



# Programme

## Philosophy and Theory of Artificial Intelligence

St Antony's College, 62 Woodstock Road, Oxford, OX2 6JF, UK

**Saturday, 21.09.2013**

**Registration I - Nissan Lecture Theatre, Foyer**

**9:00 Introduction**

**9:00-10:30 Keynote: Stuart J Russell (UC Berkeley)**

"Rationality and Intelligence"

(Nissan Lecture Theatre) Chair: Müller

ABSTRACT: The notion of bounded optimality has been proposed as a replacement for perfect rationality as a theoretical foundation for AI. I will review the motivation for this concept, including similar ideas from other fields, and describe some research undertaken within this paradigm to address the problems faced by intelligent agents in making complex decisions over long time scales.

**10:30-11:00 Coffee break (Buttery) & Registration II**

**11:00-13:30 Sections (5 x 2)**

	Session A: <b>Computing</b> (Nissan Lecture Theatre) Chair: Bonsignorio	Session B: <b>Cognitive Science</b> (Pavilion Room, Gateway Building) Chair: Votsis
11:00 11:30	<b><u>Marcin Miłkowski</u></b> <u>Computation and Multiple Realizability</u>	<b><u>Aaron Sloman</u></b> <u>Why is it so hard to make human-like AI mathematicians?</u>

### Computation and Multiple Realizability

ABSTRACT: Multiple realizability (MR) is traditionally conceived as the feature of computational systems, and has been used to argue for irreducibility of higher-level theories. I will show that there are several ways a computational system may display MR, and none of them is particularly helpful in arguing for irreducibility. These ways correspond to (at least) three ways one can conceive of the function of the physical computational system. I will conclude that for this reason, MR is of no importance for computationalism, and argue that it should rather appeal to organizational invariance or substrate neutrality of computation.

### Why is it so hard to make human-like AI mathematicians?

ABSTRACT: I originally got involved in AI many years ago, not to build new useful machines, nor to build working models to test theories in psychology or neuroscience, but with the aim of addressing philosophical disagreements between Hume and Kant about mathematical knowledge, in particular Kant's claim that mathematical knowledge is both non-empirical (apriori, but not innate) and non-trivial (synthetic, not analytic) and also concerns necessary (non-contingent) truths. I thought a "baby robot" with innate but extendable competences could explore and learn about its environment in a manner similar to many animals, and learn the sorts of things that might have led ancient humans to discover Euclidean geometry. The details of the mechanisms and how they relate to claims by Hume, Kant, and



other philosophers of mathematics, could help us expand the space of philosophical theories in a deep new way. Decades later, despite staggering advances in automated theorem proving concerned with logic, algebra, arithmetic, properties of computer programs, and other topics, computers still lack human abilities to think geometrically, despite advances in graphical systems used in game engines and scientific and engineering simulations. (What those do can't be done by human brains.) I'll offer a diagnosis of the problem and suggest a way to make progress, illuminating some unobvious achievements of biological evolution.

11:30	<b><u>Tarek Richard Besold</u> and <u>Robert Robere</u></b>	<b><u>Mark Bickhard</u></b>
12:00	<u>When Thinking Never Comes to a Halt: Tractability, Kernelization and Approximability in AI</u>	<u>The Predictive Brain: A Critique</u>

**When Thinking Never Comes to a Halt: Tractability, Kernelization and Approximability in AI**

ABSTRACT: The recognition that human minds/brains are finite systems with limited resources for computation has led researchers in cognitive science to advance the Tractable Cognition thesis: Human cognitive capacities are constrained by computational tractability. As also artificial intelligence (AI) in its attempt to recreate intelligence and capacities inspired by the human mind is dealing with finite systems, transferring this thesis and adapting it accordingly may give rise to insights that can help in progressing towards meeting the goals of AI. We therefore develop the "Tractable Artificial and General Intelligence Thesis" by applying notions from parametrized complexity theory and approximation theory to a general AI framework, also showing connections to recent developments within cognitive science and to long-known results from cognitive psychology.

**The Predictive Brain: A Critique**

ABSTRACT: The Predictive Brain is a general term for a family of related approaches to modeling perceptual and cognitive processes in the brain. There have been several additions and elaborations to a basic initial framework over the last several decades resulting in a complex and sophisticated set of models. I will argue, nevertheless, that the initial framework for these developments is flawed, and that the recent additions have compounded those initial problems. I will address several progressive elaborations that have been made, though there is not time to do a complete critique of every variant. In particular, the predictive brain models have developed within a general Helmholtzian framework of modeling perception in terms of inference from input sensations to representations of the world.

12:00	<b><u>Carlos Eduardo Brito</u> and <u>Victor X. Marques</u></b>	<b><u>Stefano Franchi</u></b>
12:30	<u>Computation in the Enactive Paradigm for Cognitive Science and AI</u>	<u>General homeostasis as a challenge to autonomy</u>

**Computation in the Enactive Paradigm for Cognitive Science and AI**

ABSTRACT: In this paper, we attempt to reconcile the so far antagonist positions that view the organism, respectively, as a self-maintaining dynamical system and as an information processing system. For this purpose, we incorporate the notion of computation into the enactive paradigm for the cognitive sciences, introducing the notion of informational cause and making use of a naturalized account of functions. We also investigate some consequences associated with the relation between cognition and computation.

**General homeostasis as a challenge to autonomy**

ABSTRACT: The paper argues that the conception of life as generalized homeostasis developed by W.R. Ashby in Design for a Brain and his other writings is orthogonal to the traditional distinction between autonomy and heteronomy that underlies much recent work in cellular biology, (evolutionary) robotics, ALife, and general AI. The philosophical and technical viability of the general homeostasis thesis Ashby advocated, the paper argues, can be assessed through the construction of virtual cognitive agents (i.e. simulated robots) mimicking the architecture of Ashby's original homeostat and his subsequent Dams device.



12:30	<b><u>Blay Whitby</u></b>	<b><u>Cem Bozsahin</u></b>
13:00	<u>Computers, Semantics, and Arbitrariness</u>	<u>Natural Recursion doesn't work that way</u>

**Computers, Semantics, and Arbitrariness**

ABSTRACT: It has become a familiar dogma to claim that computers have syntax but no semantics. This is not only a rather unusual claim to make about any logical system it also may hide certain metaphysical confusions. It is a claim that deserves deeper analysis. Under such analysis it is reasonable to claim that suitably embodied and enactive computational systems can be described as having both semantics and syntax in the same sense as humans and animals. A conceptual clarification is attempted showing exactly how and when computers can be described as containing full semantics.

