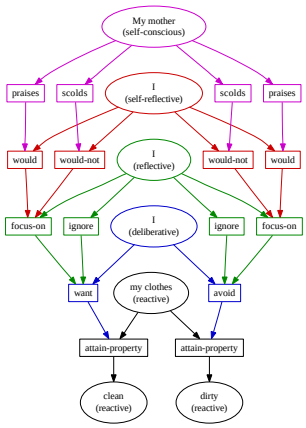
BO MORGAN

A REFLECTIVE COGNITIVE ARCHITECTURE FOR COMBINING A VARIETY OF WAYS OF LEARNING

A REFLECTIVE COGNITIVE ARCHITECTURE FOR COMBINING A VARIETY OF WAYS OF LEARNING

BO MORGAN



Ph.D. in the Media Arts and Sciences

August 2011

Don't do anything that isn't play.

— Joseph Campbell

Dedicated to the loving memory of Push Singh.

1972 – 2006

ABSTRACT

There have been two directions of research with the goal of building a machine that explains intelligent human behavior. The first approach is to build a baby-machine that learns from scratch to accomplish goals through interactions with its environment. The second approach is to give the machine an abundance of knowledge that represents correct behavior.

Each of these solutions has benefits and drawbacks. The baby-machine approach is good for dealing with novel problems, but these problems are necessarily simple because complex problems require a lot of background knowledge. The data abundance approach deals well with complicated problems requiring a lot of background knowledge, but fails to adapt to changing environments, for which the algorithm has not already been trained.

We are working on an algorithm that benefits from both of these approaches by learning from cultural language knowledge, while reflectively monitoring and recognizing the failures of this knowledge when it is used in a goal-oriented domain.

Toward this end we have developed a reflective programming language allowing us the ability to monitor the execution and interactions between large numbers of complicated lisp-like processes. Further, we have developed a cognitive architecture within our language that provides structures for layering reflective processes, resulting in a hierarchy of control algorithms that respond to failures in the layers below.

Finally, we present an example of our cognitive architecture learning in the context of a social commonsense reasoning domain with parents that teach children as they attempt to accomplish cooking tasks in a kitchen.

PUBLICATIONS

Some ideas and figures have appeared previously in the following publications:

Smith, D. and Morgan, B.; "IsisWorld: An open source commonsense simulator for AI researchers"; AAAI 2010 Workshop on Metacognition; 2010 April

Morgan, B.; "A Computational Theory of the Communication of Problem Solving Knowledge between Parents and Children"; PhD Proposal; MIT Media Lab 2010 January

Morgan, B.; "Funk2: A Distributed Processing Language for Reflective Tracing of a Large Critic-Selector Cognitive Architecture"; Proceedings of the Metacognition Workshop at the Third IEEE International Conference on Self-Adaptive and Self-Organizing Systems; San Francisco, California, USA; 2009 September

Morgan, B.; "Funk2: A Frame-based Programming Language with Causally Reflective Capabilities (draft in progress)"; Technical Note; Massachusetts Institute of Technology; 2009 May

Morgan, B.; "Learning Commonsense Human-language Descriptions from Temporal and Spatial Sensor-network Data"; Masters Thesis; Massachusetts Institute of Technology; 2006 August

Morgan, B.; "Learning perception lattices to compare generative explanations of human-language stories"; Published Online; Commonsense Tech Note; MIT Media Lab; 2006 July

Morgan, B. and Singh, P.; "Elaborating Sensor Data using Temporal and Spatial Commonsense Reasoning"; International Workshop on Wearable and Implantable Body Sensor Networks (BSN-2006); 2005 November

Morgan, B.; "Experts think together to solve hard problems"; Published Online; Commonsense Tech Note; MIT Media Lab 2005 August

Morgan, B.; "LifeNet Belief Propagation"; Published Online; Commonsense Tech Note; MIT Media Lab; 2004 January

Though there be no such thing as Chance in the world;
our ignorance of the real cause of any event
has the same influence on the understanding,
and begets a like species of belief or opinion.

— David Hume [2]

ACKNOWLEDGMENTS

Put your acknowledgments here.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

Part I

THE SOCIAL COMMONSENSE LEARNING PROBLEM

1 INTRODUCTION

> Problem-solvers must find relevant data. How does the human mind retrieve what it needs from among so many millions of knowledge items? Different AI systems have attempted to use a variety of different methods for this. Some assign keywords, attributes, or descriptors to each item and then locate data by feature-matching or by using more sophisticated associative data-base methods. Others use graph-matching or analogical case-based adaptation. Yet others try to find relevant information by threading their ways through systematic, usually hierarchical classifications of knowledge—sometimes called "ontologies". But, to me, all such ideas seem deficient because it is not enough to classify items of information simply in terms of the features or structures of those items themselves. This is because we rarely use a representation in an intentional vacuum, but we always have goals—and two objects may seem similar for one purpose but different for another purpose.
>
> — Marvin Minsky [3]

1.1 TWO POPULAR APPROACHES TO MODELLING INTELLIGENCE

Recently, there have been two directions of research with the goal of building a machine that explains intelligent human behavior. The first approach is to build a baby-machine that learns from scratch to accomplish goals through interactions with its environment. The second approach is to give the machine an abundance of knowledge that represents correct behavior.

