# Contents

## 1 Overview

## 2 Notes for participants

### 2.1 Diversity and inclusion

### 2.2 Learning goals and opportunities

### 2.3 Best learning practices

### 2.4 Credit and grading

## 3 Lecture topics by date

### 3.1 Week 1 (2/9; 2/11): motivation

### 3.2 Week 2 (2/16; 2/18): universal tools, I; methodology; levels of understanding; behavior

### 3.3 Week 3 (2/23; 2/25): universal tools, II: probability; Bayes

### 3.4 Week 4 (3/2; 3/4): universal tools, III: representation, similarity, generalization

### 3.5 Week 5 (3/11): [wellness day]; universal tools, IV: veridical representation

### 3.6 Week 6 (3/16; 3/18): memory

### 3.7 Week 7 (3/23; 3/25): actions and consequences

### 3.8 Week 8 (3/30; 4/1): higher cognition, I: language

### 3.9 Week 9 (4/6; 4/8): higher cognition, II: reasoning; induction

### 3.10 Week 10 (4/13; 4/15): higher cognition, III: intelligence; problem solving; analogy

### 3.11 Week 11 (4/20; 4/22): intro to neural computation; neurons, I

### 3.12 Week 12 (4/27; 4/29): neurons, II

### 3.13 Week 13 (5/4; 5/6): neurons, III; brains

### 3.14 Week 14 (5/11; 5/13): advanced topics

## 4 Computational concepts, principles, and methods: an abecedary
1 Overview

This course states, motivates, and offers detailed support for the observation that cognition is fundamentally a computational process \[28\]. Students are introduced to a number of conceptual tools for thinking about natural behavior and the cognitive information processing that underlies it, including statistical learning from experience and the use of patterns distilled from past experience in guiding future actions. The application of these tools to the understanding of natural minds and to advancing the goals of artificial intelligence is illustrated on selected examples drawn from the domains of perception, memory, motor control, action planning, problem solving, decision making, reasoning, and creativity.

The material is conceptually advanced and moderately to highly technical. It is aimed at advanced undergraduate students, as well as graduate students from psychology, neurobiology, computer science, and other cognitive sciences. Prior exposure to statistical concepts and the scientific method is essential.

How to use this syllabus

- For each week, there’s a list of readings with references. Some of the readings are required, others are optional. The references are also listed at the end of the syllabus, alphabetically by first author.

- For an alphabetical roster of select key ideas and topics, see section \[4\].

Readings

The recommended textbook is *Computing the Mind: How the Mind Really Works* (Oxford University Press, 2008). Additional readings (a zipped collection of PDFs) are available on the course Canvas site.

There are over 100 references listed at the end of this syllabus. Please do not be alarmed: this does not mean that you are required to read all the papers on that list. Many of the references are there to provide entry points into the technical literature on cognition for those of you who are interested in learning more about it than what this course covers.

2 Notes for participants

This section contains essential information for participants: the inclusion statement\[1\], learning goals and practices, and credit requirements.

2.1 Diversity and inclusion

*Computational Psychology* is more diverse than many other courses at Cornell, in at least two respects. First, it purposely ignores the traditional disciplinary boundaries, as cognitive science has done since its inception (see the illustration on the next page). Accordingly, my plan is for us to freely mix concepts and topics from psychology, mathematics, computer science, and neurobiology. Second, this course does not respect college and program boundaries: historically, it has been attended and successfully completed by students from different colleges and a variety of majors, among them psychology, neurobiology, engineering, information science, computer science, and English, as well as by graduate and professional masters students.

\[1\]The remarks in section \[2.1\] which are specific to this course, are intended to supplement the official Cornell statement on diversity and inclusion, which covers dimensions such as gender, race, socio-economic background, etc., and which can be found here: [http://7/diversity.cornell.edu/](http://7/diversity.cornell.edu/)
On the left, a group photograph of the (disciplinarily, but not otherwise, diverse) members of the Hixon Symposium, held in 1948, which helped shape the modern integrative approach to brain/mind science. *Left to right:* (seated) Halstead, Lashley, Klüver, Köhler, Lorente de Nó; (standing) Brosin, Jeffress, Weiss, Lindsley, von Neumann, Nielsen, Gerard, Liddell. On the right, the cover of *Embodiments of Mind* (1965; [73]), a still readable collection of papers by Warren McCulloch, the one member of the group who is absent from the group photo.