**Natural Recursion doesn't work that way**

ABSTRACT: All hierarchically organized observed behaviors are instances of recursion by value. Recursion by name can be shown to be more powerful than recursion by value: the former has infinite types, the latter does not. Any animal that can plan has recursion, so that \emph{that} kind of recursion is probably not unique to humans. Humans appear to have a certain kind of recursion which is unique, the kind that works with embedded push-down automaton, and that is probably not unique to language. Therefore it is unhelpful to build entire conception of recursion in natural language and humans on a much powerful notion of recursion than needed, and on purely syntactic terms rather than conceptual or semantic, including theories about its evolution.

13:00	<b><u>David Leslie</u></b>	<b><u>Yoshihiro Maruyama</u></b>
13:30	<u>The Lures of Imitation and the Limitations of Practice: "Machine Intelligence" and the Ethical Grammar of Computability</u>	<u>AI, Quantum Information, and Semantic Realism: On the Edge between Geometric and Algebraic Intelligence</u>

**The Lures of Imitation and the Limitations of Practice: "Machine Intelligence" and the Ethical Grammar of Computability**

ABSTRACT: Since the publication of Alan Turing's famous set of papers on "machine intelligence" over six decades ago, questions about whether complex mechanical systems can partake in intelligent cognitive processes have largely been answered under the analytical rubric of their capacity successfully to simulate symbol-mongering human behavior. While this focus on the mimetic potential of computers in response to the question "Can machines think?" has come to be accepted as one of the great bequests of Turing's reflections on the nature of artificial intelligence, I argue in this paper that the fraught legacy of this inheritance reveals a deeper conceptual ambiguity at the core of his seminal work on effective calculability. In teasing out the full implications of this ambiguity, I endeavor to show how certain underlying pragmatic and normative textures of Turing's 1936 calculability argument ultimately call into question the very idea of "machine intelligence" he later underwrites.

**AI, Quantum Information, and Semantic Realism: On the Edge between Geometric and Algebraic Intelligence**

ABSTRACT: Searle contrived two arguments on AI: the Chinese room and the one based upon the observer-relativity of computation. I aim at elucidating implications of his arguments for the quantum informational view of the universe as advocated by David Deutsch and Seth Lloyd. I argue that Searle's argument and the paradox of infinite regress yield critical challenges to the quantum view, leading us to the concept of weak and strong information physics. After looking at Wheeler's anti-realist position, I discuss Dummett's anti-realist theory of meaning and proof-theoretic semantics, arguing that his "manifestation argument" commits Searle's idea of intentionality to semantics realism. I attempt to articulate tensions between realist and anti-realist ideas on semantics, by drawing a distinction between geometric and algebraic intelligence, analogous to Cassirer's dichotomy between substance and function.

13:30-15:00 **Poster Session** & **Lunch** (Buttery)



Barkati & Rousseaux · Bello · Bianchini · Boltuć · Bonsignorio · Dewhurst · Freed · Gaudl · Hempinstall · Hodges · Laukyte

Novikova, Gaudl, Bryson & Watts · Schroeder · Shieber · Smith · Toivakainen · Toy · Vadnais · Weber · Vosgerau

## Poster session details

1. **Karim Barkati and Francis Rousseaux.** How the game apparatus (en)lighten the intelligence definition problem: Turing gesture and Caillois framework at work

Define intelligence has proved an impossible problem to settle when approached in a frontal manner. In this regard, Alan Turing had a non-trivial epistemic gesture: instead of trying to answer directly his question "can machines think?", he designed the Imitation Game, opening a wit way to approaching what "thinking" is by using a game apparatus. Opportunely, Roger Caillois' book "Man, Play and Games" provides a rich framework to explore and extend Turing's gesture: the Paidia versus Ludus two opposite tendencies, and the Separate, Regulated, Uncertain, Unproductive, Fictitious and Free six canonical characteristics. Using Caillois' framework, we systematically raise and examine questions from his canonical game qualities to reassess Turing's Imitation Game and try to differently tighten the intelligence definition.

2. **Paul Bello.** Dispositional Accounts of Implicit Believing

I suggest that belief ascription by mental simulation (Goldman 2006; Davies & Stone 1995) offers a more natural way to characterize the tacit knowledge of other agents, and of one's own implicit beliefs. Given the structure of the statement "If I were x in situation s, I'd y," I take belief ascription to be intertwined with counterfactual reasoning, capturing the idea that implicit beliefs are of a piece with dispositions; that is, they (1) are manifested in non-actual states of affairs and (2) retain the ability to support prediction and explanation. I apply this analysis to the distinction between implicit and explicit beliefs, giving an argument from cognitive plausibility to prefer a mostly-dispositional account of believing to available alternatives.

3. **Francesco Bianchini.** Superorganisms and intelligence

Superorganisms are particular biological entities that are different both from multicellular organisms and unicellular ones. They are a sort of intermediate level as regards the structure and the traits of behavior, even though, from a different point of view they seem to be superior in some capabilities to high-level organisms and they seem to be an emergent phenomenon grounded on an a particular intermediate level consisting of (partially) autonomous entities. In cognitive modeling the inspiration from biology is one of the most important set of heuristics for model architecture building in order to explain cognitive processes, event and phenomena, but also to get really intelligent machines. In this perspective, intelligence seems to be considered as the general capability of facing the external world and environmental (social and natural) events in a condition of restricted and partial information, constituted by representations. Such representations, even though partial, has to be flexible and re-structurable every time and for every task. In this paper, I suggest that the peculiarities of superorganisms and their own structure could be fruitful in artificial intelligence in order to understand how multilevel system (like also mind-brain systems) works, and how (weak) emergent phenomena are thinkable as really causing effects in a structure made by interactive levels, also for the aims of cognitive modeling.

4. **Peter Boltuć.** Non-Reductive Machine Consciousness

In what sense consciousness can be engineered? Obviously functional consciousness can. Yet, I extend the argument beyond the tradition of behaviorist or reductive views on consciousness that predominate within present-day cognitive science. Here is what I call the engineering thesis in



machine consciousness: If Nagel-Chalmers-Block-style non-reductive materialism about first-person consciousness holds true and if we can understand how such consciousness operates, the consciousness it involves can, in principle, be engineered.

