# Computational [Principles of] Psychology*

Shimon Edelman
Department of Psychology
Cornell University
Ithaca, NY 14853
`https://shimon-edelman.github.io`

## Contents



*Psych 3140/6140. Spring semester, Tuesdays and Thursdays, 10:10am-11:25am. Version 0, January 6, 2025.

# 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 Appendix C.

### 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.

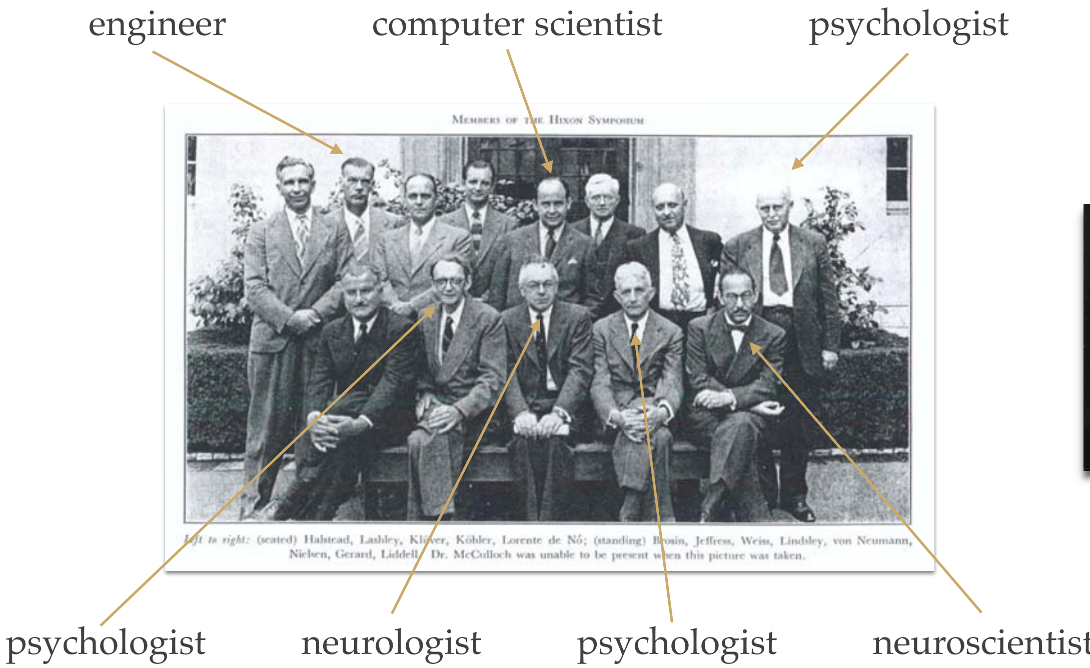
# 2 Notes for participants

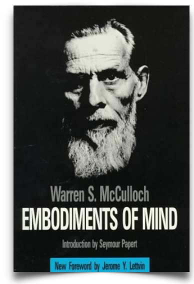
This section contains essential information for participants: the inclusion statement,[1] learning goals and practices, and credit requirements.

## 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

[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: `https://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; [78]), a still readable collection of papers by Warren McCulloch, the one member of the group who is absent from the group photo.

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.

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 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 Douglas Hofstadter described so alluringly in his Pulitzer-winning *Gödel, Escher, Bach* [55] (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 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 and participate by asking questions. Technical details (such as proofs of mathematical statements) and trivia (such as the names of researchers, with the exception of a few luminaries) are secondary. Please bring any concerns about learning practices to my office hours (or to the TA's).

## 2.4 Credit and grading

There are **no exams** in this course. To get credit, do:

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. **Weekly reading assignment: collaborative annotation**. By 11pm on each Friday, use the Perusall app, linked from the Canvas Assignment tab, to read and annotate the one PDF paper assigned for the corresponding week (in the weekly reading lists below, these papers are marked by ***). The collaborative reading assignments will be graded on a scale of 4 (excellent) to 0 (missing work). Complete up to 14 weekly collaborative readings; of these, the 10 best will count towards the final grade.

   *Guidelines for annotating the weekly collaborative readings*. When you first click on an assignment, you'll be taken to the Perusall page, which will display a list of best practices for collaborative annotation.

3. **Weekly writing/lecture assignment: micro-essay.** By 11pm on each Monday, submit via Canvas a micro-essay on the material covered in the preceding week's two lectures (see below). The essays will be graded on a scale of 6 (excellent) to 0 (missing work). Submit up to 14 micro-essays; of these, the 10 best will count towards the final grade. Please **note** that some of the topics are obligatory.

   The topic for each weekly essay is the material covered *in class* during the week preceding the due date. You should write the essay after attending the corresponding lectures. The writing prompt (motivated by a para-Buddhist parable) is the same for every week:

   (a) state and briefly motivate the **best question** you can think of pertaining to the corresponding lectures;

   (b) briefly opine on what would count as the **best answer** (or merely a good one) to your question.

   (c) For bonus points (up to the maximum score), include a brief speculation about how this week's material might be brought to bear on combatting anthropogenic global heating / climate catastrophe (Why? Because Cornell on Fire).

   For your reference, a sample bare-minimum micro-essay appears in Appendix A below.

   *Notes:* (1) Your essays will be run through Turnitin (please remember that copying and pasting from any source without attribution is a violation of academic integrity). (2) You may choose to have an "AI" (a bullshit generator, a.k.a. stochastic parrot, LLM, or Large Language Model) app such as ChatGPT write the essay in your stead. If you do so, you must indicate clearly which app/version was used to generate the text and what prompt you have used, and accompany it with a brief critique of the resulting essay. Omitting the critique would disqualify the submission, as would turning in a text that does not refer specifically to the material covered in class. (To help you decide whether or not to use an LLM, check out the papers listed in Appendix B.)

4. (Graduate students only, enrolled in Psych 6140.) In addition to the weekly readings and micro-essays, submit a written summary of your impressions and lessons from the material, in an essay form (about 1000 words), by noon, Tuesday, May 13, via email to the instructor.

Final grade components:

| | Psych 3140 | Psych 6140 |
|---|---|---|
| Weekly collab readings: | 40% | 30% |
| Weekly micro-essays: | 60% | 50% |
| Final essay: | – | 20% |

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 page 7. There is one page per week, with the "Primary" (required) and "Other" (optional) readings listed separately. The weekly collaborative readings are marked by ***. 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.

[This space intentionally left blank. The syllabus continues on the next page.]

