Spring 2018

updated: Apr. 15

Foundations of Cognitive Science

Cognitive Science 050.326/626

Professor:	Paul Smolensky	smolensky@jhu.edu	Krieger 241B	Monday 3–4
T.A.s:	Najoung Kim	n.kim@jhu.edu	Krieger 145	Thursday 1–2
	Matthias Lalisse	lalisse@jhu.edu	Krieger 155	Tuesday 3–4

Prerequisites: At least one course at the 300-level or higher in cognitive science, psychology, neuroscience, computer science, linguistics, or philosophy of mind.

Course goal: To introduce students to the broad spectrum within cognitive science of empirical methods, theoretical methods and frameworks, general conceptual issues, and major authors. The course is short on concrete content, long on understanding and integrating the basic perspectives on cognition provided by multiple disciplines. The key skill in focus is *logical argumentation*.

Course theme: Research on 'neural network' (or 'connectionist') computational models will provide one concrete context in which to consider a number of general course themes. Several cognitive domains are considered, but language is the one most frequently discussed.

Coursework: Reading (mostly major papers from the research literature); daily written work on readings; class attendance, and participation in class discussion.

Course Requirements

Attendance and participation in class discussion is extremely important, as this is an entirely discussion-based class. *Attendance will be taken.* Because of large class size, measures may be required to ensure that all students have a chance to participate.

The reading and writing demands of the course are quite high (undergraduate students receive 'W' credit for the course). Some readings (especially those towards the end of the course) are quite long and/or difficult so you are advised to plan ahead to complete readings by the assigned date.

For every reading (author), students will enter a write-up into Blackboard by the beginning of the class meeting during which that reading is to be discussed. For the student, the goal of this requirement is to gain experience in extracting and evaluating the logic of the arguments presented in cognitive science. For the class, the goal is to ensure that students have given sufficient thought to the day's reading prior to class to be able to contribute to a high-quality discussion. Thoughtful contribution to class discussion is an important part of the course and of the grade. Students may be called upon to contribute to discussions.

Write-ups fall into three categories: 'prose', 'diagram', and 'questions'.

'Prose' write-ups

The prose write-ups do not have a prescribed format, but they should by and large provide answers to the following sorts of questions:

(1) With what questions is the author mainly concerned?

(2) What are the author's main claims concerning these questions?

(3) What arguments are given to support these claims? Are they sound?

(4) On what methodologies are these arguments based?

(5) How do these claims relate to the major course issues?

(6) How do these claims relate to those of other authors you've read (in class or elsewhere)?

(7) How does the author's discipline shape their question and their argument? How might a similar question be addressed from the perspective of a different discipline?

Prose write-ups should NEVER reiterate the reading. The write-ups address the CONTENT of the paper, not the paper itself. (**Not** 'First the author talks about this and then she talks about that and I don't think

she did a good job talking about either one.') If directly addressed to the point, and concise, such write-ups may require only a single page (single spaced). Two to three (single-spaced) pages may be more realistic much of the time. One well-thought-out point is worth much more than a dozen superficial comments. When several readings are assigned for a single day, you must do a separate write-up for each reading, although on those occasions each write-up is expected to be more in the one- than the two-to-three-page range.

'Diagram' write-ups

The format for the 'diagram' write-ups is more prescribed. These are formal diagrams of the logic of the reading—or at least a key portion of it. These argument diagrams will be presented and discussed in class.

'Questions' write-ups

For a 'questions' write-up, you are to offer one or two good questions for discussion (per reading).

The 'questions' write-ups will be graded 1 or 0. For the 'prose' and 'diagram' write-ups, you will indicate at the top of each one either "for a grade" or "not for a grade". In the former case, the write-up will be graded on a 5-point scale; in the latter case, it will graded on a 3-point scale. Students are always welcome to bring their written work to the professor or TA's office hours for more detailed feedback.

Number of write-ups

The required number of write-ups is as follows (these are minimal requirements):

4 graded 'prose' write-ups: 1 before Feb 15, another before Spring Break, 2 after Spring Break

4 graded 'diagram' write-ups: 1 before Feb 15, another before Spring Break, 2 after Spring Break

6 ungraded 'prose' write-ups: 1 before Feb 15, another 2 before Spring Break, 3 after Spring Break

6 ungraded 'diagram' write-ups: 1 before Feb 15, another 2 before Spring Break, 3 after Spring Break

For each reading, a write-up is required; a 'questions' write-up will be handed in whenever a 'prose' or 'diagram' write-up is not handed in (thus 'questions' can be submitted for about a third of the readings).

