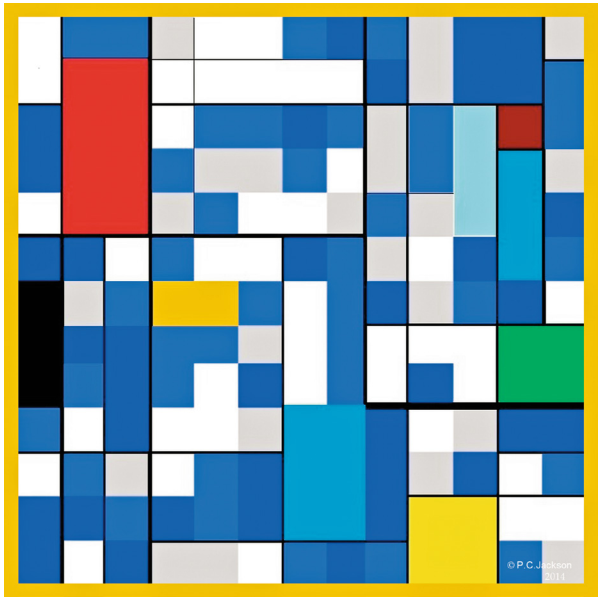
Toward Human-Level Artificial Intelligence

Representation and Computation of Meaning in Natural Language

Philip C. Jackson, Jr.



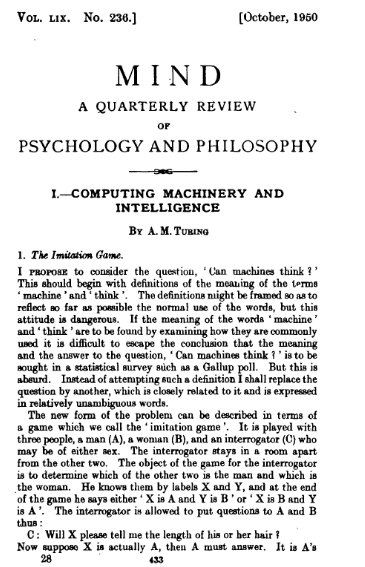
Doctoral Thesis Introduction
April 22, 2014
Tilburg University

Topics

- History and Issues for Human-Level AI
- Thesis Approach
- Tala Conceptual Language
- TalaMind Systems Architecture
- TalaMind Prototype Demonstration
- Thesis Evaluation

Can Machines Have Human-Level Intelligence?

- In 1950, Turing's paper on *Computing Machinery and Intelligence* challenged scientists to achieve human-level artificial intelligence…



Vol. LIX. No. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

1. *The Imitation Game.*

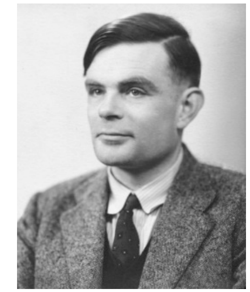
I propose to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:

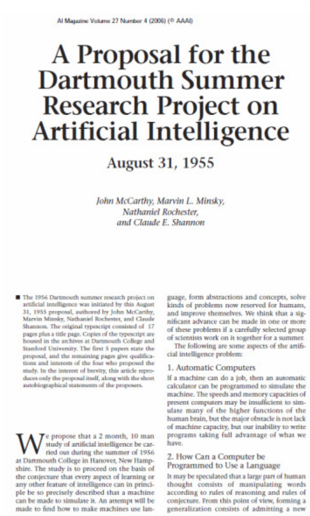
C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's

28 433



- …although the term *artificial intelligence* was not officially coined until 1955, in the Dartmouth summer research project proposal by McCarthy, Minsky, Rochester, and Shannon.



AI Magazine Volume 27 Number 4 (2006) (© AAAI)

A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence

August 31, 1955

John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon

■ The 1956 Dartmouth summer research project on artificial intelligence was initiated by this August 31, 1955 proposal, authored by John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. The original typescript consisted of 17 pages plus a title page. Copies of the typescript are housed in the archives at Dartmouth College and Stanford University. The first 5 papers state the proposal, and the remaining pages give qualifications and interests of the four who proposed the study. In the interest of brevity, this article reproduces only the proposal itself, along with the short autobiographical statements of the proposers.