# 3 Lecture topics by date

## 3.1 Week 1 (1/21; 1/23): 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 [59, sections 1,2]; Turing Machines [6].

1.2 The blind men and the elephant. Four case studies: an abstract computational theory [1]; a single-cell electrophysiology study [97]; an imaging study [52]; an engineering hack [79].

**Primary readings**

[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008, chapters 1,2.

***[52] **Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293: 2425–2430, 2001**.

**Other readings**

[120] Wigner, E. P. The unreasonable effectiveness of mathematics in the natural sciences. *Comm. Pure Appl. Math.*, XIII:1–14, 1960.

[6] Barker-Plummer, D. Turing machines. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. 2007. Available online at http://plato.stanford.edu/archives/win2007/entries/turing-machine/.

[97] Salzman, C. D., Britten, K. H., and Newsome, W. T. Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346:174–177, 1990.

[59] Hotton, S. and Yoshimi, J. Extending dynamical systems theory to model embodied cognition. *Cognitive Science*, 35:444–479, 2010, sections 1,2.

[1] Anderson, J. R. ACT: A simple theory of complex cognition. *American Psychologist*, 51:355–365, 1996.

Alumit Ishai

[79] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. Human-level control through deep reinforcement learning. *Nature*, 518:529–533, 2015.

### 3.2 Week 2 (1/28; 1/30): universal tools, I; methodology; levels of understanding; behavior

2.1 The nature of behavior. Dewey on the "reflex arc" concept [23]. Thurstone on the "stimulus-response" fallacy [113]. The life of behavior [48].

2.2 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 [66, 77, 53].

**Primary readings**

[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008, chapters 4,5.

[76] Marr, D. and Poggio, T. From understanding computation to understanding neural circuitry. *Neurosciences Res. Prog. Bull.*, 15:470–488, 1977.

[66] Joris, P. X., Smith, P. H., and Yin, T. C. T. Coincidence detection in the auditory system: 50 years after Jeffress. *Neuron*, 21:1235–1238, 1998.

[23] Dewey, J. The reflex arc concept in psychology. *Psychological Review*, 3:357–370, 1896.

[113] Thurstone, L. L. The stimulus-response fallacy in psychology. *Psychological Review*, 30:354–369, 1923.

***[48] **Gómez-Marín, A. and Ghazanfar, A. A. The life of behavior. *Neuron*, 104: 25–36, 2019**.

Alex Gomez-Marin

**Other readings**

[30] Edelman, S. Vision, reanimated and reimagined. *Perception*, 41:1116–1127, 2012. Special issue on Marr's *Vision*.

[32] Edelman, S. The minority report: some common assumptions to reconsider in the modeling of the brain and behavior. *Journal of Experimental and Theoretical Artificial Intelligence*, 28:751–776, 2016.

[53] Hazan, Y., Kra, Y., Yarin, I., Wagner, H., and Gutfreund, Y. Visual-auditory integration for visual search a behavioral study in barn owls. *Frontiers in Integrative Neuroscience*, 9(11):1–12, 2015

[39] Embar, K., Mukherjee, S., and Kotler, B. P. What do predators really want? The role of gerbil energetic state in determining prey choice by Barn Owls. *Ecology*, 95:280–285, 2014

[77] **Massa, C., Gabelli, F. M., and Cueto, G. R. Using GPS tracking to determine movement patterns and foraging habitat selection of the common barn-owl (Tyto alba). *Hornero*, 30:7–12, 2015**

Carolina Massa

### 3.3 Week 3 (2/4; 2/6): universal tools, II: probability; Bayes

3.1 A probabilistic formulation of cognition [15]. The Bayesian framework [50].

3.2 The Bayesian approach, applied to lightness perception [10, sections 1,2,4].

#### Primary readings

[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008, chapter 5; Appendix A.

***[15] Chater, N., Tenenbaum, J. B., and Yuille, A. Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, 10:287–291, 2006.

[50] Griffiths, T. L. and Yuille, A. Technical introduction: A primer on probabilistic inference. *Trends in Cognitive Sciences*, 10, 2006. Supplementary material. DOI: 10.1016/j.tics.2006.05.007.

[10] Brainard, D. H. and Freeman, W. T. Bayesian color constancy. *J. Opt. Soc. Am. A*, 14:1393–1411, 1997.

#### Other readings

[60] **Hurlbert, A. Colour constancy. *Current Biology*, 17:R906–R907, 2007**.

Anya Hurlbert

### 3.4 Week 4 (2/11; 2/13): universal tools, III: representation, similarity, generalization

4.1 Perceptual representation spaces [28, ch.5]. The face space [65, 44]. Multidimensional scaling (MDS).

4.2 Similarity and generalization [28, ch.5]. Shepard's Law [101, up to section "Mathematical Formulation" only (inclusive)]. Deriving Shepard's Law from the principle of efficient coding [104].

#### Primary readings

Alice O'Toole


[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008, chapter 5.

***[65] **Jiang, F., Blanz, V., and O'Toole, A. J. Probing the visual representation of faces with adaptation: A view from the other side of the mean. *Psychological Science*, 17:493–500, 2006**.

[101] Shepard, R. N. Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323, 1987.

[104] Sims, C. R. Efficient coding explains the universal law of generalization in human perception. *Science*, 360:652–656, 2018.

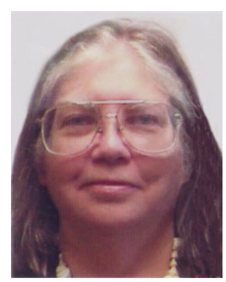

#### Other readings


[102] Shepard, R. N. How a cognitive psychologist came to seek universal laws. *Psychonomic Bulletin & Review*, 11(1):1–23, 2004.

[111] Tenenbaum, J. B. and Griffiths, T. L. Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, 24:629–641, 2001.

[31] Edelman, S. Varieties of perceptual truth and their possible evolutionary roots. *Psychonomic Bulletin and Review*, 22:1519–1522, 2015. doi: 10.3758/s13423-014-0741-z.

[103] **Shepard, R. N. and Chipman, S. Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1:1–17, 1970**.

Susan Chipman

### 3.5 Week 5 (2/18, 2/20): [February break]; universal tools, IV: veridical representation

5.1 [No class.]

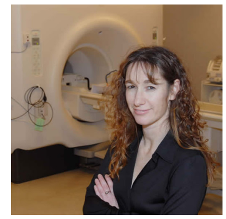
5.2 Veridical representation in perception [38, 19]. Representation is representation of similarities [26].