This diversity makes inclusion particularly important; interestingly, it also makes it easier to attain, because we can always “triangulate,” and thus better understand, every issue and concern from multiple disciplinary and personal points of view. In such circumstances, the more each of you contributes to the discussion, the more we all learn. Both myself and the TA are fully committed to having everyone’s questions and concerns heard — both in class and outside — and to helping everyone succeed who is willing to invest the effort to do so. In addition to our office hours, we are always available for meetings by appointment; we will also answer promptly any questions posted on Canvas or emailed to us. Let us know if at any point during the semester you have suggestions for making things work better; these may include additions to the present statement, which will be included in periodic revisions, to be shared on Canvas.

### 2.2 Learning goals and opportunities

Active participation in this course should improve significantly your understanding of how the brain/mind works, regardless of your disciplinary and personal background. In addition, upon its completion, you should be able to:

- reflect on how different disciplines provide complementary insights into how the brain/mind works, and how scholarship that cuts across disciplines can help integrate those insights;
- reflect on how people with different personal and educational backgrounds can join forces in solving complex problems.
Some of the instructional strategies and learning opportunities that we plan to use and make available to promote these outcomes are as follows:

- The key conceptual framework that is stated, discussed, and applied in this course — the “levels of understanding” of complex information-processing systems (introduced in week 2) — is inherently geared toward integrating multiple sources of data, methods of analysis, and formal models.

- The examples that will be used in class to illustrate various concepts will be drawn from a variety of naturally diverse real-life situations.

- You will have opportunities to come up with your own examples, thus adding to the diversity of viewpoints, enriching the abstract mathematical and computational concepts that are necessarily impersonal.

### 2.3 Best learning practices

In creating this course and keeping it up to date, I have in a sense been reprising the trajectory of my own multidisciplinary educational background and research experience. For my undergraduate degree, I studied electrical engineering. Following a stint in the military, I went back to school for an MSc and then a PhD in computer science, so as to be able to work at the intersection of mathematics, philosophy, art, and science that Doug Hofstadter described so alluringly in his Pulitzer-winning *Gödel, Escher, Bach* [52] (in cognitive science and AI, this became known eventually as “the book that launched a thousand careers”). The main text for this course, *Computing the Mind* [28], has been written with that great book as an inspiration.

The readings that are drawn from the textbook are specified by chapter, section, and subsection, but ideally the book should be read in its entirety. The key concepts are marked in the text by SMALL CAPITALS and are usually included in the index. The many margin notes may be ignored on the first reading to avoid distraction, or you may want to take them in for an extra helping of (sometimes only marginally relevant) information or an entertaining flight of fancy.

Additional readings come from a selection of more than 100 academic papers, some of which are decades-old classics and some published less than a year ago. Don’t panic! Only a few of the papers, listed in the references at the end of this document, are required reading. These are clearly marked in the week-by-week syllabus. Furthermore, for some of the required papers, it will suffice to read only a few select sections. If in doubt, please feel free to ask (in class or on Canvas).

When reading, always focus on the “big picture” and the key concepts and principles. Concepts and principles are also what I shall be trying to bring out during the lectures, which is why it is important that you attend. Neither technical details (such as proofs of mathematical statements) nor trivia (such as the names of researchers, with the exception of a few luminaries) will appear in the microessay questions. Please bring any concerns about learning practices to my office hours (or to the TA’s).

### 2.4 Credit and grading

To get credit for this course:

1. Attend the lectures; ask questions and participate in the discussion. If you must miss a lecture, please send along an email with an explanation.

2. Every Tuesday by 8am, submit your questions on the weekly readings (via email, to the TA). Each set of questions will be graded on a scale of 5 (excellent) to 0 (missing work). Altogether 10 sets of weekly questions should be submitted by the end of the semester.
3. Every Sunday by noon, submit a **micro-essay** on the weekly topic, announced on the preceding Thursday at the lecture. Each essay will be graded on a scale of 5 (excellent) to 0 (missing work). Submit up to 14 sets of micro-essays; of these, the 10 best will count towards the final grade.

4. (Graduate students only, enrolled in Psych 6140.) Submit a written summary of your impressions and lessons from the material, in an essay form (about 1000 words), by noon, Tuesday, May 18, via email to the instructor.

Final grade components:

<table>
<thead>
<tr>
<th></th>
<th>Psych 3140</th>
<th>Psych 6140</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly question sets:</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Weekly micro-essays:</td>
<td>50%</td>
<td>30%</td>
</tr>
<tr>
<td>Final essay:</td>
<td>–</td>
<td>20%</td>
</tr>
</tbody>
</table>

Up to 3 extra credits can be accrued by participating in the Psychology Department subject pool ([SONA](#)). These will be added to the final score before it is converted into a letter grade.

