generally interpretated through statistical methods and algorithms are extracted. All this information is thus mathematically expressed. My question is though whether this epistemologically necessary mathematical expression refers to facts and values that are also ontologically mathematical or rather ontologically non-mathematical and thus expressed just out of epistemological concern or necessity. In other words, if we can consider consciousness mathematical, can we extend that also to collective intentionality or this is rather an 'ant trap'?

51 Serena Doria

University of Chieti-Pescara, Department of Engineering and Geology, University of Chieti-Pescara, Italy

Conditional probability models based on fractal measures to represent the awareness process

In the subjective approach to probability, conditional probability and its extensions represent the level of partial knowledge we have about a phenomenon based on partial information. A new model of coherent upper and lower conditional probabilities defined by Hausdorff outer and inner measures has been proposed to represent respectively the unconscious and conscious activities of the human brain. In the model uncertainty measures are defined according to the complexity of the conditioning event that represents a piece of information. According to the model the partial knowledge is updated when the conditioning event has positive and finite Hausdorff measure in its Hausdorff dimension. In that case, if the Hausdorff dimension s of the conditioning event is less than the Hausdorff dimension t of Omega, the conditioning event is an event with zero probability with respect to the prior probability; the new Hausdorff outer measure is considered to define conditional probability and to update partial knowledge. According to the model the unexpected events, represented by sets with zero probability, are those which really update the knowledge. The awareness process of human beings is described by a class of coherent conditional defined by different fractal measures

Lightning talk titles and abstracts

Brent Allsop

Theories of Consciousness, Consciousness Consensus Project, Canonizer.com LLC, United States

Physicists don't understand color qualities

If you put a red/green signal inverter anywhere in the chain of perception, the strawberry will seem to have a greenness quality. This proves that the greenness quality is not a quality of the strawberry, it is a subjective quality of our conscious knowledge of the strawberry. The question is, what is responsible for that subjective greenness quality? Additionally, an objective description of the behavior of a subjective quality would tell us nothing about what it is like. This is similar to the way text alone can't tell you what redness is like. In reality, experimentalists are objectively describing everything with text. They just aren't telling us which of all those descriptions is a description of subjective redness. Discovering and demonstrating what it is that has a subjective redness quality, even if that is some mathematical function that is responsible for redness, that would objectively ground the definition of words like redness and bridge the explanatory gap. An example of a sufficiently grounded statement would be: "My redness is like your greenness, both of which we call red." Experimentally demonstrating what subjective redness is would prove consciousness isn't a 'hard problem', it is just a color quality problem.

Bryant Johnson

Independent researcher, United Kingdom

Consciousness - A standard reference model

A model of existence including consciousness is described, based on the totality of records created by human actions, thoughts and speech. It replaces confused and contradictory opinions of individuals or groups of these, however large, with a coherent description matched to the knowledge and belief of the human race as a whole. It relates the inanimate to the animate and the conscious, and reality (the objective) to belief and imagination (the subjective). It explains the relation between order, disorder and the random state. It accounts for the world of objects, the conscious world and the world of ideas, the nature of intelligence both human and artificial and the entire range of human philosophy, characteristics and activities providing answers to questions that individuals continually pose. In mathematical terms it is based on the Central Limit Theorem and the Argand diagram but in addition allows for both unfolding (explicit local) and enfolding (implicit non local) dimensional processes which are associated with the principles of wave motion, the hologram and the fractal. The rise of consciousness from the inanimate is then

associated with the increasing asymmetry (irreflexivity) imposed by the conflict between these unfolding and enfolding processes.

Nikolaos Tzagkarakis

SiSaf Ltd, United Kingdom

Energy limitation as the precondition for modelling the self

In our paper we consider how subjective experience distinguishes self from environment. We approach the question from an AI perspective, asking what cognitive mechanisms would enable artificial agents to generate self-reports similar to those of their biological counterparts. We argue that the key mechanisms include attention, environmental modelling, and memory, all operating under the key constraint of energy limitations. Attention filters environmental stimuli through internal goals/costs and restricts the range of environmental models the agent can generate. This energy limit leads to the narrowed correlation of goals/costs with specific models of self as separated from the rest of the environment. Memory plays an important role in the process, enabling the current model to be revised in the light of changing environmental conditions and agential needs. Unlike more traditional accounts, ours sees the self/world separation as largely a function of energy limitations; the energy capacities of an agent establish a threshold that defines the boundaries of the self as one of a group of models the agent constructs. The larger the energy resources, the larger the self.

Dian Lu

School of Medicine, Stanford University, USA

Exploring the solution space of conscious states using nonlinear dynamical analysis

Recent computational work proved it feasible to simulate the brain's intrinsic connectivity patterns (ICNs) by utilizing whole-brain structural connectivity and neural mass models, both of which are empirically grounded. This large-scale model is highly dimensional and numerally intractable, but we can gain insights by (1) changing the parameters in a physiologically meaningful way and observe how the system collapses; (2) analytically exploring the local two-dimensional dynamical system before adding global connectivity. In this study we simulated the propofol-induced unconscious state by increasing the time constant of the inhibitory neural population's dynamics (as propofol is GABAergic). By exploring the bifurcation diagram, we can see that the solutions mostly fall in the parameter set which generates a mixed patch of stable and unstable fixed points, both oscillatory. As the simulated propofol dosage gets higher, the set shrinks, suggesting that the safety zone allowing for a recovery excursion gets narrowed, until non-existent. The non-linear dynamical analysis can help sketching the scope of the dynamical reservoir that the brain is endowed with, but it is limited in the reliance on a complete knowledge of the dynamical equations. Great ideas are needed to bridge the two fields in a meaningful and fruitful way.