#### Primary readings


[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008, chapter 5.

[38] **Edelman, S., Grill-Spector, K., Kushnir, T., and Malach, R. Towards direct visualization of the internal shape representation space by fMRI. *Psychobiology*, 26:309–321, 1998**.

***[19] Cutzu, F. and Edelman, S. Faithful representation of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Science*, 93:12046–12050, 1996.

Kalanit
Grill-Spector

#### Other readings


[26] Edelman, S. Representation is representation of similarity. *Behavioral and Brain Sciences*, 21:449–498, 1998.

[27] Edelman, S. *Representation and recognition in vision*. MIT Press, Cambridge, MA, 1999.

### 3.6 Week 6 (2/25; 2/27): memory

6.1 Memory for things, places, and events [28, ch.6]; [114, 119].

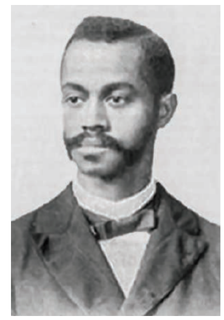
6.2 Associative memory [28, ch.6] and locality-sensitive hashing [3, 36].

#### Primary readings

[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008, chapter 6.

[114] **Turner, C. H. The behavior of a snake. *Science*, 30:563–564, 1909**.

[119] **Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., and Behrens, T. E. J. The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183:1–15, 2020. doi: 10.1016/j.cell.2020.10.024**.



Charles H. Turner

#### Other readings

[98] Samadi, H., Dona, G., and Chittka, L. Charles H. Turner, pioneer in animal cognition. *Science*, 370: 530–531, 2020. doi: 10.1126/science.abd8754.

[3] Andoni, A. and Indyk, P. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. *Communications of the ACM*, 51:117–122, 2008.

[36] Edelman, S. and Shahbazi, R. Renewing the respect for similarity. *Frontiers in Computational Neuroscience*, 6:45, 2012 (through section 7).

Guifen Chen

### 3.7 Week 7 (3/4; 3/6): actions and consequences

7.1 Managing action. The basics of motor control [121]. Bayesian motor decision making [69].

7.2 Action and reward. Reinforcement learning (RL) [123] and its relationship to Bayes. Hierarchical RL [8].

#### Primary readings

[121] Wise, S. P. and Shadmehr, R. Motor control. In Ramachandran, V. S., editor, *Encyclopedia of the Human Brain*, volume 3, pages 137–157. Academic Press, San Diego, CA, 2002.

***[69] Körding, K. P. and Wolpert, D. M. Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, 10:319–326, 2006.

[123] Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12:739–751, 2011.

#### Other readings

Yael Niv

[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008, chapter 6.

[122] Woergoetter, W. and Porr, B. Reinforcement learning. *Scholarpedia*, 3(3):1448, 2007.

[9] Botvinick, M. M., Niv, Y., and Barto, A. C. Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113:262–280, 2009.

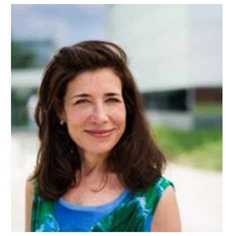
[110] **Solway, A., Diuk, C., Córdova, N., Yee, D., Barto, A. G., Niv, Y., and Botvinick, M. M. Optimal behavioral hierarchy. *PLOS Computational Biology*, 10(8):e1003779, 2014**.

### 3.8 Week 8 (3/11; 3/13): higher cognition, I: language

8.1 The structure of language [28, ch.7] and its use [25]. A functionalist approach to linguistic structure [47].

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

#### Primary readings



Adele Goldberg

[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008, chapter 7.

[25] Du Bois, J. W. Towards a dialogic syntax. *Cognitive Linguistics*, 25:359–410, 2014.

**[47] Goldberg, A. E. Subtle implicit language facts emerge from the functions of constructions. *Frontiers in Psychology*, 6:2019, 2016.**

***[33] Edelman, S. Language and other complex behaviors: unifying characteristics, computational models, neural mechanisms. *Language Sciences*, 62:91–123, 2017.

#### Other readings

[108] Solan, Z., Horn, D., Ruppin, E., and Edelman, S. Unsupervised learning of natural languages. *Proceedings of the National Academy of Science*, 102:11629–11634, 2005.

[68] Kolodny, O., Lotem, A., and Edelman, S. Learning a generative probabilistic grammar of experience: a process-level model of language acquisition. *Cognitive Science*, 39:227–267, 2015.

#### Selected recent readings on Large Language Models and "AI"

See Appendix B for an annotated selection of recent papers concerning Large Language Models and Artificial Intelligence.

### 3.9 Week 9 (3/18; 3/20): higher cognition, II: reasoning; induction

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

9.2 Induction [112].

#### Primary readings

[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008, chapter 8.

***[112] Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. How to grow a mind: statistics, structure, and abstraction. *Science*, 331:1279–1285, 2011.

#### Other readings

[86] **Pearl, J. Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv e-prints*, art. arXiv:1801.04016, January 2018**.

Judea Pearl

### 3.10 Week 10 (3/25; 3/27): higher cognition, III: intelligence; problem solving, analogy

10.1 General intelligence and IQ [49, 81, 5].

10.2 Problem solving [28, ch.8]. Analogy [57] and creativity [56].

#### Primary readings

[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008, chapter 8.

[56] Hofstadter, D. R. Variations on a theme as the crux of creativity. In *Metamagical Themas*, chapter 12, pages 232–259. Viking, Harmondsworth, England, 1985.

[57] Hofstadter, D. R. Analogy as the core of cognition. In Gentner, D., Holyoak, K. J., and Kokinov, B. N., editors, *The Analogical Mind: Perspectives from Cognitive Science*, pages 499–538. MIT Press, Cambridge MA, 2001.

[20] **Davis, H. E. *Variable Education Exposure and Cognitive Task Performance Among the Tsimane, Forager-Horticulturalists*. PhD thesis, University of New Mexico, 2014. URL `https://digitalrepository.unm.edu/anth_etds/17`.**

Helen E. Davis

***[5] Baker, D. P., Eslinger, P. J., Benavides, M., Peters, E., Dieckmann, N. F., and Leon, J. The cognitive impact of the education revolution: A possible cause of the Flynn Effect on population IQ. *Intelligence*, 49:144–158, 2015.

#### Other readings

[49] Gottfredson, L. S. Life, death, and intelligence. *Journal of Cognitive Education and Psychology*, 1: 23–46, 2004.