## 5. **Fabio Bonsignorio.** Some Observations on the Embodiment of Turing Machines

Godel's undecidability theorem and Turing's halting problem underline the inherent limits of the axiomatic method. If we look at the extreme implementation of a computer program as a formal machine proving theorems about 'reality', it is natural to predict that no computing program will be able to cope deterministically with the challenges of a changing and open ended world. For this reason, the GOFAI (Good Old Fashioned AI) paradigm had and still has a role in clarifying the minimum capabilities of a knowledge representation 'logical machine' and the likes. However, it is unlikely to provide an insight about 'what cognition/intelligence' is in nature. Does the embedding of a Turing machine on a multi rigid body help? Probably not. Does the embedding of a Turing machine on a soft robotics platform change something? This is maybe less obvious, but probably not. We argue that the 'body' is part of the Turing machine program (or program repertory) and extend Chaitin's recent proposal on 'meta-biology' by considering the program's embodiment. We suggest a principled way to model natural and artificial intelligent/cognitive agents which might not suffer of the limitations of mainstream approaches.

## 6. **Joe Dewhurst.** Alien Intelligence

I argue that sophisticated embodied robots will employ conceptual schemes that are radically different to our own, resulting in what might be described as "alien intelligence". Here I introduce embodied robotics and conceptual relativity, and consider the implications of their combination for the future of artificial intelligence. This argument is intended as a practical demonstration of a broader point: that our interaction with the world is fundamentally mediated by the conceptual frameworks with which we carve it up.

## 7. **Sam Freed.** Sub-gadamerian AI

In their book "Understanding Computers and Cognition" Winograd and Flores (1986) characterize the "rationalist" viewpoint prevailing in AI, and summarize the three main critiques of the rationalist project, not least the phenomenological critique. This critique, mainly by Dreyfus (1979, 2007) is based on the philosophy of Heidegger and Merleau-Ponty. Since this critique was voiced there has been a search for "Heideggerian AI" (Dreyfus 2007, Wheeler 2005). Winograd & Flores briefly mention the possibility of using the hermeneutics of Gadamer as a basis for a different phenomenological approach to AI (pp 28-32). In this paper I explore a possible interpretation of their call for Gadamerian AI. In this exploration, I expand Gadamer's view of interpretation to the sub-linguistic level, and show how much more practical it may be than Heideggerian AI, as an alternative to rationalist AI.

## 8. **Swen Gaudi.** Evolving Behaviour Arbitration for Real-Time Controllers

Digital games are well known for high quality graphics and virtual worlds. In contrast to that, games are not known for their exceptionally good AI. This is due to the ongoing divide between academia and industry in that area. Over the last decade game AI research gained a lot of momentum dealing mostly with finding optimal strategies. Existing approaches addressed some of the issues of believable agent generation but removed a lot of initial design freedom. In this work an approach is presented combining dynamic planning with an evolutionary method to restructure the created dynamic plan. Thus, focusing on maintaining the power of human creativity to create an easy to grasp dynamic plan which additionally incorporates adaptation towards the players.

## 9. **Susan Hempinstall.** Human Memory and External Storage Implications



My model deals with current challenges in Philosophy of Memory and also addresses necessary emergent functional capabilities of AI agents. The model engages issues in the Philosophy of Memory, more specifically the challenges of parity, objectification, and integration. The model resolves parity issues, clarifies objectification, and explains integration of internal and external memory sources. From an AI perspective the model provides an architecture of necessary mechanisms around memory recall that are present in human agents, and required in non-human agents to be recognized as intelligent. The model has four functions in two inter-cooperating pairs that allow recognition of a feature-set as part of its own prior experience. The model has two advantages in that it allows projection of social entailments, and provides a rich means of philosophical and empirical testing of external memory within the larger frame of Extended Mind Hypothesis.

## 10. **Dan Hodges.** Evolution: The Origins of Intelligence

It is suggested that probability modeling shall be integral to the anticipated future arrival of Artificial Intelligence. This paper looks to the forward "unknowns" of A.I. probability modeling in an effort to examine the distant-past conceptual and effectual beginnings of our own intelligence. In pursuing this endeavor, it is necessary to examine the process of Evolution one-step removed from it's own two-pronged beginning. A thought experiment is used to systematically dismantle evolutionary process from our present point of understanding to its "inert" pre-Evolutional physical components (without the benefit or presence of organization, purpose or intelligence - as given by the hypothesis ). When this process is then paradoxically reconstructed and examined with forward, conceptual and effectual probability modeling, a common source origin for organic and inorganic intelligence must now be considered.

## 11. **Migle Laukyte.** Kant and Kurzweil: Transcendental Philosophy Meets the Pattern Recognition Theory of Mind

I draw a parallel between two theories of the way our mind works: on the one hand is Kant's metaphysical theory, postulating a set of pure transcendental intuitions and categories; on the other is Kurzweil's neuroscientific pattern-recognition theory of the mind. These two theories approach the mind from two completely different standpoints, the first being a general, top-down account describing the conditions failing which we would not be able to make sense of reality, while the second is a bottom-up account of how the mind recognizes, connects, and organizes patterns through pattern-recognition modules and pattern recognizers. My argument is that, despite these basic differences, the two accounts can profitably be worked into a coherent whole, yielding a picture of the mind as a combination of two different but interfacing types of processes.

## 12. **Jekaterina Novikova, Swen Gaudi, Joanna Bryson and Leon Watts.** Artificial Emotions in Social Robots

With the increasing demand for robots to assist humans in shared workspaces and environments designed for humans, research on human-robot interaction (HRI) gains more and more importance. Robots in shared environments must be safe and act in a way understandable by humans, through the way they interact and move. As visual cues such as facial expressions are important in human-human communication, research on emotion recognition, expression, and emotionally enriched communication is of great importance to HRI and has also gained increasing attention during the last two decades. The goal of our work-in-progress is to further explore the idea of emotion as a strong influence of action selection and to analyse the role of emotion in HRI and human-robot cooperation. We present here a model for incorporating emotions and emotional expressions into POSH dynamic plans. We intend to measure three conditions: emotional action selection with and without expression and emotion free action selection, to observe whether emotional communication will empower the robot to influence a humans behavior.



## 13. **Marcin J. Schroeder.** Searching for Mechanisms of Autonomous Intelligence: Natural vs. Artificial Computation

Homunculus fallacy, the main obstacle in the study of consciousness, is also present in the study of computation. When searching for the mechanisms of autonomous intelligence, it is necessary to separate what constitutes the “natural” computational process, and where begins our human intervention through the interpretation of its input and output. To eliminate this intervention, we can consider operation of a Turing machine as transition between arbitrary elements of a Boolean algebra induced by manipulation of its atoms. Operation of Turing machine traditionally described in a digital way involving assignment of natural numbers as the meaning for digital configurations, becomes analog, i.e. independent from conventional interpretation. We can find similar analog computational processes in the geometric straightedge and compass constructions.

