 Last edited Thursday, February 1, 2001

# Empirical Critique
## (of the ontological sub-reading of the negative reading of formality)

## I. Preliminaries

A. Administrivia
   1. Problem set #3 is now posted on the web site
   2. Note: the due date is February 13—12 days from now. (It looks as if we are settling into a pattern of one problem set every second week.)
   3. I am still behind on posting lecture notes, and grading the first two. However, I've reserved time this weekend to get caught up on that. So by next week we should be caught up).

B. Review
   1. Last Thursday, we distinguished *positive* and *negative* readings of the 'formality' predicate, in the "formal symbol manipulation" construal of computing.
   2. On Tuesday, we examined the negative—i.e., "independent of semantics"—sub-reading of formality, claiming that this was (historically and conceptually) the most important.
   3. We saw that there were two versions of it
      a. **Conceptual** (or "vertical"): computation can be *characterised in non-semantical terms*—a reading motivated by a background desire to naturalise computing; and
      b. **Ontological** (or "horizontal": that computation *works* independently of the exemplification of semantical properties—a reading based on intuitions of semantic *disconnection*.
   4. We saw, too, that these readings are not only different, but in tension: the intuition underwriting the truth (or plausibility) of the latter undermines the truth of the former.
   5. Moreover, saw that the conceptual reading is not obviously true.

C. Positive reading
   1. There is a question, which we will get to later, about whether the positive reading of "formality," ontologically, is compatible with a *negative* reading, conceptually: that is, whether a non-semantical characterisation can be given of the "positively formal" (i.e., effective or potent) properties of the symbols or processes in terms of which a computer works.
   2. *This is what Haugeland does*, in giving his account of computers (as being *automatic, formal, digital systems*).
   3. To my knowledge, he is the only person who has a clean and clear commitment to the conceptual reading of "formality" (in its negative sense)
   4. It's possible this works, as a positive characterisation of potency (though I have my doubts)
   5. However, if to be a computer is to be an automatic, formal, digital system *with an interpretation*, as he seems to think then he hasn't naturalised computing: he has only naturalised the

*positive* (potent, effective, "body") half of the dialectic.

6. Thing is, *this* was not the part that was naturalistically challenging!
7. Cf. the voodoo doll. (cf. AOS II·2·16).

D. Status
   1. So semantics remains to be explained
   2. What about model theory? Don't we have a theory of semantics, there?
   3. Not close. Not a theory of semantics (what establishes it, what it is to have a semantic interpretation, what systems can be interpreted, etc.). Just a (very narrow, once one comes to think about it) theory of how the semantics of complex expressions, in a particular (though never well-defined) set of formal languages, are derived from the semantics of individual (atomic) identifiers (constants and predicate & relation letters).

E. Summary
   1. When all is said and done: impossible to credit the idea that the notion of formal symbol manipulation can discharge the intentional mysteries, and mesh with other (allegedly more acceptable) forms of understanding of the world.
   2. Semantic ("mind") issues, brought onto the table by this construal (so I claim), are not yet answered. All the constitutive questions remain open.

F. Plan
   1. Today, turn to the second part of the critique: an *empirical* critique, of whether the (ontological) reading of (negative) formality is actually *true*.
   2. Answer: no. The ontological reading fails, because it is too *narrow*
   3. Fundamental issue: **Involvement / participation**

> 1) The ontological sub-reading of formality claims that, wrt. effective operations, computers are *separate* from their semantic realms;
> 2) In point of fact, computers (in the wild) are *involved* in their subject matters
> 3) It is this involvement, ultimately, that will defeat the FSM view

G. Note: from here on out the notes are merely telegraphic, because we will develop the analysis collaboratively, in class. For a fuller analysis cf. AOS chapter II·3.