[81] Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., and Turkheimer, E. Intelligence: new findings and theoretical developments. *American Psychologist*, 2012.

[96] Ritchie, S. J. and Tucker-Drob, E. M. How much does education improve intelligence? A meta-analysis. *Psychological Science*, 29:1358–1369, 2018.

### 3.11 Week 11 (4/8; 4/10): intro to neural computation; neurons, I

11.1 Introduction to neural computation. Brains and neurons. Cortical receptive fields (RFs), maps [119], and hierarchies [28, ch.2,3]. Channel coding and hyperacuity [106].

11.2 What do neurons do?

Projection:

- neurons as readout devices [12];
- neurons, random projections, and similarity [36, 73].

Kernels:

- landmarks in representation spaces, similarity, and kernels [100].

**Primary readings**


[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008, chapters 2,3.

***[12] **Buzsáki, G. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, 68:362–385, 2010**.

[119] Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., and Behrens, T. E. J. The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183: 1–15, 2020. doi: 10.1016/j.cell.2020.10.024.


György Buzsáki


[36] Edelman, S. and Shahbazi, R. Renewing the respect for similarity. *Frontiers in Computational Neuroscience*, 6:45, 2012 (through section 7).


**Other readings**


[73] Lillicrap, T. P., Cownden, D., Akerman, C. J., and Tweed, D. B. Multilayer controllers can learn from random feedback weights. In *Proc. Symp. on Translational and Computational Motor Control (TCMC)*, pages 83–84, 2013. Satellite to the annual Society for Neuroscience meeting.

[100] Shahbazi, R., Raizada, R., and Edelman, S. Similarity, kernels, and the fundamental constraints on cognition. *Journal of Mathematical Psychology*, 70:21–34, 2016.

### 3.12 Week 12 (4/15; 4/17): neurons, II

12.1 What do neurons do?

Time-dependent dynamic learning:

- spike timing dependent plasticity (STDP), and Hebbian learning (see the Scholarpedia article [105]);
- history-dependent learning with the BCM rule [17];
- learning dynamic embodied control modes [22].

12.2 What do neurons do?

Population dynamics:

- neural trajectories and classification [11].

**Primary readings**

[17] Cooper, L. N. and Bear, M. F. The BCM theory of synapse modification at 30: interaction of theory with experiment. *Nature Reviews Neuroscience*, 13:798–810, 2012.

***[105] Sjöström, J. and Gerstner, W. Spike-timing dependent plasticity. *Scholarpedia*, 5(2):1362, 2010.

[11] Buonomano, D. V. and Maass, W. State-dependent computations: spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, 10:113–125, 2009.

**Other readings**


Sophie Deneve

[62] Izhikevich, E. M. and Desai, N. S. Relating STDP to BCM. *Neural Computation*, 15:1511–1523, 2003

[21] **Deneve, S. Bayesian spiking neurons I: Inference. *Neural Computation*, 20: 91–117, 2008**.

[61] Izhikevich, E. M. Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, 17:2443–2452, 2007.

[83] Peña, B. J. F. J. L. Owl's behavior and neural representation predicted by Bayesian inference. *Nature Neuroscience*, 14:1061–1067, 2011.

[22] Der, R. and Martius, G. Novel plasticity rule can explain the development of sensorimotor intelligence. *Proceedings of the National Academy of Science*, pages E6224–E6232, 2015.

### 3.13 Week 13 (4/22; 4/24): neurons, III; advanced topics I

13.1 What do neurons do?

Population dynamics (cont.):

- ongoing dynamics and chaotic itinerancy [94].

13.2 Bayes and the real world [99, 70].

**Primary readings**


[94] Rabinovich, M. I., Simmons, A. N., and Varona, P. Dynamical bridge between brain and mind. *Trends in Cognitive Sciences*, 19:453–461, 2015.

***[99] Sanborn, A. N. and Chater, N. Bayesian brains without probabilities. *Trends in Cognitive Sciences*, 20: 883–893, 2016. doi: 10.1016/j.tics.2016.10.003.


**Other readings**


[70] **Kwisthout, J., Wareham, T., and van Rooij, I. Bayesian intractability is not an ailment that approximation can cure. *Cognitive Science*, 35:779–784, 2011**.

[116] van Rooij, I. How the curse of intractability can be cognitive science's blessing. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., and Maglio, P. P., editors, *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, Austin, TX, 2015. Cognitive Science Society.


Iris van Rooij

### 3.14 Week 14 (4/29; 5/1): advanced topics II

14.1 Predictive processing and the free energy principle [45, 82, 16].

14.2 Affect and emotions [4, 42].

#### Primary readings


***[45] **Friston, K. J. The free-energy principle: a unified brain theory? *Nature Neuroscience*, 11:127–138, 2010**.

[4] Bach, D. R. and Dayan, P. Algorithms for survival: a comparative perspective on emotions. *Nature Reviews Neuroscience*, 18:311–319, 2017.


Karl Friston

#### Other readings


[16] Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36:181–204, 2013.

[82] Park, H.-J. and Friston, K. J. Structural and functional brain networks: from connections to cognition. *Science*, 342:1238411, 2013. doi: 10.1126/science.1238411.

[71] Kwisthout, J., Bekkering, H., and van Rooij, I. To be precise, the details don't matter: On predictive processing, precision, and level of detail of predictions. *Brain and Cognition*, 112:84–91, 2017.

[41] **Feldman Barrett, L. Psychological construction: the Darwinian approach to the science of emotion. *Emotion Review*, 5:379–389, 2013**.

[42] Fernandez Velasco, P. and Loev, S. Affective experience in the predictive mind: a review and new integrative account. *Synthese*, 2020. doi: 10.1007/s11229-020-02755-4.


Lisa Feldman Barrett

### 3.15 Week 15 (5/6): wrapping up

15.1 Some conclusions. Also: AMA (Ask Me Anything).

---

## Appendix A Sample micro-essay

Note: in a regular week, which has two lectures (unlike week 5, half of which is the February break), the essay must relate to both.
**WEEK 5: veridical representation.**

- *A sketch of the "best" question:* How does the claim that perception is veridical square with the existence of perceptual illusions?
- *A sketch of a possible "best" answer:* The only "enforcer" of veridicality in perception is evolution; and evolutionary pressure is always about ... rather than ... [Complete this answer as an exercise.]
- *Potential relevance to the climate catastrophe:* Humans find it hard to perceive slow changes and long-term trends at all (let alone veridically). Therefore, ... [Complete this answer as an exercise.]