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

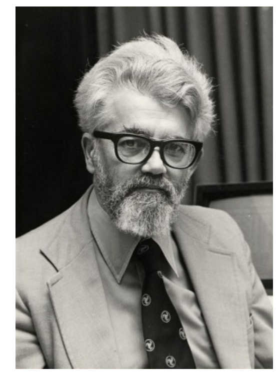
The following are some aspects of the artificial intelligence problem:

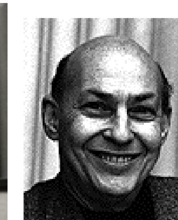
1. Automatic Computers

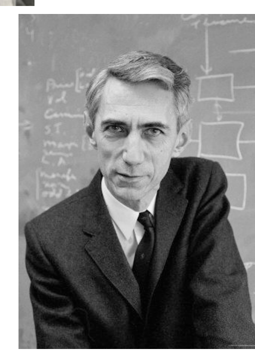
If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

2. How Can a Computer be Programmed to Use a Language

It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture. From this point of view, forming a generalization consists of admitting a new







Turing’s 1950 Paper

- Suggested we could say a computer thinks if it cannot be reliably distinguished from a human being in an “imitation game”, which is now known as a Turing Test.

- Suggested programming a computer to learn like a human child, calling such a system a “child machine”.

- Understanding natural language would be important for educating a child machine and for playing the imitation game.

McCarthy in the 1955 Dartmouth Proposal

- Said his research in the summer project would focus on "*the relation of language to intelligence*":*
  - "The English language has a number of properties which every formal language described so far lacks"
  - "The user of English can refer to himself in it and formulate statements regarding his progress in solving the problem he is working on".

- McCarthy proposed to construct an artificial language with properties similar to English, for "problems requiring conjecture and self-reference":
  - "It should correspond to English in the sense that short English statements … should have short correspondents in the language…"
  - "[and such that] using this language it will be possible to program a machine to learn to play games well and do other tasks."

* - Italics added.

Turing & McCarthy's Expectations

Turing, McCarthy, and several other early researchers expected human-level AI would be achieved before 2000.

- Turing predicted that by the year 2000, people would "be able to speak of machines thinking without expecting to be contradicted."

- In 2006, McCarthy gave a lecture titled *Human-level AI is harder than it seemed in 1955*. He said "If my 1955 hopes had been realized, human-level AI would have been achieved before many (most?) of you were born"

Turing & McCarthy's Expectations Were Not Met

- Nowadays, while people do informally speak of machines thinking, it is widely understood that computers do not yet really think or learn with the generality and flexibility of humans.

- It has been clear for many years the challenges to achieving human-level artificial intelligence are very great.

- It has become clear the challenge is somewhat commensurate with achieving fully general machine understanding of natural language.

- Indeed, while many scientists continue to believe human-level AI will be achieved, some scientists and philosophers have for many years argued that human-level AI is impossible in principle, or for practical reasons.

Central Question of the Thesis

- In sum, the question remains unanswered:

How could a system be designed to achieve human-level artificial intelligence?

- The purpose of the thesis is to help answer this question, by describing a novel research approach to design of systems for human-level AI.

- The thesis presents hypotheses to address this question, and presents arguments and evidence (a demonstration system) to support the hypotheses.

- There is much future research to be done in the approach I propose. So, I cannot prove this approach will achieve AI. I can only argue that it is likely to be successful, and that it can in principle be successful.

Thesis Approach

- Since the challenges are great, and progress has been much slower than early researchers expected, my thesis reconsiders Turing and McCarthy's original suggestions:

  – The Turing Test – how to define and recognize human-level AI?

  – Turing's suggestion of the "child machine" idea (now also called "baby machines").

  – McCarthy's proposal for a formal language having properties similar to English.

Reconsidering the Turing Test

While a Turing Test can facilitate recognizing human-level AI if it is created, it does not serve as a good definition of the goal we are trying to achieve:

1. As a behaviorist test it does not ensure the system being tested actually performs internal processing we would call intelligent.

   In 2011, "Cleverbot" was judged to be human 59% of the time in Turing Tests at the Indian Institute of Technology, by retrieving responses previously typed by humans in other dialogs.

2. The Turing Test is subjective: A behavior one observer calls intelligent may not be called intelligent by another observer, or even by the same observer at a different time.

3. It conflates human-level intelligence with human-identical intelligence. In creating human-level AI, we don't necessarily want to create systems that mimic logical errors in human thinking.