## II. Internal counterexamples (small)

A. Host of familiar small examples
   1. Quotation
   2. Meta-level variables
   3. (CAR (CDR '(A B C))
   4. DERIVABLE-FROM(S*, {$S_1$, $S_2$, … $S_k$})
   5. FIRST("tumblehome") ⇒ 'T'
   6. =('a', 'b') ⇒ FALSE

B. In each case:
   1. States of affairs in what we take to be the *realm of interpretation*

2. Play a causal (effective) role in *determining the outcome*

C. Are these counter-examples to the formalist's claim?
   1. That depends on what the formalist's claim comes to

D. Two possible counter-arguments
   1. Property version
      a. These things aren't having the effect they do *in virtue* of being the semantical referents
      b. I.e., should operate at the level of properties
      c. Look: change the semantics—e.g., take the system to represent the decay of family values—and the machine works just the same!
      d. As a response, this won't do. The assumption that you can change the semantics without affecting the operation of the machine is an *assumption* that the system is formal
         i. Not quite: assumption that the *syntax (workings) are independent of the semantics*
         ii. Rather: assumption that the *semantics are independent of the syntax (workings)*
         iii. But that is sort of telling.
         iv. Is "independent" meant to be *symmetrical* or *asymmetrical*?
         v. We will have to come back to this
      e. For now: be careful that one isn't assuming the conclusion
      f. Moreover, lots to be learned by staying here.
   2. Get rid of internal reference
      a. Revised thesis would be something like: *independent of the semantics of the ingredient symbols, except for those symbols designating other syntactic entities, or other internal entities within the causal reach of the on-going process.*
      b. Still have other counter-examples to come, so this will get still more revision
      c. But can tell already what is wrong with it: it is going to fall of the "too broad" cliff.

## III. Internal examples (Large)

A. Examples
   1. Internal reference (reference to *entities within reach of—within the scope of—the machine*) isn't exactly outré.
   2. Lots of obvious cases:
      a. Pointers and array refs
      b. File names
      c. URLs
      d. Routing tables
      e. Consistency predicates on data bases
      f. Transaction logs
      g. GOTO instructions
      h. Reference counts
      i. Compilers
      j. Linkers and loaders
      k. File names

l. Desktop publishing systems
m. E-mail
n. Servers
o. Timers and clocks
p. … other

B. What do these indicate?
1. On reflection, it is apparent that a large percentage of computational "symbol systems" are *about a realm that is within causal reach of the machine itself*
2. That is, once we take internal subject matters into account, seem to be *loads* of them.
3. Extraordinarily important: real-world systems have huge amount of *participatory involvement*
4. Perhaps even the majority of the cases.
5. This is a moral that will stay with us until the bitter (sweet?) end
6. What is it to say that a computer is a system that deals with all of those sorts of symbol structure *independently of their semantics*, when the semantics is *constituted* in (and perhaps even by) the machine?
7. Will come back to when we try to distill out the moral of this investigation

C. But continue on our survey

## IV. Mathematical examples

A. Simple examples
1. LENGTH("ABC") ⇒ 3
2. FOR I FROM 1 TO 1BEGIN PRINT("IS ANYONE HOME?") END
3. Error correction circuits

B. What is going on?
1. Situation is similar to the cases we saw above (of internal reference)
2. All sorts of mathematical properties are *internally exemplified*
3. May or may not want to say that they are doing anything effective
4. Depends on your philosophy of mathematics.
5. But extraordinarily important nevertheless

C. Status of formality?

D. Two possibilities:
1. Mathematical properties can *never* be effective
a. Okay ⇒ in these cases, would imply that symbols don't work in virtue of the exemplification of semantic properties
b. Note, though, that *the same must be true of any material entity* (including people)
c. Not a *problem* (but should be admitted)
d. Implies that *with respect to mathematical properties*, people must be formal, too
2. Mathematical properties *can* sometimes be effective
a. Seems that these cases we have rehearsed are as a good a case as any
b. Then, either
i. Computers aren't formal after all (case closed); or

ii. Adjust reading of formality, again, to something like the following
iii. Independent of the semantics of all ingredient symbols except for
α. Those (meta-level) symbols that designate other internal structures; and
β. Those symbols that designate mathematical properties of internal structures
iv. This is starting to get baroque