---

## Appendix B Large Language Models and "AI"

The following papers may help you understand what LLMs do [54, 72]; why current generative "AI" lacks understanding [75, 89]; and whether or not training on texts or samples of behavior can in principle lead to human-level AI [117].

[72] Lappin, S. Assessing the strengths and weaknesses of large language models. *Journal of Logic, Language and Information*, 33:9–20, 2024. doi: 10.1007/s10849-023-09409-x.

The transformers that drive chatbots and other AI systems constitute large language models (LLMs). These are currently the focus of a lively discussion in both the scientific literature and the popular media. This discussion ranges from hyperbolic claims that attribute general intelligence and sentience to LLMs, to the skeptical view that these devices are nomore than "stochastic parrots." I present an overview of some of the weak arguments that have been presented against LLMs, and I consider several of the more compelling criticisms of these devices. The former significantly underestimate the capacity of transformers to achieve subtle inductive inferences required for high levels of performance on complex, cognitively significant tasks. In some instances, these arguments misconstrue the nature of deep learning. The latter criticisms identify significant limitations in the way in which transformers learn and represent patterns in data. They also point out important differences between the procedures through which deep neural networks and humans acquire knowledge of natural language.

[54] Hicks, M. T., Humphries, J., and Slater, J. ChatGPT is bullshit. *Ethics and Information Technology*, 26:38, 2024. doi: 10.1007/s10676-024-09775-5.

Recently, there has been considerable interest in large language models: machine learning systems which produce human-like text and dialogue. Applications of these systems have been plagued by persistent

inaccuracies in their output; these are often called "AI hallucinations". We argue that these falsehoods, and the overall activity of large language models, is better understood as *bullshit* in the sense explored by Frankfurt [43].

[75] Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., and Fedorenko, E. Dissociating language and thought in large language models. *Trends in Cognitive Sciences*, 28(6):517–540, 2024. doi: 10.1016/j.tics.2024.01.011.

Large language models (LLMs) have come closest among all models to date to mastering human language, yet opinions about their linguistic and cognitive capabilities remain split. Here, we evaluate LLMs using a distinction between formal linguistic competence (knowledge of linguistic rules and patterns) and functional linguistic competence (understanding and using language in the world). We ground this distinction in human neuroscience, which has shown that formal and functional competence rely on different neural mechanisms. Although LLMs are surprisingly good at formal competence, their performance on functional competence tasks remains spotty and often requires specialized fine-tuning and/or coupling with external modules. We posit that models that use language in human-like ways would need to master both of these competence types, which, in turn, could require the emergence of separate mechanisms specialized for formal versus functional linguistic competence.

[89] Pezzulo, G., Parr, T., Cisek, P., Clark, A., and Friston, K. Generating meaning: active inference and the scope and limits of passive ai. *Trends in Cognitive Sciences*, 28(2):97–112, 2024. doi: 10.1016/j.tics.2023.10.002.

Prominent accounts of sentient behavior depict brains as generative models of organismic interaction with the world, evincing intriguing similarities with current advances in generative artificial intelligence (AI). However, because they contend with the control of purposive, life-sustaining sensorimotor interactions, the generative models of living organisms are inextricably anchored to the body and world. Unlike the passive models learned by generative AI systems, they must capture and control the sensory consequences of action. This allows embodied agents to intervene upon their worlds in ways that constantly put their best models to the test, thus providing a solid bedrock that is – we argue – essential to the development of genuine understanding.

[117] van Rooij, I., Guest, O., Adolfi, F., de Haan, R., Kolokolova, A., and Rich, P. Reclaiming AI as a theoretical tool for cognitive science. *Computational Brain & Behavior*, 7:616–636, 2024. doi: 10.1007/s42113-024-00217-5.

The idea that human cognition is, or can be understood as, a form of computation is a useful conceptual tool for cognitive science. It was a foundational assumption during the birth of cognitive science as a multidisciplinary field, with Artificial Intelligence (AI) as one of its contributing fields. One conception of AI in this context is as a provider of computational tools (frameworks, concepts, formalisms, models, proofs, simulations, etc.) that support theory building in cognitive science. The contemporary field of AI, however, has taken the theoretical possibility of explaining human cognition as a form of computation to imply the practical feasibility of realising human(-like or -level) cognition in factual computational systems, and the field frames this realisation as a short-term inevitability. Yet, as we formally prove herein, creating systems with human(-like or -level) cognition is intrinsically computationally intractable. This means that any factual AI systems created in the short-run are at best decoys. When we think these systems capture something deep about ourselves and our thinking, we induce distorted and impoverished images of ourselves and our cognition.

## Appendix C Computational concepts, principles & methods: a selective 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 [57, 58]. Analogy is central to general cognitive function (general fluid intelligence, often referred to as *gF* or IQ [107]) 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 [67, 124, 35, 111, 88, 69].

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 [106]. 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 [91].

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 [84]. One type of such model is the BAYES Network [87]. Graphical models map naturally onto the architecture of the brain [74].

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 [93]. 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., 7). 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 [85].

KERNELS — A family of mathematical methods that arise from measurements of SIMILARITY of two vectors and that are widely applicable in modeling cognition [63, 64, 100]. 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., 115, 118).

MINUMUM DESCRIPTION LENGTH (MDL) PRINCIPLE — The fundamental principle of the computational theory of learning, due to Solomonoff [109], 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 [95], 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 [51]. A related principle is that of SIMPLICITY [14].

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 [16], 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 [92].

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 [101, 26, 36].

**T**UNING — 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].

**U**NCERTAINTY — "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" [46]. The fundamental uncertainty in dealing with the world is the central motivation for the use of PROBABILISTIC representations and processes by cognitive systems.

**V**ERIDICALITY — 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].

**W**EIGHT 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 [18]. 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 [13].

MA**X**IMUM 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 [80].

S**Y**NTAX — The system of PATTERNS and CONSTRAINTS that governs the composition of utterances in a natural language [90, 37].

DOB**Z**HANSKY — "Nothing in biology makes sense except in the light of EVOLUTION" [24].

# References


[1] Anderson, J. R. ACT: A simple theory of complex cognition. *American Psychologist*, 51:355–365, 1996.

[2] Anderson, M. L. Embodied cognition: A field guide. *Artificial Intelligence*, 149:91–130, 2003.

[3] Andoni, A. and Indyk, P. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. *Communications of the ACM*, 51:117–122, 2008.