## 14. **Stuart Shieber.** There Can Be No Turing-Test-Passing Memorizing Machines

Anti-behaviorist arguments against the validity of the Turing Test as a sufficient condition for attributing intelligence are based on a \emph{memorizing machine}, which has recorded within it responses to every possible Turing Test interaction of up to a fixed length. The mere possibility of such a machine is claimed to be enough to invalidate the Turing Test. I consider the nomological possibility of memorizing machines and how long a Turing Test they can pass. I replicate the analysis of this \emph{critical Turing Test length} based on the age of the universe, show how considerations of communication time shorten that estimate and allow eliminating the sole remaining contingent assumption, and argue that the bound is so short that it is incompatible with the very notion of the Turing Test. I conclude that the memorizing machine objection to the Turing Test as a sufficient condition for attributing intelligence is invalid.

## 15. **Anthony Smith.** Augmenting Simple Behaviours with Emotions in Real-Time Strategy Games

The video games industry is continuously striving to improve the quality of its games in terms of realism, game play, and immersion. The use of artificial intelligence (AI) has become very important to achieve these goals (?), especially in providing convincing behaviours from its non-player characters (NPC). The behaviours of non-player characters in real-time strategy games tend to be simple and homogenous. This is mainly due to the large number of such characters in these games. To give these non-player characters a more individualistic set of behaviours the standard decision making scheme can be enhanced by a higher level cognitive process. We suggest that emotions is a good candidate to influence these character's behaviours. This research sets out to examine the combining of relatively simple goal oriented behaviours, with simple emotions to produce more complex and individualised behaviours. This method of character generation would be most suitable for real-time strategy and simulation type games, where the character's behaviours would be more interesting, but where these behaviours are not essential to the game's narrative.

## 16. **Niklas Toivakainen.** Man and his Invention

In this paper I will highlight how contemporary naturalistic philosophy of mind and AI theory is inherently bound to the project of modern experimental science: to its programmatic form established in the early 17th century. Importantly, this new paradigm established a strong link between the concept of knowledge/truth and power/control. My proposal is that this conceptual level has always been directly connected to the social, political and economic level. Another important issue that will be discussed is how man's self-understanding is just as much characterized by his ignorance and his repressions as by his clarity. These two central features give us an alternative way of approaching the conceptual problems of AI theory and the relation between man and machine.

## 17. **Bruce Toy.** Consciousness Protocol



This paper introduces a functional model for understanding consciousness as a part of the overall brain's structure. At the base is a novel protocol for AI structure that was recently published. One layer of that protocol, nominally used for location modeling, actually becomes a center for consciousness. The referent protocol represents a functional architecture that was developed independently of most current AI hypotheses, though it does have some common features with many of them.

## 18. **Pierre Vadnais.** Intelligence, Semiotic and Autopoiesis

The AI quest is based on the axiom that all algorithms can be executed by a universal Turing machine. Sixty years of computationalism (trying to algorithmize the mind) have led to the symbol grounding problem. As many years of connectionism (trying to algorithmize the brain) have hinted that solutions might come from embodiment and situatedness. However, Floridi (2011) raised an additional constraint: the zero semantic commitment condition. This paper will investigate the role of information (data) in the development of autopoietic semiotic systems and, paraphrasing Newell and Simon's (1976) hypothesis, propose that only autopoietic semiotic systems have the necessary and sufficient means for practical (sensorimotor, preverbal and preconscious) intelligence which is a necessary developmental step (Piaget 1936) towards general intelligent action.

## 19. **Arne M. Weber and Gottfried Vosgerau.** On the Very Notion of "Grounded Cognition"

Questions like 'what does it mean that cognition is "embodied" or "grounded"?' are still to be answered in detail, and the possibility of operationalizing this concept in empirical research – as well as its explanatory advantage for theoretical progress – remain unclear. To achieve some progress in this regard, we present grounded action cognition as a theoretical framework for understanding the interdependencies of motor control and action-related cognitive processes. Motor control mechanisms have to be understood in terms of acquisition and constitution conditions for cognitive abilities. On this basis, we formulate three possible – conceptual and empirical more or less plausible – theoretical conceptions. The most plausible conception allows for specific insights into the human mental architecture and provides important explanatory tools.

## 15:00-16:00 Keynote: **Jean-Christophe Baillie** (CSO, Aldebaran Robotics, Paris)

"AI: The Point of View of Developmental Robotics"

(Nissan Lecture Theatre) Chair: Shieber

**ABSTRACT:** At the crossroad of robotics and AI, the grounding problem has received much attention lately through a particular approach called developmental robotics. This direction explores how our current understanding of child cognitive and social intelligence development can inspire roboticists into building more robust, grounded and autonomously learning agents. Aldebaran Robotics, a major player in the field of humanoid robotics is launching a private research lab (A-Lab) on developmental robotics, focusing in particular on the long term goal of language acquisition in autonomous robots. We will review the challenges of the field and the particular directions that the A-Lab is going to explore.





**16:00-17:00 Keynote: Selmer Bringsjord** (Rensselaer Polytechnic Institute, Troy, NY)

"What Does Watson 2.0 Tell Us About the Philosophy & Theory of AI?"

(Nissan Lecture Theatre) Chair: Shieber

ABSTRACT: When IBM's Deep Blue beat Kasparov in 1997, I (1998) complained that chess is too easy, relative to what the human mind can muster. But then, in 2011, playing a game based in human language, IBM's Watson (1.0) beat the best human /Jeopardy!/ players on the planet. Now, thanks to a grant from IBM to RPI, a team at the latter is working with IBM to make Watson smarter; i.e., we're working toward Watson n.0, and beyond. What does Watson's prowess, and the engineering now underway to make him smarter, tell us about the philosophy, theory, and future of AI? We present and defend an answer to this question, one that among other things implies that a key, foundational hierarchy among computational formal logics (some of which constitute a mathematization of Watson 1.0's UIMA-based mind) provide a framework for predicting the future of the interaction between /homo sapiens sapiens/ and increasingly intelligent computing machines. And it turns out that this future was probably anticipated and called for in no small part by Leibniz, who thought that God, in giving us formal logic for capturing mathematics, sent thereby a hint to humans that they should search for a formal logic able to capture cognition across the full span of human intelligence.