The list of lecture topics and readings by week number and date begins on the next page. There is one page per week, with the “Primary” (required) and “Other” (optional) readings listed separately. I have also included a few photographs of the authors of the papers, to offer you a glimpse of the collective face of cognitive science; bibliographical entries for papers thus distinguished are marked in **boldface**. Advanced readings (optional), which may be too long or too technical for casual perusal, are marked in **gray**.

[The syllabus continues on the next page.]
3 Lecture topics by date

3.1 Week 1 (2/9; 2/11): motivation

1.1 The subject matter of psychology. The fundamentality of computation [28]. Examples: (i) perception — lightness, or estimating surface reflectance from images; (ii) thinking — planning, or estimating the actions needed to attain a goal; (iii) action — motor control, or estimating the signals that need to be sent to the muscles to execute the desired motion. A quick overview of computation: dynamical systems [56] sections 1,2]; Turing Machines [5].

1.2 The blind men and the elephant. Four case studies: an abstract computational theory [1]; a single-cell electrophysiology study [91]; an imaging study [50]; an engineering hack [74].

Primary readings


Other readings


Volodymyr Mnih
3.2 Week 2 (2/16; 2/18): universal tools, I; methodology; levels of understanding; behavior

2.1 Open your eyes! The general methodology: the Marr-Poggio program for neurosciences [28, ch.4]. A worked-out example: sound localization in the barn owl [63, 72, 51].

2.2 The nature of behavior. Dewey on the “reflex arc” concept [23]. Thurstone on the “stimulus-response” fallacy [107]. The life of behavior [46].

Primary readings


Other readings


3.3 Week 3 (2/23; 2/25): universal tools, II: probability; Bayes

3.1 A probabilistic formulation of cognition [14]. The Bayesian framework [48].

3.2 The Bayesian approach, applied to lightness perception [8 sections 1,2,4].

Primary readings


Other readings

3.4 Week 4 (3/2; 3/4): universal tools, III: representation, similarity, generalization

4.1 Perceptual representation spaces [28, ch.5]. The face space [62, 41]. Multidimensional scaling (MDS).

4.2 Similarity and generalization [28, ch.5]. Shepard’s Law [95, up to section “Derivation…” only]. Deriving Shepard’s Law from the principle of efficient coding [98].

Primary readings


Other readings


3.5 Week 5 (3/11): [wellness day]; universal tools, IV: veridical representation

5.2 Veridical representation in perception \[38\] [18]. Representation is representation of similarities \[26\].

Primary readings


Other readings


3.6 Week 6 (3/16; 3/18): memory


6.2 Memory for concepts, places, and events [28] ch.6; [108, 112].

Primary readings


Other readings

3.7 Week 7 (3/23; 3/25): actions and consequences

7.1 Managing action. The basics of motor control \[114\]. Bayesian motor decision making \[66\].

7.2 Action and reward. Reinforcement learning (RL) \[116\] and its relationship to Bayes. Hierarchical RL \[7\].

Primary readings


Other readings


3.8 Week 8 (3/30; 4/1): higher cognition, I: language

8.1 The structure of language [28] ch.7 and its use [25]. A functionalist approach to linguistic structure [45].

8.2 A computational framework for language and other sequential behaviors [33].

Primary readings


Other readings


3.9  Week 9 (4/6; 4/8): higher cognition, II: reasoning; induction

9.1  Graphical models (Bayesian Networks) and reasoning [28 ch.8].

9.2  Induction [106].

Primary readings


Other readings


Judea Pearl
3.10  Week 10 (4/13; 4/15): higher cognition, III: intelligence; problem solving, analogy

10.1  General intelligence and IQ [47, 76, 4].

10.2  Problem solving [28 ch.8]. Analogy [54] and creativity [53].

Primary readings


Other readings


3.11 Week 11 (4/20; 4/22): intro to neural computation; neurons, I

11.1 Introduction to neural computation. Brains and neurons. Cortical receptive fields (RFs), maps, and hierarchies [28 ch.2,3]. Channel coding and hyperacuity [100].

11.2 What do neurons do?

Projection:

– neurons as readout devices [11];
– neurons, random projections, and similarity [36 69].

Kernels:

– landmarks in representation spaces, similarity, and kernels [34].

Primary readings


Other readings


Reza Shahbazi
3.12 Week 12 (4/27; 4/29): neurons, II

12.1 What do neurons do?

Time-dependent dynamic learning:

– spike timing dependent plasticity (STDP), and Hebbian learning (see the Scholarpedia article [99]);
– history-dependent learning with the BCM rule [16];
– learning dynamic embodied control modes [22].