[4] Bach, D. R. and Dayan, P. Algorithms for survival: a comparative perspective on emotions. *Nature Reviews Neuroscience*, 18:311–319, 2017.

[5] Baker, D. P., Eslinger, P. J., Benavides, M., Peters, E., Dieckmann, N. F., and Leon, J. The cognitive impact of the education revolution: A possible cause of the Flynn Effect on population IQ. *Intelligence*, 49:144–158, 2015.

[6] Barker-Plummer, D. Turing machines. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. 2007. Available online at http://plato.stanford.edu/archives/win2007/entries/turing-machine/.

[7] Bertero, M., Poggio, T., and Torre, V. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76: 869–889, 1988.

[8] Botvinick, M. M. Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, 12:201–208, 2008.

[9] Botvinick, M. M., Niv, Y., and Barto, A. C. Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113:262–280, 2009.

[10] Brainard, D. H. and Freeman, W. T. Bayesian color constancy. *J. Opt. Soc. Am. A*, 14:1393–1411, 1997.

[11] Buonomano, D. V. and Maass, W. State-dependent computations: spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, 10:113–125, 2009.

[12] Buzsáki, G. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron*, 68:362–385, 2010.

[13] Caporale, N. and Dan, Y. Spike timing-dependent plasticity: A Hebbian learning rule. *Annual Review of Neuroscience*, 31:25–46, 2008.

[14] Chater, N. and Vitányi, P. Simplicity: a unifying principle in cognitive science? *Trends in Cognitive Sciences*, 7:19–22, 2003.

[15] Chater, N., Tenenbaum, J. B., and Yuille, A. Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, 10:287–291, 2006.

[16] Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36:181–204, 2013.

[17] Cooper, L. N. and Bear, M. F. The BCM theory of synapse modification at 30: interaction of theory with experiment. *Nature Reviews Neuroscience*, 13:798–810, 2012.

[18] Cooper, L. N., Intrator, N., Blais, B. S., and Shouval, H. Z. *Theory of Cortical Plasticity*. World Scientific, Singapore, 2004.

[19] Cutzu, F. and Edelman, S. Faithful representation of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Science*, 93:12046–12050, 1996.

[20] Davis, H. E. *Variable Education Exposure and Cognitive Task Performance Among the Tsimane, Forager-Horticulturalists*. PhD thesis, University of New Mexico, 2014. URL `https://digitalrepository.unm.edu/anth_etds/17`.

[21] Deneve, S. Bayesian spiking neurons I: Inference. *Neural Computation*, 20:91–117, 2008.

[22] Der, R. and Martius, G. Novel plasticity rule can explain the development of sensorimotor intelligence. *Proceedings of the National Academy of Science*, pages E6224–E6232, 2015.

[23] Dewey, J. The reflex arc concept in psychology. *Psychological Review*, 3:357–370, 1896.

[24] Dobzhansky, T. Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, 35:125–129, 1973.

[25] Du Bois, J. W. Towards a dialogic syntax. *Cognitive Linguistics*, 25:359–410, 2014.

[26] Edelman, S. Representation is representation of similarity. *Behavioral and Brain Sciences*, 21:449–498, 1998.

[27] Edelman, S. *Representation and recognition in vision*. MIT Press, Cambridge, MA, 1999.

[28] Edelman, S. *Computing the mind: how the mind really works*. Oxford University Press, New York, NY, 2008.

[29] Edelman, S. On look-ahead in language: navigating a multitude of familiar paths. In Bar, M., editor, *Prediction in the Brain*, chapter 14, pages 170–189. Oxford University Press, New York, 2011.

[30] Edelman, S. Vision, reanimated and reimagined. *Perception*, 41:1116–1127, 2012. Special issue on Marr's *Vision*.

[31] Edelman, S. Varieties of perceptual truth and their possible evolutionary roots. *Psychonomic Bulletin and Review*, 22:1519–1522, 2015. doi: 10.3758/s13423-014-0741-z.

[32] Edelman, S. The minority report: some common assumptions to reconsider in the modeling of the brain and behavior. *Journal of Experimental and Theoretical Artificial Intelligence*, 28:751–776, 2016.

[33] Edelman, S. Language and other complex behaviors: unifying characteristics, computational models, neural mechanisms. *Language Sciences*, 62:91–123, 2017.

[34] Edelman, S. and Intrator, N. Learning as extraction of low-dimensional representations. In Medin, D., Goldstone, R., and Schyns, P., editors, *Mechanisms of Perceptual Learning*, pages 353–380. Academic Press, 1997.

[35] Edelman, S. and Shahbazi, R. Survival in a world of probable objects. *Behavioral and Brain Sciences*, 34:197–198, 2011. A commentary on *Bayesian Fundamentalism or Enlightenment? On the Explanatory Status and Theoretical Contributions of Bayesian Models of Cognition* by Jones & Love.

[36] Edelman, S. and Shahbazi, R. Renewing the respect for similarity. *Frontiers in Computational Neuroscience*, 6:45, 2012.

[37] Edelman, S. and Waterfall, H. R. Behavioral and computational aspects of language and its acquisition. *Physics of Life Reviews*, 4:253–277, 2007.

[38] Edelman, S., Grill-Spector, K., Kushnir, T., and Malach, R. Towards direct visualization of the internal shape representation space by fMRI. *Psychobiology*, 26:309–321, 1998.

[39] Embar, K., Mukherjee, S., and Kotler, B. P. What do predators really want? The role of gerbil energetic state in determining prey choice by Barn Owls. *Ecology*, 95:280–285, 2014.

[40] Evgeniou, T., Pontil, M., and Poggio, T. Regularization networks and support vector machines. *Advances in Computational Mathematics*, 13:1–50, 2000.

[41] Feldman Barrett, L. Psychological construction: the Darwinian approach to the science of emotion. *Emotion Review*, 5:379–389, 2013.

[42] Fernandez Velasco, P. and Loev, S. Affective experience in the predictive mind: a review and new integrative account. *Synthese*, 2020. doi: 10.1007/s11229-020-02755-4.

[43] Frankfurt, H. G. *On Bullshit*. Princeton University Press, Princeton, NJ, 2005.

[44] Freiwald, W. A., Tsao, D. Y., and Livingstone, M. S. A face feature space in the macaque temporal lobe. *Nature Neuroscience*, 12:1187–1199, 2009.

[45] Friston, K. J. The free-energy principle: a unified brain theory? *Nature Neuroscience*, 11:127–138, 2010.