Each of these solutions has benefits and drawbacks. The baby-machine approach is good for dealing with novel problems, but these problems are necessarily simple because complex problems require a lot of background knowledge. The data abundance approach deals well with complicated problems requiring a lot of background knowledge, but fails to adapt to changing environments, for which the algorithm has not already been trained.

1.2 THE COMMONSENSE REASONING PROBLEM DOMAIN

Commonsense reasoning is a long-standing goal of the field of artificial intelligence. One of the difficulties in developing algorithms for dealing with a commonsense reasoning domain is that the algorithm needs a lot of background knowledge about a given domain before it can answer even simple questions about

it. However, this knowledge is often only true in very specific situations and has many exceptional cases. For example, the knowledge that most birds can fly is generally true, but we also know that many birds are flightless, such as penguins, ostriches, and road runners. Also, we have knowledge about the typical behavior of objects; for example, we know that refridgerators keep things cold, but we also reason efficiently about exceptional cases, such as when the refridgerator is not plugged in, or when the power goes out.

1.2.1 *The Valley of Complex Adaptability*

We would like to build intelligent machines that are able to perform household tasks, such as cooking, cleaning, and doing the laundry, but these tasks seem deep within the "valley of complex adaptability".

1.2.2 *Representations for Commonsense Reasoning*

There have been many approaches to artificial intelligence that use first-order logic as a representation for these types of knowledge and their exceptions, but these systems become cumbersome in their inability to express "fuzzy" sorts of relationships, such as when the knowledge is applicable, for example the modifiers, "most of the time", "usually", and "almost never", are difficult to express in first-order logic. When we have a lot of knowledge, we need ways to keep track of in which situations this knowledge is useful. This is a form of "meta-knowledge", or knowledge about knowledge. Meta-knowledge about first-order logic cannot be expressed in first-order logic, so logic fails us in this regard. Therefore, we need other ways to represent our knowledge in addition to logic.

1.2.3 *The Agent-Environment Model*

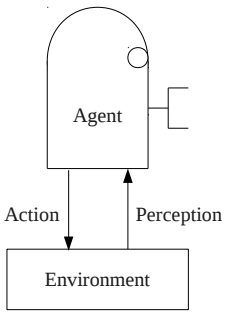


Figure 1: The agent-environment model.

Figure 1 shows the basic agent-environment model. In this model, we make a distinction between the environment and the agent. At any given time, the agent and the environment

are each represented as a specific static form of data. Further, these representations change over time, according to a given transfer function. We will treat this system as a deterministic system, although one could imagine adding random variables to the transfer function, but the basic theory is the same. The two processes communicate information along two channels: (1) an action channel from the agent to the environment, and (2) a perception channel from the environment to the agent.

1.2.4 *The Reinforcement Learning Model*

Figure 2 shows the basic reinforcement learning model. In this model, we make a distinction between a computational process that is the environment and a computational process that is the agent. These two processes communicate information along three channels: (1) an action channel from the agent to the environment, (2) a perception channel from the environment to the agent, and (3) a reward channel from the environment to the agent.

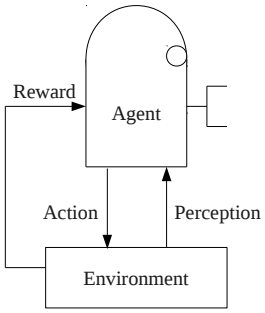


Figure 2: The reinforcement learning model.

The reinforcement learning model is a useful model because it is one of the simplest formulations of goal-oriented learning. Figure 3 shows how the simplest types of categorizations of perceptions can be learned. Note that any categorization is more or less useful to us based on how well it approximates the reward function.

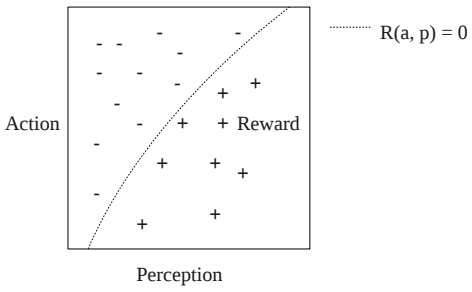


Figure 3: Categorizing perceptions and actions by knowing rewards.

1.2.5 *The Origins of Knowledge*

If we are going to be clear about what we mean by meta-knowledge, we first must be more precise about what we mean by knowledge in the first place.

1.2.6 *Layers of Knowledge about Knowledge*

1.3 USING BACKGROUND KNOWLEDGE IN A GOAL-ORIENTED DOMAIN

We are working on an algorithm that benefits from both of these approaches by learning from cultural language knowledge, while reflectively monitoring and recognizing the failures of this knowledge when it is used in a goal-oriented domain.