**17:00-17:30** **Poster Session** & *Coffee break* (Buttery)

Barkati & Rousseaux · Bello · Bianchini · Boltuć · Bonsignorio · Dewhurst · Freed · Gaudl · Hempinstall · Hodges · Laukyte

Novikova, Gaudl, Bryson & Watts · Schroeder · Shieber · Smith · Toivakainen · Toy · Vadnais · Weber · Vosgerau

**17:30-19:00 Sections (3 x 2)**

	Session A: <b>Information</b> (Nissan Lecture Theatre) Chair: Boltuć	Session B: <b>Modelling</b> (Pavilion Room, Gateway Building) Chair: Bickard
17:30 18:00	<b><u>Anderson De Araújo</u></b> <u>Semantic information and artificial intelligence</u>	<b><u>Richard Evans</u></b> <u>Computer Models of Constitutive Social Practices</u>

**Semantic information and artificial intelligence**

ABSTRACT: To measure the degree of informativity of a deduction, it has been proposed by Floridi and others to analyze the static semantic content of propositions. Nonetheless, databases of computational systems are dynamic. For this reason, the present article provides a definition of the dynamic strong semantic information of a logical formula. First, a measure of the informational complexity of a first-order formula in a dynamic database is defined. Thus, the semantic informativity of a formula with respect to a given database is analyzed as the ratio between the number of consequences of this formula in the database and its informational complexity. According to this definition, a deduction could be informative, despite the fact the conjunction of its propositions is not. Moreover, it is possible to measure the deductive power of a computational system in terms of semantic parameters.

**Computer Models of Constitutive Social Practices**

ABSTRACT: The distinction between regulative and constitutive concepts of practice is familiar to philosophers, but relatively unknown within the AI community. This talk will show how the constitutive view can be put to use in AI applications. I will give live demonstrations of two multi-agent simulations that



are based on the constitutive interpretation of social practices: the idea that there are certain actions we can only perform because of the practices we are participating in. The first is a simulation of the Game of Giving and Asking for Reasons, as described in “Making it Explicit”. The second simulation models the pragmatic import of utterances as a set of concurrent practices.

18:00	<b><u>Gordana Dodig Crnkovic</u></b>	<b><u>Bruce Toy</u></b>
18:30	<u>Information, Computation, Cognition. Agency-based Hierarchies of Levels</u>	<u>Behavior Models, an Architectural Analysis</u>

## **Information, Computation, Cognition. Agency-based Hierarchies of Levels**

ABSTRACT: Nature can be seen as informational structure with computational dynamics (info-computationalism), where an (info-computational) agent is needed for the potential information of the world to actualize. Starting from the definition of information as the difference in one physical system that makes a difference in another physical system – which combines Bateson and Hewitt's definitions, the argument is advanced for natural computation as a computational model of the dynamics of the physical world where information processing is constantly going on, on a variety of levels of organization. This setting helps elucidating the relationships between computation, information, agency and cognition, within the common conceptual framework, which has special relevance for biology and robotics.

## **Behavior Models, an Architectural Analysis**

ABSTRACT: This paper describes a functional model for understanding the brain's approach to behavior modeling as it relates to other animate and inanimate objects. Using a protocol for AI structure that was recently published, we can look at the individual's process for understanding the behavior of the objects in its world. The referent protocol represents a functional architecture that was developed independently of most current AI hypotheses, though it does have some common features with many of them.

18:30	<b><u>Ana-Maria Olteteanu</u></b>	<b><u>Remco Heesen</u></b>
19:00	<u>From Simple Machines to Eureka in Four Not-So-Easy Steps. Towards Creative Visuospatial Intelligence</u>	<u>Interaction Networks With Imperfect Information</u>

## **From Simple Machines to Eureka in Four Not-So-Easy Steps. Towards Creative Visuospatial Intelligence**

ABSTRACT: We are still far from building machines that match human-like visuospatial intelligence or creativity. Humans are able to make visuospatial inferences, make creative use of affordances, use structures as templates for new problems, generate new concepts compositionally out of old ones and have moments of insight. We discuss each of these cognitive abilities and the features they presuppose in a cognitive system. We propose a core set of mechanisms that could support such cognitive features. We then suggest an artificial system framework in which the knowledge-encoding supports these processes efficiently, while being in line with a variety of cognitive effects.

## **Interaction Networks With Imperfect Information**

ABSTRACT: This paper uses agent-based modeling to explore questions concerning collaborative learning. In this model small differences in the initial information of agents lead to large differences in how desirable it is to collaborate with them. Interpreting the agents as scientists with different interests and competence, the model suggests explanations for the phenomenon of “academic superstars”. While the existence of superstars (individuals with a large number of collaborators and citations) could be explained using epistemically irrelevant sociological factors, the model proves that superstars can arise even in the absence of such factors. The model is consistent with the idea that superstars are simply more competent. However, it also suggests a novel explanation, where superstars arise purely in virtue of epistemic luck.



**19:00-20:00 Keynote: Theodore Berger (University of Southern California, L.A.)**

"Engineering Memories: A Cognitive Neural Prosthesis for Restoring and Enhancing Memory Function"

(Nissan Lecture Theatre) Chair: Bishop

**ABSTRACT:** Theodore Berger leads a multi-disciplinary collaboration with Drs. Marmarelis, Song, Granacki, Heck, and Liu at the University of Southern California, Dr. Cheung at City University of Hong Kong, Drs. Hampson and Deadwyler at Wake Forest University, and Dr. Gerhardt at the University of Kentucky, that is developing a microchip-based neural prosthesis for the hippocampus, a region of the brain responsible for long-term memory. Damage to the hippocampus is frequently associated with epilepsy, stroke, and dementia (Alzheimer's disease), and is considered to underlie the memory deficits characteristic of these neurological conditions. The essential goals of Dr. Berger's multi-laboratory effort include: (1) experimental study of neuron and neural network function during memory formation -- how does the hippocampus encode information?, (2) formulation of biologically realistic models of neural system dynamics -- can that encoding process be described mathematically to realize a predictive model of how the hippocampus responds to any event?, (3) microchip implementation of neural system models -- can the mathematical model be realized as a set of electronic circuits to achieve parallel processing, rapid computational speed, and miniaturization?, and (4) creation of conformal neuron-electrode interfaces -- can cytoarchitectonic-appropriate multi-electrode arrays be created to optimize bi-directional communication with the brain? By integrating solutions to these component problems, the team is realizing a biomimetic model of hippocampal nonlinear dynamics that can perform the same function as part of the hippocampus

