Course No: CS786

Title: Computational cognitive science

About the course:

To what degree functions of the mind can be reproduced or simulated by a computer is a question that has become volubly prominent in recent years. It is often posed for public consumption as a matter of *ends* – when will computer performance surpass human performance on interesting challenges. From a scientific standpoint, it is more useful to ask this question from the standpoint of *means* – what physical and informational resources do biological organisms have available to respond adaptively to situations they encounter in the world?

This is the question that computational cognitive scientists seek to answer. Such research frequently starts from parametric characterizations of empirical behavior. Theorists then develop computational models of the phenomenon that capture the quantitative relationship between these parameters and various experimental conditions. A good empirical fit permits further questions of biological and epistemic plausibility to be asked of the model. Models that pass these quasi-philosophical checks graduate to the status of theories. These accounts are, inevitably, challenged as incomplete or erroneous by further iterations of experiments and models.

The cycle of research in cognitive science, therefore, encompasses, in order of the workflow presented above, neuroscience and psychology, statistics, computer science, and philosophy. This course is meant to introduce students interested in the computational aspects of cognitive science a relatively comprehensive overview of the discipline.

Previous iterations of this course have covered the material listed below in the course outline segmented by *cognitive* processes. In this iteration of the course (Winter 2022), my plan is to cover more or less the same material, but organized according to *algorithmic* processes. The five key algorithmic processes involved in cognition, by my reckoning, are association, reinforcement, accumulation, embodiment and learning. So we will explore our material thematically by first looking at models of cognition that heavily use association, or explain fundamentally associative phenomena, then turn our attention to reinforcement and so on.

Over 4-5 meetings per area, these algorithmic primitives of cognition will be introduced via descriptions of empirical studies, followed by the chronology of models seeking that utilize them. Instruction in each topic will conclude with an instructor-led discussion of the merits of competing models, terminating in an appreciation of promising future directions of research in the area. The instructor's emphasis will slant towards approximate Bayesian approaches to such challenges.

This course has historically used continuous assessment, eschewing mid-sem and end-sem exams. I am hoping to retain that aspect of the course. So, the evaluation plan of the course is as follows:

1. 40% of course credit will come from performance in 4 assignments throughout the course (10% each)

2. 20% of course credit will come from attendance in the course, as measured passively by Zoom attendance statistics
3. 20% of course credit will come from performance on a 5000 word research paper each one of you will submit at the end of the course, either replicating an existing model of cognition, or extensively reviewing literature in an area of cognitive science.
4. 10% of course credit will come from participation in various cognitive experiments that will be notified on the course webpage from time to time
5. 10% of course credit will come from my subjective assessment of your activity level and engagement in the classes throughout the conduct of the course

Note: Because of the large emphasis on participation in online classes, students who anticipate difficulty in participating in the live classes at any point in the semester for logistical reasons are strongly urged to drop the course.

Course outline

Foundations – evidence for invariants in behavior – associativity – Pavlovian conditioning – Minsky, Newell and the strong AI program – the framing problem – production system architectures of the mind – the Bayesian revolution – inference, learning and causation – compositionality and probabilistic programs – approximate computation in the mind – algorithmic accounts of sub-optimal inference

Perception – James, Helmholtz, Wundt – classical psychophysics – perceptual modalities – quantification and analysis methods – Gestalt principles – assimilation and contrast effects – poverty of stimulus – Gibsonian psychophysics – Anne Treisman's feature integration account – recognition by components – David Knill & Eero Simoncelli's Bayesian visual perception work

Memory – early experiments – Miller and the magic number 7 – classical experiment settings and analyses – signal detection theory – Tulving's memory types – Baddeley and the discovery of working memory – Rich Shiffrin's line of models and their problems – Austerweil's random walk model – Standing and the fidelity of visual long-term memory connecting to Tim Brady's recent work – Tom Hills and memory search

Decision-making – von-Neumann and *homo economicus* – Rescorla-Wagner, Hall-Pearce and classical conditioning findings – operant conditioning and skill learning – Sutton-Barto-Singh and reinforcement learning building up to skill-learning – cognitive biases in decision-making – Tversky's non-compensatory models – the Gigerenzer fast-and-frugal school of heuristics – fast and slow decisions and their consequences – drift diffusion models and their competition – frugal preference learning

Language – semantics and semiotics – neurobiological foundations of language with empirical evidence – language universals and typology – Sapir-Whorf hypothesis, evidence for and against – pragmatics and social signaling – nativist vs emergent models of language learning – Bayesian accounts of structure learning – non-human languages – Wittgenstein and philosophy as language

Motor control and learning – systemic principles, feedback, redundancy, coordination – physiological basis – information processing problems – Peter Dayan's model-free vs model-based motor models – Danny Wolpert's hierarchical motor control models – Paul Schrater's Bayesian structure learning – hierarchical reinforcement learning – Karl Friston's free energy approach – Nikolai Bernstein's beautiful ideas on the value of noise in the motor system – Jeff Beck's rational sub-optimal inference account

Similarity & categorization – Luce, Shepherd and empirical foundations – exemplars vs prototypes debate with empirical data – Nosofsky, Shiffrin and the rise of cluster models – Anderson's rational model – Josh Tenenbaum's Bayesian program – hierarchical Dirichlet models of categorization – compositionality and the generation of new categories – Liane Gabora's computational models of creativity

References Mostly research papers assigned ahead of lectures.