12.2 What do neurons do?

Population dynamics:

– neural trajectories and classification [10].

Primary readings


Other readings


3.13 Week 13 (5/4; 5/6): neurons, III; brains

13.1 What do neurons do?
Population dynamics (cont.):
   – ongoing dynamics and chaotic itinerancy [88].

13.2 What does the brain do? Prediction [42, 77, 15].

Primary readings


Other readings

3.14 Week 14 (5/11; 5/13): advanced topics

14.1 Bayes and the real world [93]

14.2 Real behavior and real brains [111] [32]

Primary readings


Other readings


4 Computational concepts, principles, and methods: an abecedary

Note: the number of important principles in cognitive science is somewhat larger than the number of letters in the alphabet. To make up for this inconvenience, I have highlighted in SMALL CAPITALS every concept of interest that is mentioned, whether or not it has its own entry.

**ANALOGY** — a comparison that involves a structure mapping between complex entities, situations, or domains [54, 55]. Analogy is central to general cognitive function (general fluid intelligence, often referred to as $gF$ or IQ [101]) and has been hypothesized to underlie structural LEARNING in vision and in language.

**BAYES THEOREM** — a direct consequence of the definition of conditional probability; the basis for the so-called rational theories of cognition. The Bayes Theorem prescribes a way of integrating prior beliefs with
new data, in a way that proves useful in all domains, from perception, through thinking and language, to motor control [64, 117, 35, 103, 83, 66].

**Channel Coding** — measuring a stimulus with a set of graded, overlapping filters (receptive fields or “channels”) supports a high degree of resolution, or hyperacuity, that cannot be achieved through dense sampling by pointlike filters [100]. This principle is at work throughout cognition [28, p.90].

**Dimensionality Reduction** — A high-dimensional perceptual measurement space is advantageous because it may capture more of the useful structure of the problems that a cognitive system needs to deal with, such as categorization. Learning by “tiling” the representation space with examples is, however, infeasible in a high-dimensional space, because the number of required examples grows exponentially with dimensionality [34]. This necessitates dimensionality reduction prior to learning, which, moreover, needs to be done so as to lose as little as possible of the useful information.

**Embodiment and Situatedness** — Evolution fine-tunes the computations carried out by natural cognitive systems to the mechanics of the bodies that they control and the ecological niche in which they are situated [2].

**Function Approximation** — Learning from examples and generalization to new queries is equivalent to function approximation, a problem in which the values of an unknown function are given at a number of points in its domain and are used to form an estimate that can then support generalization [85].

**Graphical Models** — The relationships among a set of variables of interest to a cognitive system can be conveniently represented in the form of a directed graph, in which the vertices stand for variables (of which some may be observable and others hidden, corresponding to the properties of the world that need to be inferred from sensory data) and the edges — for probabilistic dependencies between pairs of variables [79]. One type of such model is the Bayes Network [82]. Graphical models map naturally onto the architecture of the brain [70].

**Holism** — The pattern of causal dependencies in a system of knowledge about the natural world is such that any two items may be potentially interdependent; in this sense, rich cognitive representations are holistic [87]. This property of world knowledge gives rise to serious algorithmic challenges in truth maintenance systems, where facts newly acquired through learning can potentially interact with, and cause the revision of, the entire knowledge base.

**Ill-Posedness** — Problems arising in perception, thinking, and action planning are typically ill-posed in the sense that they do not possess a unique solution (e.g., [6]). Function approximation, which is central to learning, is ill-posed because an infinite number of mappings may be consistent with a given set of input-output pairs, and so is probability density estimation. Such problems can be made well-posed by regularization.

**Joint Probability** — The most that can be learned about the world by observing or tracking a set of variables of interest is an approximation to their joint probability density (note that the problem of probability estimation is ill-posed). To learn the causal dependencies among the variables, one must go beyond mere observation and intervene on variables of interest [80].

**Kernels** — A family of mathematical methods that arise from measurements of similarity of two vectors and that are widely applicable in modeling cognition [60, 61, 94]. Formally, a positive definite kernel is a function of two arguments that represents an inner product (dot product) in some feature space.

**Learning and Learnability** — Most of the detailed knowledge about how the world works that animals with advanced cognitive systems need to master cannot be “squeezed” through the genomic bottleneck and must therefore be learned from experience. The field of machine learning has amassed a wealth of insights into the computational nature of this process, including constraints and limitations on learning and learnability (e.g., [109, 110]).