Aleksandr Groznykh

Independent researcher, Germany

Neural Cellular Automata

NCA is a differentiable self-organized system composed of a set of individual agents, each executing a local rule to achieve a global objective. Every agent recurrently operates under the same rule, enabling cells to acquire distributed, local algorithms with minimal parameters. This system demonstrates exceptional versatility in solving various tasks, encompassing feedback control and generative modeling. The presentation will include a diverse array of live demos, showcasing the practical applications and capabilities of NCA. These demonstrations serve to illustrate the efficacy of the system and its potential contributions to the field of cognitive modeling.

Uziel Awret

Inspire Institute, United States

Philosophical advantages of 'Suskind Sphere' scenario

The purpose of this short presentation is to argue that the AdS/CFT revolution (especially its surprising relevance to Condensed Matter Theory) can provide us with unique philosophical advantages that we cannot get from either QM and QFT. While the proposed mechanism is strange it provides enough philosophical advantages to merit consideration. I will begin with what I call the 'Susskind's Sphere' scenario (from "Dear Qbitzers ER=EPR") in which a massive two-di spherical quantum computer shell is governed by conformal QFT. Susskind argues that the dual AdS space is not the lab space and can only be accessed by uploading a 'technician' onto the quantum computer shell. Here I will argue for a more conservative scenario and explore the possibility that there exist physical correlates of consciousness that are governed by a conformal QFT (and harboring massive entanglement). Relating consciousness to their AdS dual space has many philosophical advantages: A way of placing the 'space of consciousness' 'in' ordinary space; Solving the Meta Problem of consciousness (meta-correlational); Protecting physicalism from the conceivability argument; Providing geometry and geometric meaning for free; Minimizes metaphysical commitments; Consciousness is constituted like an AdS space as a robust medium in.

Yair Pinto

Brain and Cognition, Psychology, University of Amsterdam, Netherlands

Agent-causal libertarian free will in the light of the halting problem

In this talk I will outline that the "prediction paradox" can be extended to the "algorithm dodging paradox". The prediction paradox highlights that iff we consider voluntary human behavior undodgeable predictions seem impossible. E.g. if a super-scientist informs you that "you must raise your hand 1 minute from now", then you can dodge (i.e. act as to falsify) this prediction. The prediction paradox has a straightforward solution. Determinism can assert that you are a "must dodge" robot, i.e. when you receive A as input, you must do B,

and vice versa. Here I outline that we can expand from predictions to must-dodge algorithms. Roughly put, there are only two "dodge"-algorithms. Either "you must dodge everything, except this algorithm", or "you must dodge everything, including this algorithm". The first option can be dodged by complying with subsequent input. The second option is the most interesting one. I will outline that this algorithm with itself as input implies an infinite loop - thus this algorithm can be dodged by doing anything after learning of it. Interestingly, this "must dodge all" algorithm seems closely related to the halting problem. Finally, I sketch how the current reasoning can be experimentally tested.

Andy E. Williams

Nobeah Foundation, Kenya

The necessity of Human-Centric Functional Modeling in converging on a predictive model for consciousness

Human-Centric Functional Modelling (HCFM) is a technique that represents consciousness as navigating a functional state space that consists of all possible functional states consciousness might occupy, along with all possible processes through which it might navigate from one functional state to another. Because this functional state space does not make any assumptions about the structures through which those functions are implemented, it can hypothetically represent the functionality of any possible model of consciousness. Comparing any two or more models of consciousness requires that those models have representations of the same functionality, and it requires that observations of the functionality of consciousness are modelled in a common way, so that any predicted functionality might be compared with any observations. Any two or more models of consciousness cannot be compared in order to converge on a better common understanding where this is not the case. If any given model is not a complete model of all possible functions of consciousness, then that model is not completely comparable and cannot be made to converge to a more correct model where it is incomplete. HCFM provides a method for modelling consciousness that can potentially be generalized to apply to any model.

Discussion sessions

Next to talks by invited speakers and participants, this conference will feature discussion sessions in parallel groups. The goal is to create an open and friendly atmosphere in which thoughts and ideas can be exchanged.

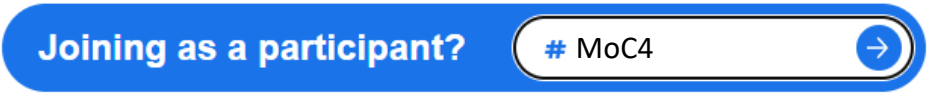
Each discussion session is devoted to one topic, either a specific question/idea related to models of consciousness, or a general question concerning progress and visions of the field as a whole. In addition to various topics which have been proposed in advance by the organizers and the advisory board, new topics can be added during the conference in response to talks or based on general interest.

In order to let those questions emerge that receive the most interest by all participants, a web interface will be used, which allows attendees to indicate their interest in a question, as well as to propose new questions of their own. 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

#MoC4



for “Models of Consciousness 4”.

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 symbol.
- In order to add a new discussion topic, simply enter the question into the field at the top. (Please word questions carefully. If questions are not worded well, they might be blocked by moderators.)

The web interface is accessible also in advance of the conference.

Discussions are an integral component of science and we hope that the sessions held during this 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 following two quotes by Isaac Asimov and a collaborator of Werner 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

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

NOTES

Association for Mathematical Consciousness Science

Mathematical Institute
University of Oxford
Andrew Wiles Building
Radcliffe Observatory Quarter (550)
Woodstock Road
Oxford
OX2 6GG