Philosophy 3456
Foundations of AI

Background Notes

Stanford University
Winter Quarter, 1989–90

Description:	Background notes, issues, reading list, etc. for Spring 1990 Philosophy of AI course
File path:	… : …
Copyright:	© 1990 Brian Cantwell Smith
Last edited	Friday, January 19, 1990 — 8:44:05 AM
Last visited:	1999-11-15 — 9:37 PM

- **Morals, claims, and other intellectual points**

  - Cognitive science (cognitive AI)
    - Vulnerable to the underlying model of computation
    - These are people we're studying: hold true of friends & lovers
  - Computation
    - We don't yet really know what it is
    - Empirical inquiry to find out: do justice to practice
  - Intentionality
    - Subject matter in its own right (related, but ≠, to computation or cognition)
    - Non-effectiveness of semantic relations
    - Two fundamental issues:
      a) Reach (reference) — non-effective
      b) Registration (meaning) — aspectual
    - Different possible base cases:
      - Information (Dretske, B&P, Rosenchein, etc.)
      - Representation
    - Mention (but don't take on here) other (especially naturalistic) sources of intentionality in surrounding intellectual territory:
      - Biological function (Milikan);
      - Physical entropy notions of information;
      - Consciousness (Searle);
      - Etc.
  - Conceptual schemes / registration
    - What
  - Theory
    - (Type/token) reduction, supervenience, etc.
    - Explanation: not necessarily causal (naturalism)
    - Relation of theorist's and subject's registration schemes
    - What's to be explained? How much. Origins.
    - When dealing with intentional phenomena: no type distinction among theory, model, subject matter (cf. Kepler)
  - Miscellaneous morals
    - That representing and manifesting states of affairs are (intentionally) different. E.g., probabilities, digitality, etc.
    - Correspondence continuum

- **Assigned readings**

1 ⇒ 2: Introduction/Lay of the Land/The intentional challenge
- Haugeland "Semantic Engines"
- First part of B&E's "model theoretic semantics"
- Searle's "Minds, Brains, and Programs"
- Turing's original article
- Dennett's "Intentional Systems"
- Newell & Simon's Turing award lecture: "Computer science as empirical inquiry: symbols and search"
- Taylor's "Cognitive Psychology"

2 ⇒ 3: Computation in the wild
- Handout from the Encyclopedia of Computer Science
-

3 ⇒ 4: Digital state machines
- Chapter 4 of Goodman's Languages of Art
- Chapter ?? of Haugeland's AI: The Very Idea
- Haugeland's "Analog and Analog"
- Lewis' "Analog and Digital"
- First few pages of introduction to Mead's Analog VLSI

4 ⇒ 5: Formal symbol manipulation
- Fodor's "Methodological solipsism"
- Newell and Simon's "Physical Symbol System"
- Newell's "Knowledge Level"

5 ⇒ 6: Recursive function theory
-

6 ⇒ 7: The mind-body problem for machines
- Searle's "reduction, supervenience, and emergent properties" chapter
- Smith's "Correspondence Continuum"
- Haugeland's "Weak Supervenience"
-

7 ⇒ 8: Connectionism and the rise of concepts
- Cussins' "C3"
- Smolensky's "On the proper treatment of connectionism"
- Fodor & Pylyshyn's reply to Smolensky: "Connections and cognitive architecture: a critical analysis"
- Clark, Andy "Connectionism and Cognitive Science"
- McLelland, Rumelhardt, et al: chapters 2 ("A general framework for parallel distributed processing") and 4 ("PDP models and general issues in Cognitive Science").

- **Intellectual points**

— Strong/weak AI, etc.
  — Ivan Blair: "Intentionality, Mind, & Matter"