[46] Gibson, J. J. Survival in a world of probable objects. *Contemporary Psychology*, 2:33–35, 1957.

[47] Goldberg, A. E. Subtle implicit language facts emerge from the functions of constructions. *Frontiers in Psychology*, 6:2019, 2016.

[48] Gómez-Marín, A. and Ghazanfar, A. A. The life of behavior. *Neuron*, 104:25–36, 2019.

[49] Gottfredson, L. S. Life, death, and intelligence. *Journal of Cognitive Education and Psychology*, 1: 23–46, 2004.

[50] Griffiths, T. L. and Yuille, A. Technical introduction: A primer on probabilistic inference. *Trends in Cognitive Sciences*, 10, 2006. Supplementary material. DOI: 10.1016/j.tics.2006.05.007.

[51] Grunwald, P. Introducing the minimum description length principle. In *Advances in minimum description length: theory and applications*, pages 3–22. MIT, Cambridge, MA, 2005.

[52] Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293:2425–2430, 2001.

[53] Hazan, Y., Kra, Y., Yarin, I., Wagner, H., and Gutfreund, Y. Visual-auditory integration for visual search a behavioral study in barn owls. *Frontiers in Integrative Neuroscience*, 9(11):1–12, 2015.

[54] Hicks, M. T., Humphries, J., and Slater, J. ChatGPT is bullshit. *Ethics and Information Technology*, 26: 38, 2024. doi: 10.1007/s10676-024-09775-5.

[55] Hofstadter, D. R. *Gödel, Escher, Bach: an Eternal Golden Braid*. Basic Books, New York, 1979.

[56] Hofstadter, D. R. Variations on a theme as the crux of creativity. In *Metamagical Themas*, chapter 12, pages 232–259. Viking, Harmondsworth, England, 1985.

[57] Hofstadter, D. R. Analogy as the core of cognition. In Gentner, D., Holyoak, K. J., and Kokinov, B. N., editors, *The Analogical Mind: Perspectives from Cognitive Science*, pages 499–538. MIT Press, Cambridge MA, 2001.

[58] Hofstadter, D. R. and Mitchell, M. The Copycat project: a model of mental fluidity and analogy-making. In Hofstadter, D. R., editor, *Fluid Concepts and Creative Analogies*, chapter 5, pages 205–265. Basic Books, NY, 1995.

[59] Hotton, S. and Yoshimi, J. Extending dynamical systems theory to model embodied cognition. *Cognitive Science*, 35:444–479, 2010.

[60] Hurlbert, A. Colour constancy. *Current Biology*, 17:R906–R907, 2007.

[61] Izhikevich, E. M. Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, 17:2443–2452, 2007.

[62] Izhikevich, E. M. and Desai, N. S. Relating STDP to BCM. *Neural Computation*, 15:1511–1523, 2003.

[63] Jäkel, F., Schölkopf, B., and Wichmann, F. A. Generalization and similarity in exemplar models of categorization: Insights from machine learning. *Psychonomic Bulletin & Review*, 15:256–271, 2008.

[64] Jäkel, F., Schölkopf, B., and Wichmann, F. A. Does cognitive science need kernels? *Trends in Cognitive Sciences*, 13:381–388, 2009.

[65] Jiang, F., Blanz, V., and O'Toole, A. J. Probing the visual representation of faces with adaptation: A view from the other side of the mean. *Psychological Science*, 17:493–500, 2006.

[66] Joris, P. X., Smith, P. H., and Yin, T. C. T. Coincidence detection in the auditory system: 50 years after Jeffress. *Neuron*, 21:1235–1238, 1998.

[67] Knill, D. and Richards, W., editors. *Perception as Bayesian Inference*. Cambridge University Press, Cambridge, 1996.

[68] Kolodny, O., Lotem, A., and Edelman, S. Learning a generative probabilistic grammar of experience: a process-level model of language acquisition. *Cognitive Science*, 39:227–267, 2015.

[69] Körding, K. P. and Wolpert, D. M. Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, 10:319–326, 2006.

[70] Kwisthout, J., Wareham, T., and van Rooij, I. Bayesian intractability is not an ailment that approximation can cure. *Cognitive Science*, 35:779–784, 2011.

[71] Kwisthout, J., Bekkering, H., and van Rooij, I. To be precise, the details don't matter: On predictive processing, precision, and level of detail of predictions. *Brain and Cognition*, 112:84–91, 2017.

[72] Lappin, S. Assessing the strengths and weaknesses of large language models. *Journal of Logic, Language and Information*, 33:9–20, 2024. doi: 10.1007/s10849-023-09409-x.

[73] Lillicrap, T. P., Cownden, D., Akerman, C. J., and Tweed, D. B. Multilayer controllers can learn from random feedback weights. In *Proc. Symp. on Translational and Computational Motor Control (TCMC)*, pages 83–84, 2013. Satellite to the annual Society for Neuroscience meeting.

[74] Litvak, S. and Ullman, S. Cortical circuitry implementing graphical models. *Neural Computation*, 21: 3010–3056, 2009.

[75] Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., and Fedorenko, E. Dissociating language and thought in large language models. *Trends in Cognitive Sciences*, 28(6):517–540, 2024. doi: 10.1016/j.tics.2024.01.011.

[76] Marr, D. and Poggio, T. From understanding computation to understanding neural circuitry. *Neurosciences Res. Prog. Bull.*, 15:470–488, 1977.

[77] Massa, C., Gabelli, F. M., and Cueto, G. R. Using GPS tracking to determine movement patterns and foraging habitat selection of the common barn-owl (Tyto alba). *Hornero*, 30:7–12, 2015.

[78] McCulloch, W. S. *Embodiments of mind*. MIT Press, Cambridge, MA, 1965.

[79] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. Human-level control through deep reinforcement learning. *Nature*, 518:529–533, 2015.

[80] Myung, I. J. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47: 90–100, 2003.

[81] Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., and Turkheimer, E. Intelligence: new findings and theoretical developments. *American Psychologist*, 2012.

[82] Park, H.-J. and Friston, K. J. Structural and functional brain networks: from connections to cognition. *Science*, 342:1238411, 2013. doi: 10.1126/science.1238411.

[83] Peña, B. J. F. J. L. Owl's behavior and neural representation predicted by Bayesian inference. *Nature Neuroscience*, 14:1061–1067, 2011.

[84] Pearl, J. Causal inference in statistics: A gentle introduction. In *Computing Science and Statistics: Proceedings of Interface '01*, volume 33, 2001. Avaliable online at `http://ftp.cs.ucla.edu/pub/stat_ser/R289.pdf`.