Working together on any of the class assignments is strongly encouraged. For prose and questions write-ups, each student is expected to turn in their own work; **the other students worked with must always be identified** (or the work will be considered plagiarism). For diagrams, groups can submit a single diagram listing all group members.

The principal criteria for the grading of written work and class discussion in this class are (roughly in decreasing order of importance):

Identifying the main points
Correctly stating the main points
Identifying the main argument
Correctly stating the main argument's logic
Relating to the 'bigger picture', e.g.,
    course issues, covert aphorisms [see below]
    previous authors read in the course
    other knowledge outside the class
Providing an incisive, logically sound critique of the main point and/or argument

Final project (required for 626, optional for 326)**.** The final project can take the form of an essay, a diagram, a poster, or any other format approved by the professor. Students are strongly encouraged to work together; a single project may be submitted for the entire group. The purpose of the project is to address a set of course issues from the multiple perspectives represented in the readings. Since every other piece of written work focuses on a single reading, this is the students' opportunity to focus on relating the readings to one another, as they relate to specific common issues.

Late work/incompletes. Work handed in late will be marked down (one point per day); work more than a few days late will not be accepted. No incompletes will be given, except for documented medical emergencies. ***Cheating will be severely punished, following School procedures.***

GENERAL METHODOLOGICAL TOPICS

Research methods

Empirical methods	Theoretical methods
Introspection	Philosophical analysis
Linguistic generalization	Linguistic analysis
Empirical neuroscience	Computational and mathematical modeling
Experimental psychology	Theoretical psychology

Major distinctions and issues [and 'East-pole' position]

Theory- vs. data-centering

Should the primary goal of cognitive science be to construct a comprehensive, insightful theory or to collect and describe a comprehensive set of data? [theory]

Theories of brain function vs. theories of behavior vs. theories of mental knowledge

What is cognitive science the science of? [mental knowledge]

Levels of cognitive analysis

At what level does the explanation of cognitive phenomena lie: the level of individual neurons, or a more fine-grained, coarse-grained, or abstract level? [highly abstract level]

Explanation vs. description

What should be the descriptive or explanatory goals of cognitive science? [deep explanation, not superficial description]

Internal vs. external/functional explanation

Should cognitive explanation derive from principles internal to the cognitive system, or from external factors such as the function served by cognition or the physical constraints under which it operates? [internal explanation]

Formal vs. non formal theories

What is the appropriate level of formality for cognitive theory? Is human knowledge a formal system? [highly formal; yes]

Computational vs. non computational frameworks

Should a cognitive theory be a computational theory, viewing the mind as a machine that takes data in, processes it according to an algorithm, and outputs behavior? [yes]

Representational vs. non representational frameworks

Does cognition deploy representations of the outside world, on which it computes? [yes]

Competence vs. performance

Should cognitive science seek theories that explain the details of actual human performance, or theories of 'competence' — the knowledge underlying performance, ignoring the vagaries of particular instances of use of this knowledge? [competence first]

Models vs. theories

Should cognitive science strive to build models or theories of cognition, and how are these two related? [theories]

Mind as statistical processor vs. mind as structure processor

Is the mind primarily a device for processing highly structured, symbolic information, or primarily a device for performing statistical analysis of experience? [structure]

Nativism vs. empiricism

Does knowledge derive from experience? [not the crucial knowledge]

Abduction vs. data-driven induction

Is abstract knowledge generalized from experience (data-driven induction) or does experience select among *a priori* abstract hypotheses in the form of data-generating models (abduction)? [abduction]

Conscious vs. unconscious processes

What roles do conscious and unconscious processes play in an overall theory of cognition? [unconscious knowledge is key]

Explicit vs. implicit knowledge/rule-following vs. rule-governed

Is knowledge (e.g., a rule) explicitly encoded in the mind as a declarative statement *S* for a general-purpose interpreter, or implicitly encoded in a system whose behavior can be described by *S*? [implicit]