Thesis Alternative to the Turing Test

Inspection of a system's design and operation to verify support of *higher-level mentalities*, such as:

- Natural language understanding.
- Higher-level forms of learning and reasoning, e.g.:
  - Learning about new domains by developing analogies and metaphors with previous domains.
  - Learning by reflection and self-programming
  - Learning by invention of languages and representations.
- Imagination.
- Consciousness.

Other topics include Sociality, Emotions, Values, etc.

Reconsidering Child/Baby Machines

- To date this idea has not been successful, since a baby machine needs to have very flexible ways of representing new knowledge. (Minsky, 2006)

- It is not the case that people have been trying and failing to build baby machines for the past sixty years.
  - Rather, most AI research has been on lower-level areas such as problem-solving, theorem-proving, game-playing, machine learning, etc.
  - It has been clear any attempts to build baby machines with the lower-level techniques would fail, because of representational problems.

- What we may draw from this is that the baby machine approach has not yet been adequately explored. More attention needs to be given to the architecture and design of a child or baby machine, and to the representation of thought and knowledge. This motivates Hypothesis I.

Thesis Approach for Baby Machines

Hypothesis I:

> Intelligent systems can be designed as *intelligence kernels*, i.e. systems of concepts that can create and modify concepts to behave intelligently within an environment.

- I call concepts that describe how to perform processes "executable concepts" or "xconcepts". The prototype illustrates that an xconcept can create and modify an xconcept to solve a problem.
- The ability of an intelligence kernel to be a baby machine depends on the representational flexibility of its conceptual language(s). Chapters 3 and 4 discuss how the approach of the thesis addresses the representational issues Minsky identified for baby machines.
- I first stated a version of this hypothesis in my 1979 Master's thesis. It is a variant of the Physical Symbol System Hypothesis (Newell & Simon, 1976).

McCarthy's Work Relative to Natural Language

Although McCarthy in 1955 proposed to develop a formal language with properties similar to English:

- Beginning in 1958, his papers concentrated on use of predicate calculus for representation and inference in AI systems.
- In 2005, he wrote "English is important for its semantics – not its syntax."
- In 2008, he wrote that grammar is secondary, that the language of thought for an AI system should be based on logic, and gave objections to using natural language as a language of thought.
- McCarthy was far from alone: Almost all AI research on natural language understanding has attempted to translate natural language into a formal language such as predicate calculus, frame-based languages, conceptual graphs, etc. Some approaches have constrained and 'controlled' natural language, so that it may more easily be translated into formal languages.

Thesis Approach to Natural Language

My thesis investigates whether an AI system can perform cognitive processing directly with unconstrained natural language.

Hypothesis II:

> The concepts of an intelligence kernel may be expressed in an open, extensible conceptual language, providing a representation of natural language semantics based very largely on the syntax of a particular natural language such as English, which serves as a language of thought for the system.

- The thesis discusses theoretical issues for this approach, including McCarthy's objections. I argue there is no reason in principle why natural language syntax cannot be used to support inference in an AI system.

Thesis Terminology

- The name *Tala* for the conceptual language is taken from the Indian musical framework for cyclic rhythms.
  - The musical term *tala* is also spelled *taal* and *taala*, and coincidentally *taal* is Dutch for “language”.
- *TalaMind* refers to:
  - The theoretical approach of the thesis and its hypotheses.
  - An architecture the thesis discusses for design of systems according to the hypotheses.
  - The prototype system illustrating this approach.
- *Tala Agent* refers to a system having the TalaMind architecture.

An example of a Tala conceptual expression

```
(ask (wusage verb)
     (subj Leo)
     (indirect-obj Ben)
     (obj
        (turn
           (wusage verb)
           (modal can)
           (sentence-class question)
           (subj you)
           (obj (grain (wusage noun)))
           (into
               (fare (wusage noun)
                   (for
                      (people (wusage noun))
                      )))))
     (tense present)
     (subj-person third-singular)]
```

Leo asks Ben 'Can you turn grain into fare for people?'

Tala can represent unconstrained, complex English sentences, involving self-reference, conjecture, and higher-level concepts – anything that can be expressed in English, with underspecification and semantic annotation.

Tala sentences can describe how to perform processes and be executable concepts, i.e. Tala is a simple programming language, with variables, pattern-matching, etc.