[85] Pearl, J. Causal inference in statistics: an overview. *Statistics Surveys*, 3:96–146, 2009.

[86] Pearl, J. Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv e-prints*, art. arXiv:1801.04016, January 2018.

[87] Pearl, J. and Russel, S. Bayesian networks. In Arbib, M. A., editor, *Handbook of Brain Theory and Neural Networks*, pages 157–160. MIT Press, Cambridge, MA, 2003. Available online at `http://ftp.cs.ucla.edu/pub/stat_ser/R277.pdf`.

[88] Perfors, A., Tenenbaum, J. B., and Regier, T. Poverty of the stimulus? A rational approach. In Sun, R. and Miyake, N., editors, *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, pages 663–668, Austin, TX, 2006. Cognitive Science Society.

[89] Pezzulo, G., Parr, T., Cisek, P., Clark, A., and Friston, K. Generating meaning: active inference and the scope and limits of passive ai. *Trends in Cognitive Sciences*, 28(2):97–112, 2024. doi: 10.1016/j.tics.2023.10.002.

[90] Phillips, C. Syntax. In Nadel, L., editor, *Encyclopedia of Cognitive Science*, volume 4, pages 319–329. Macmillan, London, 2003.

[91] Poggio, T. A theory of how the brain might work. *Cold Spring Harbor Symposia on Quantitative Biology*, LV:899–910, 1990.

[92] Pothos, E. M. and Busemeyer, J. R. Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences*, 36:255–327, 2013.

[93] Quine, W. V. O. Two dogmas of Empiricism. In *From a Logical Point of View*, pages 20–46. Harvard University Press, Cambridge, MA, 1953.

[94] Rabinovich, M. I., Simmons, A. N., and Varona, P. Dynamical bridge between brain and mind. *Trends in Cognitive Sciences*, 19:453–461, 2015.

[95] Rissanen, J. Minimum description length principle. In Kotz, S. and Johnson, N. L., editors, *Encyclopedia of Statistic Sciences*, volume 5, pages 523–527. J. Wiley and Sons, 1987.

[96] Ritchie, S. J. and Tucker-Drob, E. M. How much does education improve intelligence? A meta-analysis. *Psychological Science*, 29:1358–1369, 2018.

[97] Salzman, C. D., Britten, K. H., and Newsome, W. T. Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346:174–177, 1990.

[98] Samadi, H., Dona, G., and Chittka, L. Charles H. Turner, pioneer in animal cognition. *Science*, 370: 530–531, 2020. doi: 10.1126/science.abd8754.

[99] Sanborn, A. N. and Chater, N. Bayesian brains without probabilities. *Trends in Cognitive Sciences*, 20: 883–893, 2016. doi: 10.1016/j.tics.2016.10.003.

[100] Shahbazi, R., Raizada, R., and Edelman, S. Similarity, kernels, and the fundamental constraints on cognition. *Journal of Mathematical Psychology*, 70:21–34, 2016.

[101] Shepard, R. N. Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323, 1987.

[102] Shepard, R. N. How a cognitive psychologist came to seek universal laws. *Psychonomic Bulletin & Review*, 11(1):1–23, 2004.

[103] Shepard, R. N. and Chipman, S. Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1:1–17, 1970.

[104] Sims, C. R. Efficient coding explains the universal law of generalization in human perception. *Science*, 360:652–656, 2018.

[105] Sjöström, J. and Gerstner, W. Spike-timing dependent plasticity. *Scholarpedia*, 5(2):1362, 2010.

[106] Snippe, H. P. and Koenderink, J. J. Discrimination thresholds for channel-coded systems. *Biological Cybernetics*, 66:543–551, 1992.

[107] Snow, R. E., Kyllonen, P. C., and Marshalek, B. The topography of ability and learning correlates. In Sternberg, R. J., editor, *Advances in the psychology of human intelligence*, volume 2, pages 47–104. Erlbaum, Hillsdale, NJ, 1984.

[108] Solan, Z., Horn, D., Ruppin, E., and Edelman, S. Unsupervised learning of natural languages. *Proceedings of the National Academy of Science*, 102:11629–11634, 2005.

[109] Solomonoff, R. J. A formal theory of inductive inference, parts A and B. *Information and Control*, 7: 1–22, 224–254, 1964.

[110] Solway, A., Diuk, C., Córdova, N., Yee, D., Barto, A. G., Niv, Y., and Botvinick, M. M. Optimal behavioral hierarchy. *PLOS Computational Biology*, 10(8):e1003779, 2014.

[111] Tenenbaum, J. B. and Griffiths, T. L. Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, 24:629–641, 2001.

[112] Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. How to grow a mind: statistics, structure, and abstraction. *Science*, 331:1279–1285, 2011.

[113] Thurstone, L. L. The stimulus-response fallacy in psychology. *Psychological Review*, 30:354–369, 1923.

[114] Turner, C. H. The behavior of a snake. *Science*, 30:563–564, 1909.

[115] Valiant, L. G. A theory of the learnable. *Communications of the ACM*, 27:1134–1142, 1984.

[116] van Rooij, I. How the curse of intractability can be cognitive science's blessing. In Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., and Maglio, P. P., editors, *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, Austin, TX, 2015. Cognitive Science Society.

[117] van Rooij, I., Guest, O., Adolfi, F., de Haan, R., Kolokolova, A., and Rich, P. Reclaiming AI as a theoretical tool for cognitive science. *Computational Brain & Behavior*, 7:616–636, 2024. doi: 10.1007/s42113-024-00217-5.

[118] Vapnik, V. N. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10: 988–999, 1999.

[119] Whittington, J. C. R., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., and Behrens, T. E. J. The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183:1–15, 2020. doi: 10.1016/j.cell.2020.10.024.

[120] Wigner, E. P. The unreasonable effectiveness of mathematics in the natural sciences. *Comm. Pure Appl. Math.*, XIII:1–14, 1960.

[121] Wise, S. P. and Shadmehr, R. Motor control. In Ramachandran, V. S., editor, *Encyclopedia of the Human Brain*, volume 3, pages 137–157. Academic Press, San Diego, CA, 2002.

[122] Woergoetter, W. and Porr, B. Reinforcement learning. *Scholarpedia*, 3(3):1448, 2007.

[123] Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12:739–751, 2011.

[124] Yuille, A. and Kersten, D. Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, 10:301–308, 2006.