Knowledge/concepts as rules/definitions vs. as examples

In the mind, are concepts, or knowledge more generally, constituted of stored specific examples, or of general definitions or rules, as in mathematics? [definitions and rules]

Inference and decision making: logic based vs. non-rational approaches

Are human inferences and decisions based on some kind of logic? [yes]

Serial vs. parallel processing

What roles do serial and parallel processes play in an overall theory of cognition? [*abstract* seriality]

Independent mental faculties ('modules') vs. interactionism

Is the mind a collection of fairly autonomous faculties or modules, each concerned with some particular cognitive domain and governed by its own idiosyncratic principles, or is the mental processing of information from multiple domains so heavily interactive that decomposition into separate modules is not possible or useful? [modular]

Localized vs. distributed neural realization

Are localized bits of representations, knowledge, or processes realized in localized bits of the nervous system? [irrelevant]

Situated vs. non-situated cognition

Must a theory of cognition depend crucially on the way the mind is situated in the body and in the external social and physical world? [are you kidding?]

Covert aphorisms of cognitive science

Philosophy of science

Science = Data or Data ≫ Theory

Data ⇒ Experiments

Better theory = More data coverage

Serious constraint on theory ⇒ Data (∴ AI, Philosophy ⇒ BS)

Theory = Description

Theory = Empirical generalizations

Unmeasurable ⇒ Unscientific

Precise theory ⇒ Computer implementation

Precise theory ⇒ Mathematical formalism

Implausible (introspective intuition) ⇒ False

Explanation ⇒ External (e.g., functional) justification

Functional fable ⇒ Explanation

Internal explanation ⇒ Circular reasoning

Demystification ⇒ Trivialization/Denigration

Simplification ⇒ Irrelevance

Model ⇒ Theory

Substantive

Mind ⇒ Folk psychology

Mind = Brain

Cognition = Behavior

Thought ⇒ ¬Logic

Mind ⇒ ¬Formal system

Knowledge = Experience

Knowledge = Rules

Concept = Definition

Category = Categorization

¬[Computational ⇒ Intentional]

Level of representation = Level of cognitive computation

Connectionism ⇒ Associationism

Connectionism ⇒ Empiricism

Neurally informed ⇒ Neural model

Language = *X* (= Communication; Words; Speech recognition)

READINGS

All readings are available on the course Blackboard site.

Source books

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. MIT Press. [1/4 chapters]

* Cummins, R. & Dellarosa Cummins, D. 2000. Minds, Brains, and Computers: The Foundations of Cognitive Science. Blackwell. [7/35 chapters]

§ Pinker, S. & Mehler, J. eds. 1988. *Connections and Symbols.* MIT Press/Bradford Books. [2/3 chapters]

[out of print to bookstore but available via Amazon and to JHU students via cognet.mit.edu]

† Russell, B. 1945. *A History of Western Philosophy.* Simon & Schuster. [9/31 chapters]

Russell readings:

Title	Chapter	Pages:Paragraphs
[Plato:] The Theory of Ideas	XV	119–132
Plato's Theory of Immortality	XVI	138:2–140:3
Knowledge and Perception in Plato	XVIII	149–158
Aristotle's Logic	XXII	195–202
Descartes	IX	561:1–568
Locke's Theory of Knowledge	XIII	609:3–613:1
Hume	XVII	659–674
Kant	XX	706:3–708:1; 712:3–718:1

Reading Schedule (**N.B.:** Schedule as of Mar. 30. Subject to change.)