Tala responds to McCarthy's 1955 proposal for a formal language that corresponds to English. Its sentences are as concise as English, because they are isomorphic to English.

Tala Syntax Design

- The TalaMind hypotheses do not mandate any particular approach to representing the syntax of a natural language. Although I've used a dependency grammar approach in the thesis, future research could explore phrase-structure grammars.

- Chapter 5 presents an initial design for the syntax of Tala that is fairly general and flexible, addressing issues such as compound nouns, gerunds, compound verbs, verb tense, aspect and voice, nested prepositions, clitic possessive determiners, gerundive adjectives, shared dependencies, coordinating and subordinating / structured conjunctions, subject-verb agreement, etc.

- This coverage suggests a Tala syntax could be comprehensive for English.

Thesis Approach to Higher-Level Mentalities

Finally, my thesis considers the relationship of natural language understanding to other higher-level mentalities:

Hypothesis III:

> Methods from cognitive linguistics may be used for multiple levels of mental representation and computation. These include constructions, mental spaces, conceptual blends, and other methods.

- This hypothesis is used in my theoretical discussions, and in the design of the prototype demonstration system.

TalaMind System Architecture

- The TalaMind architecture parallels Gärdenfors' (1994) discussion of three levels of inductive inference, though with several differences.
- The TalaMind architecture is open to different ways of representing concepts at the three conceptual levels.
- At the linguistic level, TalaMind is open to use of other formal languages and notations if they are better than natural language for expressing some concepts.
- So, TalaMind is open to predicate calculus or conceptual graphs, in addition to Tala.

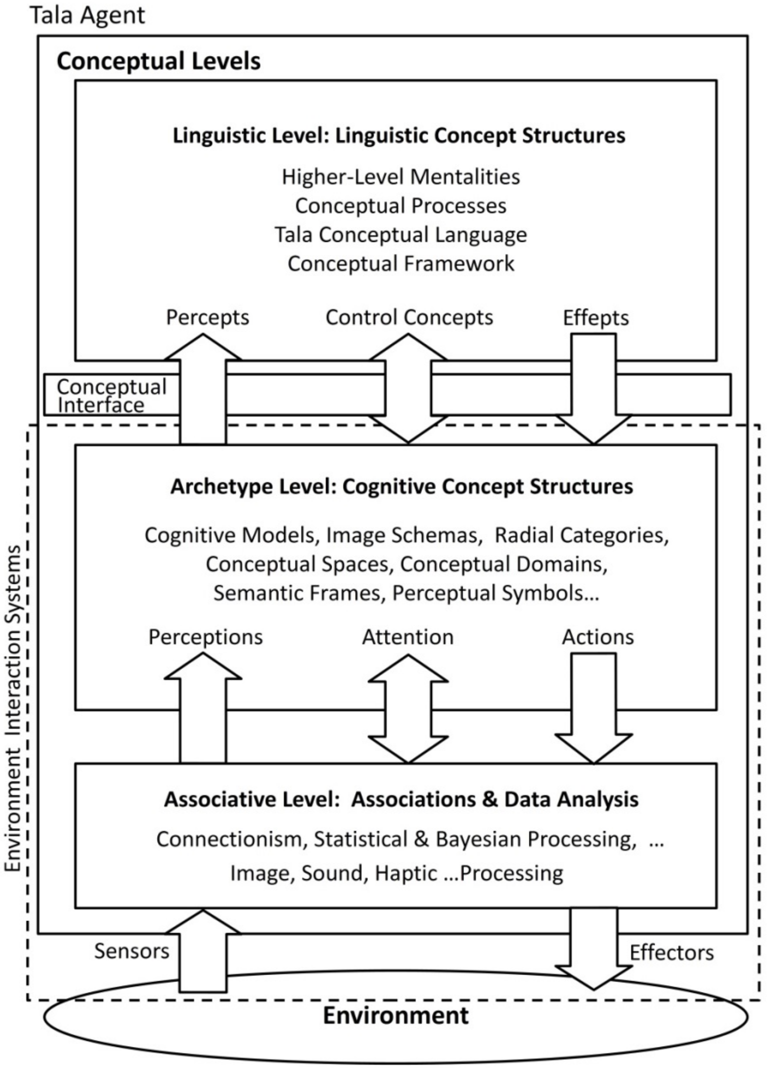


Figure 1-1. The TalaMind System Architecture

TalaMind Archetype & Associative Levels

- Tala allows words to have linguistic definitions. The architecture also allows words to have radial, context-specific meanings represented at the archetype level.
- Gärdenfors (1994) discussed "conceptual spaces" in the archetype level.
- TalaMind is open to alternative ways to represent concepts at this level.
- Ways to unify representations within or across the three levels may be worthwhile.