**Minimum Description Length (MDL) Principle** — The fundamental principle of the computational
theory of learning, due to Solomonoff [103], is that LEARNING is learning of regularities. It is derived from the observation that learning is only useful insofar as it supports generalization and that generalization is only possible if regularities are discovered in the observed data. A modern operationalization of this idea is the Minimum Description Length Principle of Rissanen [89], according to which regularities in the data are best captured by a representation that minimizes the sum of the description lengths of the code and of the training data under that code [49]. A related principle is that of SIMPLICITY [13].

NAVIGATION — Finding a route through a representation space, subject to certain CONSTRAINTS, is a paradigm for all sequential behaviors. Thus, in foraging, for instance, the SEARCH space represents the terrain in which the animal is situated; in planning, it may be a graph representing the space of possible solutions to the problem at hand; in language production, the graph would be a representation of the speaker’s knowledge of language [29].

OPTIMIZATION — A wide range of tasks in cognition, including perception, thinking (e.g., problem solving and decision making), and motor control reduce to SEARCHING a space of possible solutions for an optimal one [28]. Optimality in this context is imposed by various CONSTRAINTS, which may stem from the nature of the problem, from implementational considerations, from EVOLUTIONARY pressure, or from general requirements of tractability and uniqueness (as in REGULARIZATION).

PREDICTION — A true understanding of the world (e.g., one that takes the form of a CAUSAL PROBABILISTIC model) should allow the cognitive system to exercise FORESIGHT: to predict impending events and the consequences of its own actions [28]. Such capacity for prediction turns out to be a very general explanatory principle in cognition [15], which can be linked to other general principles, such as BAYESIAN probability theory.

QUANTUM PROBABILITY — Animal behaviors that involve probabilistic assessment of cues and outcomes are often strongly context- and order-dependent. Understanding such behaviors may require positing individual states that are superpositions (i.e., are impossible to associate with specific values), as well as composite systems that are entangled (i.e., that cannot be decomposed into their subsystems). The relevant theories are best expressed in terms of quantum probability postulates [86].

REGULARIZATION — A problem that is formally ILL-POSED in that it has no unique solution can be turned into a well-posed one by imposing external CONSTRAINTS on the solution space. One class of such constraints, which has a profound grounding in LEARNING theory, is regularization through smoothing, which is related to BAYESIAN probability, to statistical learning theory, and to the MAXIMUM LIKELIHOOD idea [40].

SIMILARITY — the most important ultimate use to which sensory data could be put involves estimating the similarity between two stimuli, which constitutes the only principled basis for GENERALIZATION of response from one stimulus to another, and therefore for any non-trivial LEARNING from experience [95] [26] [36].

TUNING — Neurons in animal brains, and neuron-like units in artificial distributed cognitive systems, are typically tuned to various features of the perceptual world, of motor behavior, or of the animal’s internal representational states [28] ch.3. Graded, shallow tuning, with a high degree of overlap between the profiles of adjacent units, is behind perceptual filters or CHANNELS that support hyperacuity. Because a tuned unit effectively represents the SIMILARITY between the actual and optimal stimuli, graded tuning also underlies VERIDICALITY [27].

UNCERTAINTY — “The percept is always a wager. Thus uncertainty enters at two levels, not merely one: the configuration may or may not indicate an object, and the cue may or may not be utilized at its true indicative value” [44]. The fundamental uncertainty in dealing with the world is the central motivation for the use of PROBABILISTIC representations and processes by cognitive systems.

VERIDICALITY — Perceptual representations based on CHANNEL CODING are provably veridical in that they faithfully reflect the SIMILARITY relationships of the represented items, such as visual objects [27].
WEIGHT LEARNING — In computational systems composed of simple, neuron-like elements, LEARNING typically proceeds by adjusting the weights of the synaptic connections, although the threshold of the non-linear transfer function that is part of the standard “formal neuron” can also be adjusted [17]. The rule for experience-based weight adjustment proposed by Hebb in 1949, according to which “neurons that fire together, wire together,” has now been recast and widely accepted as spike timing dependent plasticity, or STDP [12].

MAXIMUM LIKELIHOOD ESTIMATION (MLE) — According to the MLE principle, the parameters of a PROBABILISTIC model that is intended to reproduce a set of observations should be tuned so as to make the actual observed data set most likely [75].

S YNTAX — The system of PATTERNS and CONSTRAINTS that governs the composition of utterances in a natural language [84, 37].

DOBZHANSKY — “Nothing in biology makes sense except in the light of EVOLUTION” [24].

References


27