**20:00-22:00 Conference Dinner, St Antony's College Hall (pre-bookings online)**



# Programme

## Philosophy and Theory of Artificial Intelligence

St Antony's College, 62 Woodstock Road, Oxford, OX2 6JF, UK

**Sunday, 22.09.2013**

**9:00-10:00 Keynote: Murray Shanahan (Imperial College, London)**

"Consciousness, Artificial Intelligence, and the Frame Problem"

(Nissan Lecture Theatre) Chair: Scheutz

ABSTRACT: This paper posits an intimate relationship between the neural underpinnings of consciousness and the computational requirements for overcoming the frame problem, where the frame problem is construed as the challenge of bringing all relevant cognitive resources to bear on an ongoing situation. If the brain is considered as a dynamical system, then the hallmark of consciousness, according to the claim, is integration. Whether in respect of a perception, a sensation, or a thought, the conscious condition arises when the system's many independent parts (processes) act under the influence of the system as a whole, while the system as a whole acts in a way that aggregates the influence of all its parts. As well as correlating with conscious modes of brain activity, this dynamical regime facilitates the formation of coalitions of brain processes drawn from the full combinatorial repertoire of possibilities.

**10:00-11:00 Keynote: Michael Wheeler (University of Stirling, Scotland)**

"AI and Extended Cognition"

(Nissan Lecture Theatre) Chair: Scheutz

ABSTRACT: According to the hypothesis of extended cognition (ExC), there are actual (in this world) cases of intelligent action in which thinking and thoughts (more precisely, the material vehicles that realize thinking and thoughts) are spatially distributed over brain, body and world, in such a way that the external (beyond-the-skull-and-skin) factors concerned are rightly accorded cognitive status. Two questions that one might naturally ask in relation to ExC are 'How might research in artificial intelligence (AI) bear on the truth or otherwise of ExC?' and 'Would the adoption of ExC enable us to do AI better?'. In answer to the first question, I shall argue that although AI cannot directly show ExC to be true or false, it can make a crucial indirect contribution to the issue by helping us to articulate what, in the debate over ExC, is standardly known as the mark of the cognitive. In answer to the second question, I shall argue that if a successful case for ExC cannot be built using the strategy outlined in answer to the first question, then ExC will be of no more than heuristic value to any AI researcher who already recognizes the intimate causal dependence of cognition on environmental scaffolding.

**11:00-11:30 Coffee break (Buttery)**

**11:30-13:30 Sections (4 x 2)**

<p>Session A: <b>Morals</b> (Nissan Lecture Theatre) Chair: Sandberg</p>	<p>Session B: <b>Embodied Cognition</b> (Pavilion Room, Gateway Building) Chair: Miłkowski</p>
----------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------



11:30 12:00	<b><u>Matthias Scheutz</u></b> <u>The need for moral competency in autonomous agent architectures</u>	<b><u>Elena Spitzer</u></b> <u>Tacit Representations and Artificial Intelligence: Hidden Lessons from an Embodied Perspective on Cognition</u>
----------------	----------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------

**The need for moral competency in autonomous agent architectures**

ABSTRACT: Soon autonomous robots will be deployed in our societies for many different application domains, ranging from assistive robots for healthcare settings, to combat robots on the battlefield, and all these robots will have to have the capability to make decisions autonomously. In this paper we argue that it is imperative that we start developing moral capabilities deeply integrated into the control architectures of such autonomous agents. For, as we will show, any ordinary decision-making situation from daily life can be turned into a morally charged decision-making situation, where the artificial agent finds itself presented with a moral dilemma where any choice of action (if inaction) can potentially cause harm to other agents.

**Tacit Representations and Artificial Intelligence: Hidden Lessons from an Embodied Perspective on Cognition**

ABSTRACT: In this talk, I will explore how an embodied perspective on cognition might inform research on artificial intelligence. Many embodied cognition theorists object to the central role that representations play on the traditional view of cognition. Based on these objections, it may seem that the lesson from embodied cognition is that AI should abandon representation as a central component of intelligence. However, I argue that the lesson from embodied cognition is actually that AI research should shift its focus from how to utilize explicit representations to how to create and use tacit representations. To develop this suggestion, I provide an overview of the commitments of the classical view and distinguish three critiques of the role that representations play in that view. I provide further exploration and defense of Daniel Dennett's distinction between explicit and tacit representations. I argue that we should understand the embodied cognition approach using a framework that includes tacit representations. Given this perspective, I will explore some AI research areas that may be recommended by an embodied perspective on cognition.

12:00 12:30	<b><u>Vincent C. Müller (co-author Nick Bostrom)</u></b> <u>Future Progress in Artificial Intelligence: A Poll Among Experts</u>	<b><u>Carlos Herrera and Ricardo Sanz</u></b> <u>Heideggerian AI and the being of robots</u>
----------------	-------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------

**Future Progress in Artificial Intelligence: A Poll Among Experts**

ABSTRACT: In some quarters, there is intense talk about high-level machine intelligence (HLMI) and superintelligent AI coming up in a few decades, bringing with it significant risks for humanity; in other quarters, these issues are ignored or considered science fiction. We wanted to clarify what the distribution of opinions actually is, what probability the best experts currently assign to high-level machine intelligence coming up within a particular time-frame and which risks they see with that development. We thus designed a brief questionnaire and distributed it to four groups of experts. The results show some discrepancy between different groups but also an agreement that there is significant probability of AI systems reaching or surpassing human ability within a few decades.

**Heideggerian AI and the being of robots**

ABSTRACT: In this paper we discuss to what extent Heideggerian AI approaches are consistent with Heidegger's philosophy. We identify a number of inconsistent premises: commitment to a positive contribution to the advancement of AI techniques; exclusive attention to ontological analysis of humans-Dasein; robots as copies of natural systems; consideration of humans, animals and robots as beings of the same kind. These premises run against the Heidegger's theory of science, his views on technology, and the core of Heidegger's philosophy: the significance of ontological categories. A truly Heideggerian AI should tackle the question of what robots are. In other words, realise an ontological analysis on the being of robots.



12:30 13:00	<u><b>Marcello Guarini and Jordan Benko</b></u> <u>Order Effects, Moral Cognition, and Intelligence</u>	<u><b>Madeleine Ransom</b></u> <u>Why Emotions Do Not Solve the Frame Problem</u>
----------------	------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------

**Order Effects, Moral Cognition, and Intelligence**

ABSTRACT: Order effects have to do with how the order in which information is presented to an agent can effect how the information is processed. This paper examines the issue of order effects in the classification of moral situations. Some order effects mark a localized failure of intelligence. The hypothesis examined herein is that the processes or mechanisms that make undesirable order effects possible may also have highly desirable effects. This will be done by comparing two artificial neural networks (ANNs) that classify moral situations, one subject to order effects and another that is not subject to them. The ANN subject to order effects has advantages in learning and noise tolerance over the other ANN – features hard to ignore in modeling intelligence.