- Modelling, simulation, etc.
  - Itself an intentional notion
  - Challenge: AI has nothing particular to say about
  - What metric of equivalence?
- Implementation
  - Not an understood notion
  - Things don't cross implementation boundaries
  - …
- Concepts
  - Divide theorists into two broad classes
    a. Conceptual: accept a registration scheme (conceptual scheme) as a theoretical parameter. Analyse thinking as inter- and supra-conceptual; take concepts as unanlysed primitives. Exemplars: logic, Barwise, Fodor
    b. Pre-conceptual: concepts (constituency, formation, etc.) are part of the phenomena to be explained. Not necessarily similar across agents, time, etc. Exemplars: Smolensky, Cussins, …
- Computation claim on mind
  - Distinguish two views:
    a) that computation is tricky, more interesting than is realised (mine);
    b) that computation is FSM, but people are much more than that (Berkeley school — Dreyfus, Haugeland, Searle)
- Computation
  - Notion of effective computability
- Semantics
  - Triple semantics
  - Identifiablity of subject domain and machine (under a quotient) only if process-world relation isn't contextually dependent.
  - "Procedural" semantics
  - Two lessons of logic
  - Betrays your metaphysics
  - original/derivative
  - two factor accounts
- Shanker article

- **General notes, pedagogical points**

  - Foundations vs. architectures
    - There are really two computational "promises": the practical one, of intelligent behaviour, and the intellectual one, of explaining intentionality. Course will concentrate on the latter. Discussions of architecture tend to focus more on the former. Such a course would include such issues as:
      - Role of logic, debate between McCarthy and Newell & Simon, etc.
      - Role of the meta-level
      - Connectionism
      - Semantic networks
      - …

  - — One possibility would be to give a brief history of all this at some point.
- Assumptions
  - — That people know some computer stuff, and AI.
- Things that could (should?) be included in the "intentional challenge" lecture:
  - — Some of the registration/theorist stuff
  - — Early stuff about models, model theory, etc.
  - — First hints about different types of characterisation
  - — …
- Equip people to read rest of the literature. I.e., if you look over the bibliography, two questions you could ask:
    - a) What's the minimal amount someone should have read, so as to know what's going on in the foundations of AI?
    - b) What do I need to lead people through (and teach them), so that they could then read the rest of it on their own?
  - — Not a simple choice, but I will lean towards the second.
- People outside AI seem to concentrate on the model of computation, whereas those inside seem to focus on particular architecture or architectural families (logicist, connectionist, etc.).
- What am I going to do about the acceptance of a conceptual scheme? (Cf. Vinod's interest in notationality.) [This should be a paper.]

- **Organisational issues**

  - One possible division:
    1) Lay of the land
        - a. Introduction [1]
        - b. The intentional challenge [2]
    2) The philosophy of computation
        - a. Computation in the wild [3]
        - b. Digital state machines [4]
        - c. Formal symbol manipulation [5]
        - d. Effective computability [6]
    3) Foundational issues in AI
        - a. The mind/body problem for machines [7]
        - b. Connectionism & the rise of concepts [8]
        - c. The semantics of computation [9]
    4) An alternative account [10]
  - Possible mergers:
    - — Semantics of computation/comp'n in the wild
    - — Alternative account: could be a make-up or extra class (Tuesday night of last week).

- **Substantive issues**

  - — Strong vs. weak AI
  - — Turing machines
    - — Turing equivalence is weak: behaviourist

- Cascaded correspondence
- Visual vs. propositional representation (Kosslyn, Hayes, Block, Dennett, etc.)
- Digitality
- Formality
  - Universal allegiance, but no single reading
  - Antisemantical reading
    - motivation (disconnection)
- Programming
  - Program, process, interpreter, compiler, architecture, algorithm, implementation, [instantiation], (material) realisation, etc.
  - Triple semantics
- Two-factor accounts: causal/functional role, and reference.
- Transducers, sensory processing, "symbol grounding". "Semantical" vs. "physical" transducers. Alignment story.

- **Possible lectures**

  - Topics
    - **Part I: Philosophy of computation**
      - **Computation in the wild:** Programs, processes, interpreters, compilers, specification, architecture, realisation, implementation, etc.
      - **Digital state machines:** Digitality: discrete, continuous, indefinite, precision vs. rigour, second-order discrete, etc. Continuous programming languages. Analog VSLI. [problem: arcane]
      - **Effective computability**: Computability, complexity. Representational assumptions about the tape. Isomorphism to subject matter. "Decision" in decidability results. ⇒ branch of "materialism".
      - **Intentionality 1**: general introduction (summarise gauntlet). Reach and registration, disconnection (¬effectiveness), lessons of logic, assumed conceptual scheme. How do you see chairs, not light waves? [problem: metaphysically inaccessible to idealists]
      - **Intentionality 2**: computational models. Programming language accounts. Two-factor accounts. Triple semantics. Contextual dependence.
    - **Part II: Philosophy of cognitive science**
      - **The outside world**: modelling, simulation, etc. The closed world assumption. Contextually based semantics.
      - **Antisemantics**: Formal symbol manipulation, …
  - While "computation in the wild" isn't logically first, it might be the best one to lead with, since we can use it to generate questions that will subsequently be looked at in more detail. E.g.: spec-program-interior-world relations, internal architecture (itself symbolic?), levels of analysis, transducers, etc. Also inadequacy of current theories. Functional languages vs. imperative, etc. Fact that we don't understand Turing machines. I.e., an phenomenological map of subsequent intellectual territory.