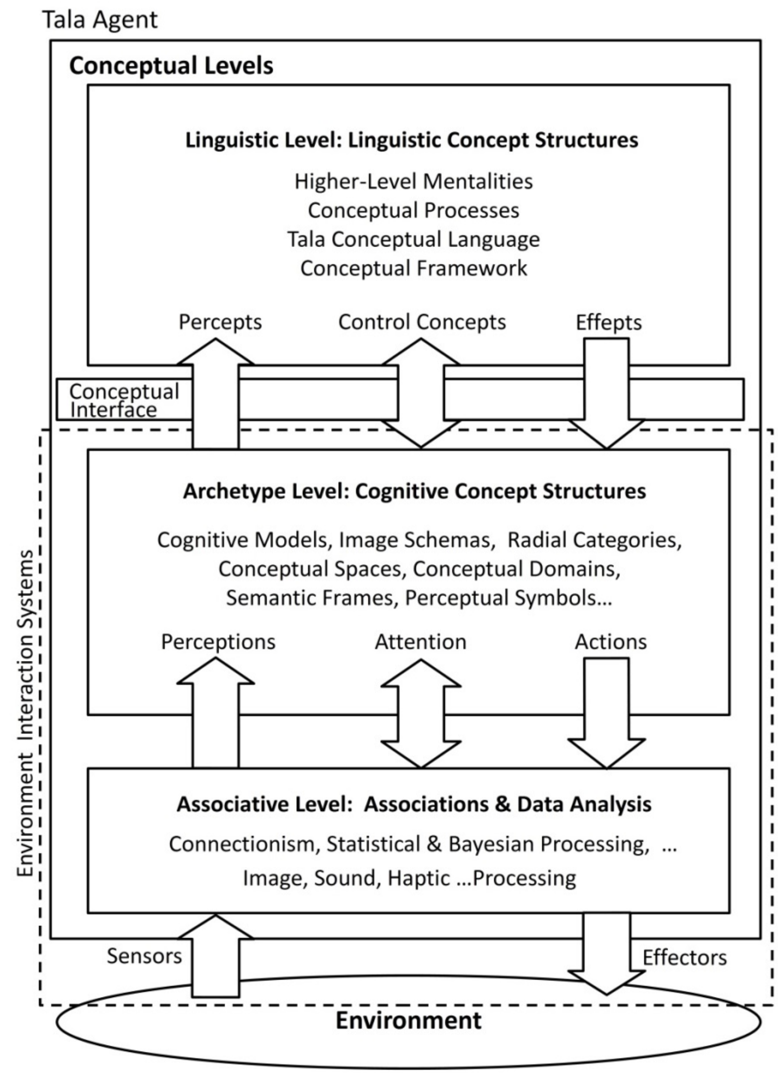


Figure 1-1. The TalaMind System Architecture

TalaMind Embodiment

- The TalaMind architecture includes interfaces with an environment.
- So, to the extent that understanding natural language requires embodiment, TalaMind supports this.
- I give arguments that embodiment does not require an intelligent system to have physical capabilities exactly matching those of human beings.

