The Blyth Institute

A Dualist Approach to Modeling Cognition

by Jonathan Bartlett
The Blyth Institute

Cognition: Physicalist and Dualist Views

- **Physicalism** - there is nothing going on in the mind that isn't describable by physics
- **Dualism** - the mind and the body are not the same thing, and the mind's mode of operation is not physically describable
- **Emergentism** - too muddy - usually winds up being either primarily physicalist or dualist with different wording

Nailing Down a Definition of "Physical"

- Many *unhelpful* definitions of "physical":
  - Physical = Material
  - Physical = Observable
  - Physical = Testable
- Usable definition of "physical"
  - Physical = Computationally Tractable

Benefits of Using Tractability for Physicality

- Gives workable boundaries for "physical"
- Since computation is equivalent on all finitary logics, computability seems to be a workable equivalent for finitary matter
- Endorsed by many physicalists who denote boundaries to the physical
  - **Principle of Computational Equivalence** (Stephen Wolfram)
  - **Tractable Cognition Thesis** (Iris van Rooij)
- Makes "physicality" a falsifiable hypothesis

Falsifying Physicality: The Halting Problem

- Halting Problem:
  Given a computer program and its input, decide if the program will ever finish. This will be denoted by H(p)
- Why it matters:
  - This is a problem for which a computer program *cannot* be written to solve
  - Therefore, if it is shown to be solvable by humans, physicality is <u>falsified</u>
  - We will provide <u>evidence</u>, not proof

A Quick Primer on Proving Unsolvability

- Given a set of symbols, number every possible function of one variable (this is a countably infinite set)
- Designate F(x) as the function that will run the function at position x in the list above with x also as the argument
- Designate G(x) as being F(x) + 1
- Assume G(x) is in our list of functions
- This leads to a contradiction, because, given its own index, it should return F(x), but it is defined as returning F(x) + 1
- Therefore, neither G(x) nor F(x) are in our list of functions
- Therefore, it is proved that F(x) is always unsolvable using the same set of symbols that are used for the numbering
- Of interest for our study, is that the operators (i.e. symbols) in all finitary logic systems are equivalent

Oracles: A Model for Solving Uncomputable Functions

- Since the H(p) cannot be solved, it is considered an “oracle” - a non-computational function
- Even without being able to implement it, H(p) can be used by itself as a logical relationship
- Turing used H(p) to prove that there were harder problems than H(p)
- Oracles allow functions which cannot be computed to at least be reasoned about

Oracles and Cognitive Models

- Humans have been compared to **oracle machines** before (Turing, Gödel, Copeland)
- My proposal is to use **specific** oracles as parts of **cognitive modeling**
- Specifically, oracles can make good models of **insight problems** in cognitive psychology

Insight vs Analysis Problems in Cognitive Psychology

Comparison

Analysis	Insight
Uses algorithms or heuristics	Deals with conceptualization
Made easier with continuous effort	Made easier with breaks
Subjects can gauge progress	Subjects cannot gauge progress

Examples

Analysis	Insight
Water Jug	Nine-Dot
Algebra	Matchsticks
Towers of Hanoi	Mutilated Chessboard

The Simplistic Argument

- The H(p) **provably** cannot be solved by a computer algorithm in the general case
- The H(p) **practically** is solved continually by computer programmers every day - otherwise they wouldn't be able to tell if their programs would function and finish
- The programs being created are of **arbitrary** complexity - i.e. not decided by the programmers but by outsiders (i.e. employers), thus it does not appear to be a selection bias
- Therefore, it appears that humans can perform the H(p) function

Problems with the Simplistic Argument

- There exist problems in number theory that have not been solved, which could be solved by knowing H(p) for certain programs
- Solving H(p) seems to depend on prior knowledge
- H(p) results can be wrong

Bringsjord's Process

1. Solve H(p) for p of size *n*
2. Use the patterns in step 1 to solve many of the programs of size *n*+1
3. Of the remaining programs of size *n*+1, solve them based on new patterns that can be proven to be non-halters

Note: Proofs of non-halters in Step 3 is usually based on previous proofs of non-halters originating from the previous (size *n*) non-halters

Chaitin: Partial H(p) Solutions Can Be Reached by Adding Axioms

- H(p) is not generally solvable, but can be solved in specific cases
- Having additional axioms can lead to additional partial solutions on H(p)

What We've Learned

- From experience - H(p) is reliably performed by humans
- From Chaitin - more halting problems can be solved if new axioms are introduced
- From Bringsjord - humans can progressively develop new proofs to expand the range of solvability for H(p)

Proposal

$$I(Q, p, B) = A$$

- Q: a decision problem (such as H)
- p: a given program
- A: a set of axioms required to solve Q
- B: a set of axioms such that $|A \cap B| = |A| - 1$
- I: an **oracle** representing human insight needed to generate the axiom to solve the problem

Issues

- Which axioms are we talking about? (top-down vs bottom-up)
- Under what circumstances does the insight oracle operate?
- What is the reliability of this oracle?
- What are the differences between individuals?

Oracles as a General Method for including Nonphysical Components in Scientific Models

- Oracles are any function which can be *described* but cannot be *computed* - matches our definition of non-physical
- Oracles can be tested by congruence but not predictability
- Oracles can be used in conjunction with normal computational elements
- Oracles therefore allow the integration of material and non-material concepts in modeling

Methodological Issues for Oracles

- Testing is dialectic rather than deductive
- Post-hoc materialist explanations can always be proposed
- Testing should be based on:
  - Whether the characteristics of the results can be well-specified
  - Whether the competing material explanations are continually post-hoc or if they define a generalized rule
  - Metrics similar to CSI may be usable to integrate these criteria

Engineering Applications

- Allows separation of problem spaces between human and machine
- Helps humans frame their own problems in a way that facilitates their solving them
- Shows which types of solutions most need to be documented (those that include new axioms)
- Integration with cognitive science allows us to determine which problems require what working conditions