- **Questions**

— Should I explicitly identify the three criteria that I want for a theory of computation? No: unfair, since that presumes a stance on strong AI, for example. But worth identifying, nonetheless. Maybe there are more.

— Go through AI's public critics & commentators; for each, identify which points are being: made, assumed, confused, ignored, etc.

- **People and their views**

— Relevance legend:
 • → specifically AI or computation;
 € → cognitive psych or intentionality, but not computational.

— Have a line on:

€ **Barwise & Perry**: Take meaning and information (unevaluated & evaluated versions, respectively) as basic intentional notions. Classification. Logic writ broad. Conceptual, non-psychologistic. Problems: pan-intentionalist.
• **Dennett**: instrumental ascription of content
• **Dreyfus**: traditional computation;
• **Fodor**: formal symbol manipulation (RTM, CTM), conceptual level of analysis.
• **Smolensky, Paul**: only moderately sophisticated philosophically, but one of the few people thinking seriously about connectionism, especially its conceptual consequences
• **Cussins, Adrian**:
• **Peacock, Christopher**:
• **Haugeland**: one of the few philosophers of computation. Automatic formal digital system with an interpretation. Hermeneutic bent.
• **Searle**: interest in high-level cognition; methodologically reductionist (causal story), but tenacity for real mental phenomena (consciousness). Computationally naïve.
• **Stich**: Buys formal syntactic line on computation, (bravely) explores consequences for cognitive science. Concludes folk and scientific psychology are incompatible. So much the worse for folk. Me: since view of computation is untenable, all rather academic.
€ **Milikan, Ruth**: Biological function
• **Winograd, Terry**: "Rationalistic" understanding of computation, but not of its use. Hermeneutically-inspired stance towards cognition, laced with an idealist (social solipsist) metaphysics.
• **Hayes, Pat**: …
• **McCarthy, John**: …
• **Minsky, Marvin**: …
• **Harel**:
• **Schank, Roger**: …
• **Jackendoff, Ray**: … (MIT school)
• **Anderson**: …
• **Rumelhardt, David**: …
€ **Lakoff, George**: …
• **Hofstadter, Douglas**: …
• **Scott, Dana**: …
• **Turing, Alan**: …
• **Craik**: …

  - **Marr, David**: …
  - **xxx**: …
  - **xxx**: …
  - **xxx**: …
  - **xxx**: …
  - **xxx**: …
  - **xxx**: …
  - **xxx**: …
  - **xxx**: …
  - **xxx**: …

— Don’t (yet) have a line on:

€ **Goodman, Nelson**: … (languages of notation, etc.)

  - **Boden, Margaret**: …
  - **Churchland, Patricia**: … (neuropsychology)
  - **Churchland, Paul**: … (radical naturalism?)
  - **Johnson-Laird**: … (traditional cognitive psych, modelling)
  - **Putnam, Hilary**: …
  - **Pylyshyn, Zenon**: …
  - **Quine, V. W. O.**: …

€ **Sellars, Wilfrid**: …

€ **Suppes, Pat**: …

  - **Taylor, Charles**: …

€ **Davidson, Donald**: …

  - **Lewis, David**: …
  - **Kripke, Saul**: …
  - **Neri-Casteneda, Hector**: …
  - **Newell, Alan**: …

— Other name Niklas (and others) have suggested

[— ***get from handout***]

  - **xxx**: …
  - **xxx**: …
  - **xxx**: …
  - **xxx**: …
  - **xxx**: …

—

- **To be read**

— New

— Pylyshyn: skim Computation & Cognition; check if a BBS précis.
— Johnson-Laird:
— Dennett, Dan: The Intentional Stance, & Beyond Belief
— Boden: … (book of lectures)
— Churchland, Paul: Mind, Matter, and Consciousness (???)
— Minsky, Marvin: Society of Mind
— Churchland, Pattie: Neurophilosophy