Date	Reading	Discipline
29-Jan	*Overview*	
31-Jan	McCarthy 1959 !, McClelland, Rumelhart & Hinton 1986 !	Comp Sci, Psych
5-Feb	Dennett 1984 [& pp. 3–5 of this syllabus], Plato (in Russell 1945†) !	Phil
7-Feb	Aristotle, Descartes (in Russell 1945†) !	Phil
12-Feb	Locke, Hume, Kant (in Russell 1945†) !	Phil
14-Feb	Skinner 1953, Chomsky 1959 !	Psych, Ling
19-Feb	Chomsky 1965 (Preface, Ch 1 !), Putnam 1967*	Ling, Phil
21-Feb	Saxe & Carey 2006	Psych
26-Feb	Hebb 1949*, Lashley 1950*	Psych, Psych
28-Feb	Caramazza 1986 ! , Marr 1982*	Psych, Comp Sci
5-Mar	Turing 1950*	Comp Sci
7-Mar	J. R. Anderson 2005 !	Comp Sci/Psych
12-Mar	Newell & Simon 1972	Comp Sci/Psych
14-Mar	Searle 1980	Phil
19-Mar	*Spring*	
21-Mar	*Break*	
26-Mar	Hofstadter 1985	
28-Mar	Rosch 1978	Psych
2-Apr	Lakoff 1987 !	Ling
4-Apr	Rumelhart & McClelland 1986 *!	Psych
9-Apr	Pinker & Prince 1988 §!	Ling
11-Apr	McCloskey 1991	Psych
16-Apr	Fodor & Pylyshyn 1988 §! [extra: Smolensky 1991* (+ *debate*)]	Phil
18-Apr	Smolensky & Legendre 2006 !	Phil & Psych
23-Apr	LeCun, Bengio & Hinton 2015; Marcus 2018	Comp Sci, Psych
25-Apr	*Course wrap-up*	
30-Apr	Griffiths et al. 2010; McClelland, Botvinick, et al. 2010	Psych
2-May	*TBD*	

Key:

! = note relevant page numbers given above under 'Readings' and below under 'References' !

† = pre-20th century philosophy: not primary reading; excerpts from B. Russell, *A History of Western Philosophy*; see preceding page for relevant page numbers

* = in the Cummins & Dellarosa Cummins (eds.) collection (*Minds, Brains, and Computers*)

§ = in the Pinker & Mehler (eds.) collection (*Connections and Symbols*)

REFERENCES

Anderson, J.R. 2005. Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science*, 29. 313-341. Skim Section 5; read the rest.

Caramazza, A. 1986. On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: The case for single- patient studies. *Brain and Cognition*, 5, 41-66. [read §§I – II, pp. 41–55:1.]

Chomsky, N. 1959. A review of B.F. Skinner's *Verbal Behavior*. *Language* **35**:26-58 [Sections 1-4,11; reprinted in Block, N. ed., 1980, *Readings in Philosophy of Psychology*, Harvard University Press, 48-63.]

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. MIT Press. Preface, pp. v-vii, and Chapter 1: Methodological Preliminaries, pp. 1-62. [read §§1-4, pp. 3-27; skim §§5-7, pp. 27-47; read §8, pp. 47-59; skip §9, pp. 60-62.]

Dennett, D. 1984/1998. The logical geography of computational approaches: A view from the East Pole. Reprinted in Dennett, D., *Brainchildren*, 215–234. MIT Press.

Fodor, J.A. & Pylyshyn, Z.W. 1988. Connectionism and cognitive architecture. *Cognition* 28, 3-71. [Reprinted in S. Pinker & J. Mehler (eds.), *Connections and Symbols*.] [Skim §§2.1, 2.2, 4; read pp. 67-69.]

Griffiths, T.L., Chater, N., Kemp, C., Perfors, A. & Tenenbaum, J.B. 2010. Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences* 14, 357–364.

Hebb, D.O. 1949. *The Organization of Behavior*. Wiley. Introduction, pp. xi-xix, Chapter 4: The first stage of perception: growth of the assembly, pp. 60-78. [From reprint, with new introduction, in J.A. Anderson & E. Rosenfeld eds., *Neurocomputing: Foundations of Research*, MIT Press/Bradford Books; pp. 43-56; read 43-44, 48-50:3, 53:5-54:5, 55:8-56.]

Hofstadter, D. 1985. *Metamagical Themas*. Basic Books. Chapter 26: Waking up from the Boolean dream, or, subcognition as computation, pp. 630-665.

Lakoff, G. 1987. *Women, Fire, and Dangerous Things*. University of Chicago Press. Excerpts: Preface, pp. xi-xvii; Ch. 1, pp. 5-11; Ch. 2, pp. 12-15, 56-57; Ch. 3, 58-67; Ch. 4, pp. 68-74; Ch. 21, pp. 370-373.