**Why Emotions Do Not Solve the Frame Problem**

ABSTRACT: It has been claimed that the emotions solve, or help solve, the frame problem (Megill & Cogburn 2005; Evans 2004; DeSousa 1987; Ketelaar & Todd 2000). However, upon careful examination, the specific proposals on offer are underspecified. The purpose of this paper is to precisify and evaluate these proposals. Specifically i) what is meant by the frame problem in these instances; ii) what are the proposed solutions; and iii) do they work? I will argue that while the emotions – best viewed as forming part of the heuristics research program – are a viable proposal for helping solve the intracontext frame problem, they cannot function to solve or help solve the intercontext frame problem, as they are themselves subject to contextual variability.

13:00 13:30	<u><b>Miles Brundage</b></u> <u>Artificial Intelligence and Responsible Innovation</u>	<u><b>Robin Zebrowski</b></u> <u>The Machine Uprising Has Been Delayed (Again): Extended Mind Theories as A(nother) Challenge to Artificial Intelligence</u>
----------------	-------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------

**Artificial Intelligence and Responsible Innovation**

ABSTRACT: Thought leaders in AI often highlight the importance of socially responsible research, but the current literature on the social impacts of AI tends to focus on particular application domains and provides little practical guidance to researchers. Additionally, there has been little interaction between the AI literature and the field of “responsible innovation,” which has developed many theories and tools for shaping the impacts of research. This paper synthesizes key insights from both of these literatures, and describes several aspects of what responsible innovation means in AI: ethically informed selection of long-term goals, reflectiveness about one’s emphasis on theoretical vs. applied work and choice of application domains, and proactive engagement with communities that may be affected by one’s research.

**The Machine Uprising Has Been Delayed (Again): Extended Mind Theories as A(nother) Challenge to Artificial Intelligence**

ABSTRACT: We have failed to make significant progress in the field of strong artificial intelligence in spite of a robust neuroscience and continued evolution of our theories of mind. The Extended Mind Hypothesis provides yet another serious challenge to AI, and ought to force a total re-examination of our assumptions in the field. This paper argues that, given the evidence and argument for extended minds, AI is due for another major shift in basic approach.

**13:30-14:30 Lunch (Buttery)**



**14:30-15:30 Keynote: Luciano Floridi** (University of Oxford)

"Enveloping the World - How Reality Is Becoming AI-Friendly"

(Nissan Lecture Theatre) Chair: Shagrir

ABSTRACT: Nanotechnology, the Internet of Things, Web 2.0, Semantic Web, Cloud computing, motion-capturing games, smart phones apps, GPS, Augmented Reality, Artificial Companions, drones... is there a unifying perspective from which all these ICT phenomena might be interpreted as aspects of a single, macroscopic trend? Part of the difficulty, in answering this question, is that we are still used to looking at ICTs as tools to interact with the world, when in fact they have become environmental forces, which are creating and shaping (that is, re-ontologising) our reality, more and more pervasively. To put it briefly, the answer may lie in realising that ICTs are enveloping the world - In robotics, an envelope (also known as reach envelope) is the three-dimensional space that defines the boundaries that the robot can reach.

**15:30-18:00 Sections (5 x 2)**

	Session A: <b>Intelligence &amp; Reasoning</b> (Nissan Lecture Theatre) Chair: Bryson	Session B: <b>Embodied Cognition</b> (Pavilion Room, Gateway Building) Chair: Nasuto
15:30 16:00	<b><u>Ioannis Votsis</u></b> <u>Science with Artificially Intelligent Agents: The Case of Gerrymandered Hypotheses</u>	<b><u>Adam Linson</u></b> <u>The Expressive Stance: Intentionality, Expression, and Machine Art</u>

**Science with Artificially Intelligent Agents: The Case of Gerrymandered Hypotheses**

ABSTRACT: Barring some civilisation-ending natural or man-made catastrophe, future scientists will likely incorporate fully fledged artificially intelligent agents in their ranks. Their tasks will include the conjecturing, extending and testing of hypotheses. At present human scientists have a number of methods to help them carry out those tasks. These range from the well-articulated, formal and unexceptional rules to the semi-articulated rules-of-thumb and intuitive hunches. If we are to hand over at least some of the aforementioned tasks to artificially intelligent agents, we need to find ways to make explicit and ultimately formal, not to mention computable, the more obscure of the methods that scientists currently employ with some measure of success in their inquiries. The focus of this talk is a problem for which the available solutions are at best semi-articulated and far from perfect. It concerns the question of how to conjecture new hypotheses or extend existing ones such that they do not save phenomena in gerrymandered or ad hoc ways. This talk puts forward a fully articulated formal solution to this problem by specifying what it is about the internal constitution of the content of a hypothesis that makes it gerrymandered or ad hoc. In doing so, it helps prepare the ground for the delegation of a full gamut of investigative duties to the artificially intelligent scientists of the future.

**The Expressive Stance: Intentionality, Expression, and Machine Art**

ABSTRACT: This paper proposes a new interpretive stance toward works of art that is relevant to AI research, termed the 'expressive stance'. This stance makes intelligible a critical distinction between present-day machine art and human art, but allows for the possibility that future machine art could find a place alongside our own. The expressive stance is elaborated as a response to Daniel Dennett's notion of the intentional stance, which is critically examined with respect to his specialized concept of rationality. The paper also shows that temporal scale implicitly serves to select between different modes of explanation in prominent intentional theories. It also highlights a relevant difference, in terms of phenomenological background, between expert systems and systems that produce art.



16:00	<u>David Davenport</u>	<u>Alex Tillas</u> and <u>Gottfried Vosgerau</u>
16:30	<u>Explaining Everything</u>	<u>Perception, Action &amp; the Notion of Grounding</u>

**Explaining Everything**

ABSTRACT: This paper looks at David Deutsch's recent claim that nothing less than a philosophical breakthrough is needed before real progress can be made in constructing an AGI (Artificial General Intelligence). For an agent to be truly intelligent, Deutsch argues, it must be able to generate new explanations about how the world works for itself. Such creativity, then, is the ingredient missing from current would-be AIs, and a problem he traces to the philosophy underpinning their implementation. While agreeing with the gist of Deutsch's argument, I take issue with certain aspects of it, including his lexicon. In doing so, I will contrast his views with those of Floridi, Bickhart and others, to suggest that at least some of the required philosophical insights do, in fact, already exist.