- — Newell, Alan: new book
- — Marr, David: (at least the introduction to) Vision
- — Richards, Whitman (ed): Natural Computation
- — Jackendoff, Ray: Computational Theory of Mind

— Re-read

- — Dreyfus: introduction to second edition of What Computers Can't Do
- — Harel: Algorithmics (as an example of "algorithm" stance towards computation)
- — Newell, Alan: "Physical Symbols Systems" and "The Knowledge Level"
- —

- **Reading list (by topic):**

- — Strong/weak AI
    - — Intro to Johnson Laird?
    - — beg. chapters of Pylyshyn's book
    - — Simon
    - — Searle's Minds, Brains, and Programs
- — Formality and formal symbol manipulation
    - — Fodor's "Methodological Solipsism"
    - — Newell's "Physical Symbol Systems" and "Knowledge Level"
    - —
- — Functionalism: part 3 of Block (1980)
- — Imagery: part 2 of Block (1981)
- — ST
    - — Chapters 2&3 of SILogic
    - — L&P interview
    - — first few chapters of S&A.
- — Digitality
    - — Exerpt from Goodman's Languages of Art
    - — Exerpt from Haugeland's AI: The Very Idea
    - — Haugeland's "Analog and Analog"
    - — Lewis' "Analog and Digital"
    - — Exerpt from introduction to Mead's Analog VLSI
- — Supervenience, type/token reductionism, etc.
    - — Searle: chapter in new book
    - — Churchland: chapter in Neurophilosophy
    - — Kim: <what was handed out>
    - — Haugeland: Weak Supervenience
    - — Putnam: article in MD
- — Connectionism
    - — Cussins' "C3"
    - — Smolensky: "On the microstructure of cognition"
    - — <book in which Fodor & Pylyshyn's reply is printed>
    - — Clark, Andy
    - — McLelland, Rumelhardt, et al: 3rd chapter (?) of PDP
- — Original(autonomous)/derivative semantics

    - Searle?
- Logico-deductive approach
    - McCarthy?
    - Hayes: “In Defense of Logic”
    - Israel: “What’s wrong with Non-monotonic Logic?”
    - McDermott: Computational Intelligence piece
    - Moore, Robert: “The Role of Logic in Knowledge Representation and Commonsense Reasoning”
- Dual-component/two-factor views
    - Colin McGinn
    - Ned Block
- Functionalism
- The Frame Problem
    - Fodor: “ … and the Music of the Intellectual Spheres”
    - Haugeland: <talk given at CSLI>
    - <check with Pat Hayes>
- Levels of analysis
    - Marr’s: 3 levels
    - logic: syntax/semantics/model theory
    - Newell: “Knowledge Level”
    - B&P: indirect classification
    - Two factor accounts
    - Smith: Φ/Ψ
- The new AI
    - Rosenschein
    - Agre and Chapman
    - Brooks
- Modelling
    - Intro to Rosen’s Anticipatory Systems
    - Cati, “Minds & Brains”
- On the relationship between scientific and folk psychology
- Other
    - Neisser
    - Winsett & Boyd: realism in biology
    - Kugler
    -

- **Notions**

    - (In)definite
    - (Semantic) reach
    - Algorithm
    - Analog
    - Artificial world
    - Aspect(ual)
    - Closed world assumption

- Compiler
- Computation
- Concept
- Conceptual scheme
- Context(ual dependence)
- Continuous
- Correspondence
- Digital
- Digital state machines
- Discrete
- Dual-component
- Effectiveness
- Embedded
- Embodied
- Formal symbol manipulation
- Formality condition
- Implementation
- Indexicality
- Information
- Information processing
- Instantiation
- Intentionality
- Intepreter
- Learning
- Meaning
- Medium independence
- Model
- Model theory
- Process
- Program
- Realisation
- Recursive function theory
- Reduction
- Reference
- Registration
- Representation
- Semantics
- Simulation
- Situated
- State
- Strong/weak AI
- Supervenience
- Symbol
- Token
- Transducer
- Triple semantics

- Turing machine
- Turing test
- Type
- Virtual machine

——end of file ——��