Lashley, K.S. 1950. In search of the engram. *Society of Experimental Biology Symposium, 4: Psychological Mechanisms in Animal Behavior*. Cambridge University Press. Pp. 454-455, 468-473, 477-480. [Reprinted, with new introduction, in J.A. Anderson & E. Rosenfeld eds., *Neurocomputing: Foundations of Research*, MIT Press/Bradford Books; pp. 57-63.]

LeCun, Y., Bengio, Y. & Hinton, G.E. 2015. Deep learning. *Nature*, 521, 436-444.

Marcus, G. 2018. Innateness, AlphaZero, and Artificial Intelligence. arXiv 1801.05667.

Marr, D. 1982. *Vision*. W. H. Freeman and Company. pp. 19-38, 329-332. *Excerpt pp. 69-83.

McCarthy, J. 1959. Programs with common sense. *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*. [Skim Section 1; Section 4 is optional.]

McClelland, J.L., Botvinick, M.M., Noelle, D.C., Plaut, D.C., Rogers, T.T., Seidenberg, M.S. & Smith, L.B. 2010. Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, 14, 348–356.

McClelland, J.L., Rumelhart, D.E., & Hinton, G.E. 1986. The appeal of parallel distributed processing. In D.E. Rumelhart, J.L. McClelland & The PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition; Vol. 1: Foundations*, 3-44. [read 3 -13, 20-25]

McCloskey, M. 1991. Networks and theories: The place of connectionism in cognitive science. *Psychological Science*, 2, 387-395.

Newell, A. & Simon, H.A. 1972. *The theory of human problem solving*. Prentiss-Hall, from Chapter 14, 787-823. [from reprint in Collins & Smith eds., pp. 33-51.]

Pinker, S. & Prince, A. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28, 73-194. [Reprinted in S. Pinker & J. Mehler (eds.), *Connections and Symbols*.] [Skim §4, skip §7 and appendix; leaves 47 pp.]

Putnam, H. 1967. The 'innateness hypothesis' and explanatory models in modern linguistics. *Synthese* 17, 12-22

Rosch, E. 1978. Principles of categorization. In E. Rosch & B.B. Lloyd eds., *Cognition and Categorization*, Erlbaum, pp. 27-48. [Reprinted in Collins & Smith eds., pp. 312-322.]

Rumelhart, D.E. & McClelland, J.L. 1986. On learning the past tenses of English verbs. In J.L. McClelland, D.E. Rumelhart, & The PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition; Vol. 2: Psychological and Biological Models*, pp. 216-271 [Skim pp. 226:1-233:0, 234:3-238:1, 248:1-260; skip the Appendix].

Saxe, S. & Carey, R. 2006. The perception of causation in infancy. *Acta Psychologica*, 123, 144-165.

Searle, J.R. 1980. Minds, brains, and programs. In *The Behavioral and Brain Sciences*, 3, 417-424. [Reprinted in Mind Design II, John Haugeland, Ed., pp. 183-204.]

Sejnowski, Terrence J. and Rosenberg, Charles R. 1987. Parallel networks that learn to pronounce English text. *Complex Systems* 1.

Skinner, B.F. 1953. *Science and Human Behavior*. Macmillan. Excerpts: pp. 27-35, 62-66, 87-90. [Reprinted in Block, N. ed., 1980, *Readings in Philosophy of Psychology*, Harvard University Press, pp. 37-47.]

Smolensky, P. 1988. On the proper treatment of connectionism. *The Behavioral and Brain Sciences*, 11, 1-23.

Smolensky, P. 1991. Connectionism, constituency, and the language of thought. In B. Loewer & G. Rey (Eds.), *Meaning in Mind: Fodor and his Critics.* Oxford: Basil Blackwell. 201–227.

Smolensky, P., & Legendre, G. 2006. Harmony optimization and the computational architecture of the mind/brain. In P. Smolensky & G. Legendre, *The harmonic mind: From neural computation to optimality-theoretic grammar*, Vol. 1, 2–61. Cambridge, MA: MIT Press. [Chapter 1; read pp. 2-47; reading boxes is optional (approx. 16 pp.)]

Turing, A.M. 1950. Computing machinery and intelligence. *Mind*, 59, 433-460. [Reprinted in Collins & Smith eds., pp. 6-19.]