**Perception, Action & the Notion of Grounding**

ABSTRACT: Traditionally, the mind has been seen as neatly divided into input; central processing; and output – almost watertight – compartments. This view has been recently challenged by theorists who argue that cognition is grounded in bodily states (sensorimotor representations). In this paper, we focus on the debate about the relation between perception and action in an attempt to flesh out the process and in turn clarify the notion of grounding. Given that, at present, the debate in question is far from settled, we attempt an assessment of the implications that possible outcomes of this debate would have on Grounding Cognition Theories. Interestingly, some of these consequences seem to threaten the very existence of the Grounded Cognition program as a whole.

16:30	<u>Sjur Kristoffer Dyrkolbotn</u> and <u>Truls Pedersen</u>	<u>Massimiliano Cappuccio</u>
17:00	<u>Arguably argumentative: A formal approach to the argumentative theory of reason</u>	<u>The Seminal Speculation of a Precursor: Elements of Embodied Cognition and Situated AI in Alan Turing</u>

**Arguably argumentative: A formal approach to the argumentative theory of reason**

ABSTRACT: We address the recently proposed argumentative theory of reason, which suggests that human cognition must be understood as having evolved in order to facilitate social interaction by way of argumentation. We note a fundamental tension between a social view of human reasoning, like that adopted by the argumentative theory, and an individualistic view, often adopted in the philosophy of rationality and intelligence, where the focus is on the mental processes of individual reasoners. Proposing a formal approach to the study of argumentative reasoning, we argue that the theory of abstract argumentation frameworks, as studied in artificial intelligence, can be used as a starting point for formalisms that will allow us to shed light on the logical principles at work in argumentative interaction. We conclude with a discussion addressing philosophical implications of this endeavour, and we argue that the argumentative theory gives rise to a fresh and interesting point of view on intelligence and rationality whereby these notions can be seen as pertaining to social rather than individual aspects of agency.

**The Seminal Speculation of a Precursor: Elements of Embodied Cognition and Situated AI in Alan Turing**

ABSTRACT: Some key notions of situated robotics find their origin in Alan Turing's seminal work. They emerge in both his foundation of computationalism (cognitive constraints of formalized symbolic systems) and theory of AI (bodily constraints of learning machines). I will show the deep link between these two parts of Turing's speculation, ultimately relatable to the embedded and extended nature of the logico-symbolic practices that are deemed capable to scaffold real intelligence. Real intelligence is both structurally limited and actively mediated by embodied, cognitive, and even cultural conditions, in accord with the cognizer's biological constitution and its historical coupling with the environment.





17:00 17:30	<u><b>Ron Chrisley</b></u>  <u>The appearance of robot consciousness</u>	<u><b>J. Mark Bishop, Slawomir J. Nasuto, Matthew Spencer, Etienne Roesch and Tomas Tanay</b></u>  <u>Playing HeX with Aunt Hilary: games with an anthill</u>
----------------	--------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------

**The appearance of robot consciousness**

ABSTRACT: This paper is a critique of some central themes in Pentti Haikonen's recent book, *Consciousness and Robot Sentience*. Haikonen maintains that the crucial question is how the inner workings of the brain or an artificial system can appear, not as inner workings, but as subjective experience. It is argued here that Haikonen's own account fails to answer this question, and that the question is not in fact the right one anyway.

**Playing HeX with Aunt Hilary: games with an anthill**

ABSTRACT: In a reflective and richly entertaining piece from 1979, Doug Hofstadter famously imagined a conversation between "Achilles" and an anthill (the eponymous "Aunt Hillary"), in which he playfully explored many ideas and themes related to cognition and consciousness. For Hofstadter, the anthill is able to carry on a conversation because the ants that compose it play roughly the same role neurons play in the brain. Unfortunately, Hofstadter's work is notably short on detail suggesting how this magic might be achieved. In this paper we demonstrate how populations of simple ant-like creatures can be organised to solve complex problems; problems that involve the use of forward planning and strategy and in so doing, introduce Hofstadter's "Aunt Hilary" to the complex charms of the strategic game 'HeX'.

17:30 18:00	<u><b>Tijn Van Der Zant and Bart Verheij</b></u>  <u>Elephants don't fly; and they know it!</u>	<u><b>Brian Cantwell Smith</b></u>  <u>Computation's Lost Promise</u>
----------------	-------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------

**Elephants don't fly; and they know it!**

ABSTRACT: Neither top-down reasoning systems nor bottom-up behavior-based robotics lead to robots that can reasonably be called intelligent. Both approaches seem to work only to a certain level. In reasoning systems there can be too many possibilities to compute, in behavior-based architectures there are often not enough possibilities to finish a task. In this paper we demonstrate the early results of a bottom-up behavior-based system that uses machine learning for behavior selection. It is integrated with a reasoning architecture that learns from experience and then restructures the behavior-based architecture. This leads to robots that understand what they can and cannot do. The experiences are stored and used within a rules-with-exceptions architecture. The entire system can be described as knowledge-based adaptive robots. It is used in the international benchmark of RoboCup@Home.

**Computation's Lost Promise**

ABSTRACT: Computation's Lost Promise

**18:00-18:30 Coffee break (Buttery)**



**18:30-20:00 Keynote: Daniel C. Dennett** (Tufts University, Boston)

"If brains are computers, what kind of computers are they?"

(Nissan Lecture Theatre) Chair: Müller

**ABSTRACT:** Our default concepts of what computers are (and hence what a brain would be if it was a computer) include many clearly inapplicable properties (e.g., powered by electricity, silicon-based, coded in binary), but other properties are no less optional, but not often recognized: Our familiar computers are composed of millions of basic elements that are almost perfectly alike--flipflops, registers, or-gates--and hyper-reliable. Control is accomplished by top-down signals that dictate what happens next. All subassemblies can be designed with the presupposition that they will get the energy they need when they need it (to each according to its need, from each according to its ability). None of these is plausibly mirrored in cerebral computers, which are composed of billions of elements (neurons, astrocytes, ...) that are no-two-alike, engaged in semi-autonomous, potentially anarchic or even subversive projects, and hence controllable only by something akin to bargaining and political coalition-forming. A computer composed of such enterprising elements must have an architecture quite unlike the architectures that have so far been devised for AI, which are too orderly, too bureaucratic, too efficient.