E. Morals
1. Same conclusion wrt mathematics as in the internal case: computers are causally involved in subject matters that exemplify properties that they reason about
2. Question: what does all this imply
i. For the philosophy of mathematics?
ii. E.g., wrt establishing the intended interpretation
b. For cognitive science?
3. Cf. brain in a vat
a. Classic image: a brain that has *no causal connection with its subject matter*
b. Condemned to a life of pure, disconnected reason
c. What is wrong with the brain?
i. … or with the vat?
ii. … or with the thoughts it has, as it whiles away its time?
iii. … or with the time that passes—the time that it is whiling away?
4. Note how classic image is starting to erode
a. Two possibilities:
i. Need transducers (or equivalent) to connect to the *rest* of the world
ii. Need transducers (or equivalent) to connect to the *world*
b. Classic image: (i)
i. Not as if the brain is in a *vat*
ii. It is more as if the brain is *nowhere*
c. Sure enough, that would be trouble (semantical and otherwise)
d. But at best, in reality, what is defensible is (a.ii).

## V. External

A. Prefatory remarks
1. Talked a few times ago about the paradigmatic image for FSM: mathematical theorem prover, or NASA system to calculate planetary orbits
2. Emphasized the separation (disconnection) between syntactic and semantic domains.
3. Another assumption: that they are *causally* isolated
4. General image is of an internal CPU or process, whizzing about shuffling symbols, with only occasional interruptions or traffic across the I/O boundary.
5. That boundary, at least in cognitive science, is viewed as the province of *transducers*
6. Wrt the formality (or "computationality") of transducers, cf. Fodor's remark: "Please don't ask me about transducers. I am particular busy just now."
7. One thing we are going to want to ask is whether that image is correct.

B. Robocrab
   1. Describe "coupled" architecture (cf. AOS·II·3)
   2. That is: continuous causal coupling, from crab through e-m radiation through sensors through control circuits through effectors through arms through crab.
   3. Questions to be asked:
      a. Is Robocrab a formal symbol manipulator?
      b. Is Robocrab a computer?
   4. Only if the answer to the first is *no* and second is *yes* do we have evidence against FSM.

C. Three possible counters:
   1. Transducer
      a. Admit that antisemanticity has been violated, but
      b. Deny that the robot is in fact a computer
      c. Probably by claiming that it works entirely in virtue of a complex string of transducers.
   2. Resolute FSM
      a. Maintain robot's claim to being computational
      b. Deny that antisemantics has been violated (superficial evidence notwithstanding)
   3. Combination
      a. Drawing a line through the machine, and claim:
      b. One side ("input/output") violates antisemantics, but *not* computational
      c. Other side ("symbol manipulation"): *is* computational, but *antisemantical after all.*

D. Don't think any work.
   1. Re transduction:
      a. What about word-processors?
      b. Data-driven machines in general?
   2. 2. Boundaries (important in second and third reply)
      a. Not moats! All sorts of causation crosses them
      b. Imagine a thread (e.g., a READ-EVAL-PRINT loop), that does some internal symbol-shuffling, projects something out across the boundary, and suspends, waiting for input.
      c. Process suspends, waiting for input to come back
      d. Then, once it *does* return, process or thread starts up again
      e. Same structure internally
      f. What warrant do we have for breaking up the process boundaries that way?
      g. *Flow of causation* goes right across the boundary (sometimes even into the world).
      h. Why don't we individuate processes so that the *process* travels out into the world?
      i. (Answer left as an exercise for the reader)
   3. Boundaries (important in second and third reply)
      a. Not moats! All sorts of causation crosses them
      b. [Draw picture, on board, of thread waiting for o/i loop to complete (read-eval-print)]

E. Note in passing
   1. Have been talking today about causal (effective) boundaries, causal (effective) flow, causal (effective) traffic

2. This causal (effective) coupling matters to the semantic interpretation of the system
3. Can perhaps start to see why the earlier proposal, of simply *re-interpreting* the system, is not (especially in general) such a sustainable option

## VI. Diagnosis

A. So far, everything has been anecdotal, and bottom up
B. Need to back off, and lay out the conceptual structure of what has been going on
C. Will do that next Tuesday.

—— *end of file* ——