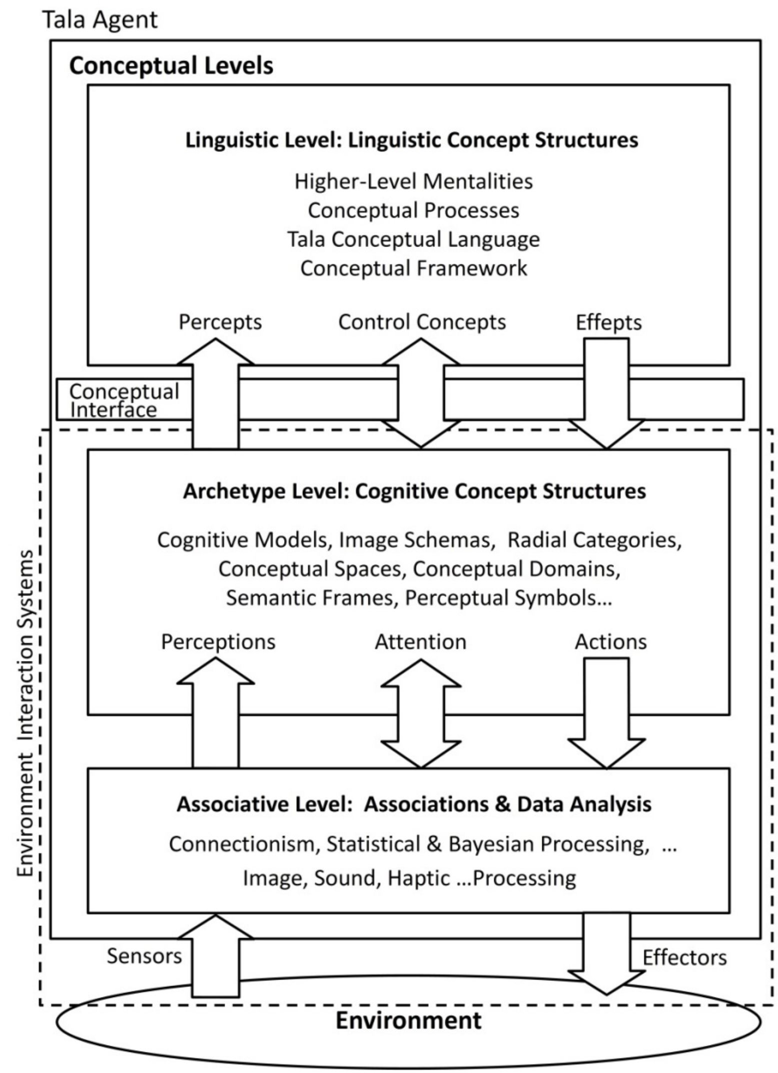


Figure 1-1. The TalaMind System Architecture

Conceptual Framework & Conceptual Processes

In addition to the Tala conceptual language, the architecture contains two other principal elements at the linguistic level:

- *Conceptual Framework*. An information architecture for managing an extensible collection of concepts, expressed in Tala.

- *Conceptual Processes*. An extensible system of processes that operate on concepts in the conceptual framework, to produce intelligent behaviors and new concepts.

The TalaMind hypotheses do not require a *society of mind* architecture in which subagents communicate using the Tala conceptual language, but it is consistent with the hypotheses and natural to implement a society of mind at the linguistic level.

The prototype system has a preliminary design for a conceptual framework and processes, including a society of mind.

Design for TalaMind Prototype System

Chapter 5 presents the design for the prototype demonstration system, which I wrote in Jscheme and Java. Jscheme is a version of Scheme implemented in Java by Anderson, Hickey, and Norvig (2006).

The conceptual framework includes preliminary representations of perceived reality, semantic domains, mental spaces, conceptual blends, and event memory.

The prototype conceptual processes include executable concepts with pattern-matching, variable binding, transmission of mental speech acts between subagents, and composable grammatical constructions.

TalaMind Prototype Demonstration

- Chapter 6 discusses how the prototype illustrates that the TalaMind approach could potentially support the higher-level mentalities of human-level intelligence.

- The demonstration illustrates learning and discovery by reasoning analogically, causal and purposive reasoning, meta-reasoning, imagination via nested conceptual simulation*, and internal dialog between subagents in a society of mind using a language of thought.

  * Conceptual processing of hypothetical scenarios, with possible branching of scenarios based on alternative events, such as choices of simulated Tala agents within scenarios. "Nested" means that simulated agents can also perform this processing within hypothetical scenarios.

- The simulations show conceptual processing without encyclopedic and commonsense knowledge, and without a scalable, efficient architecture. These are needed to achieve human-level AI.

Thesis Evaluation

Chapter 7 discusses:

- The extent to which the theoretical arguments and prototype demonstration support the thesis.
- The novelty of the research approach in relation to previous research.
- Areas for future AI research to develop the proposed approach.
- Future applications of human-level AI and related issues in economics, in particular the issue of “technological unemployment”. (Keynes, 1930)

This concludes my talk.

Thank you, very much, for your consideration of these ideas.