1.4 A NEW LISP PROGRAMMING LANGUAGE FOR MODERN COMPUTER ARCHITECTURES

Toward this end we have developed a reflective programming language allowing us the ability to monitor the execution and interactions between large numbers of complicated lisp-like processes.

1.5 A REFLECTIVE COGNITIVE ARCHITECTURE FOR LAYERING FAILURE TOLERANT CONTROL ALGORITHMS

Further, we have developed a cognitive architecture within our language that provides structures for layering reflective processes, resulting in a hierarchy of control algorithms that respond to failures in the layers below.

1.6 A DEMONSTRATION IN A SOCIAL COMMONSENSE REASONING DOMAIN

Finally, we present an example of our cognitive architecture learning in the context of a social commonsense reasoning domain with parents that teach children as they attempt to accomplish cooking tasks in a kitchen.

Part II

REFLECTIVE LAYERS OF GOALS, LEARNING, AND KNOWLEDGE

2

INTRODUCTION

In this chapter we will describe a sequence of scenarios that will demonstrate the top three layers of our theory: (1) reflective, (2) self-reflective, and (3) self-conscious.

First, we will describe examples critics and selectors in the top layers of our model.

2.1 BASIC FORMS OF FAILURE MUST BE DEBUGGED

A plan to use a resource is executed and that resource is no longer available when that step is about to be executed. This could be due to a number of types of reasons:

World Model Failure: The model of the world was incorrect. Miscategorized preconditions and postconditions for an action.

Planning Failure: The plan was incorrect. The agent had the correct knowledge regarding the actions involved in the plan, but the knowledge was not used when the plan was created.

- Control of Planning Failure:

2.2 SELF-CONSCIOUS REFLECTION

Self-conscious reflective critics look for conflicts between self- and other-models in stories and select resources that can debug those types of conflicts.

For example, when a person plans to use a resource and then another person uses that resource.

Part III

A PROGRAMMING LANGUAGE FOR SIMULATING A VARIETY OF REFLECTIVE THOUGHT PROCESSES

Part IV

A COGNITIVE ARCHITECTURE FOR COORDINATING A VARIETY OF WAYS OF LEARNING

3

EXAMPLES

Part V

AN EXAMPLE OF A SOCIAL COMMONSENSE REASONING DOMAIN INVOLVING MULTIPLE WAYS OF LEARNING

4

MATH TEST CHAPTER

Ei choro aeterno antiopam mea, labitur bonorum pri no. His no decore nemore graecis. In eos meis nominavi, liber soluta vim cu. Sea commune suavitate interpretaris eu, vix eu libris efficiantur.

4.1 SOME FORMULAS

Due to the statistical nature of ionisation energy loss, large fluctuations can occur in the amount of energy deposited by a particle traversing an absorber element[1]. Continuous processes such as multiple scattering and energy loss play a relevant role in the longitudinal and lateral development of electromagnetic and hadronic showers, and in the case of sampling calorimeters the measured resolution can be significantly affected by such fluctuations in their active layers. The description of ionisation fluctuations is characterised by the significance parameter κ, which is proportional to the ratio of mean energy loss to the maximum allowed energy transfer in a single collision with an atomic electron:

You might get unexpected results using math in chapter or section heads. Consider the `pdfspacing` *option.*

$$\kappa = \frac{\xi}{E_{max}} \mathbb{ZNR}$$

E_{max} is the maximum transferable energy in a single collision with an atomic electron.

$$E_{max} = \frac{2m_e\beta^2\gamma^2}{1 + 2\gamma m_e/m_x + (m_e/m_x)^2},$$

where $\gamma = E/m_x$, E is energy and m_x the mass of the incident particle, $\beta^2 = 1 - 1/\gamma^2$ and m_e is the electron mass. ξ comes from the Rutherford scattering cross section and is defined as:

$$\xi = \frac{2\pi z^2 e^4 N_{Av} Z \rho \delta x}{m_e \beta^2 c^2 A} = 153.4 \frac{z^2}{\beta^2} \frac{Z}{A} \rho \delta x \quad \text{keV},$$

where

z	charge of the incident particle
N_{Av}	Avogadro's number
Z	atomic number of the material
A	atomic weight of the material
ρ	density
δx	thickness of the material

κ measures the contribution of the collisions with energy transfer close to E_{max}. For a given absorber, κ tends towards large values if δx is large and/or if β is small. Likewise, κ tends towards zero if δx is small and/or if β approaches 1.

1 Examples taken from Walter Schmidt's great gallery: http://home.vrweb.de/~was/mathfonts.html

The value of κ distinguishes two regimes which occur in the description of ionisation fluctuations:

1. A large number of collisions involving the loss of all or most of the incident particle energy during the traversal of an absorber.

   As the total energy transfer is composed of a multitude of small energy losses, we can apply the central limit theorem and describe the fluctuations by a Gaussian distribution. This case is applicable to non-relativistic particles and is described by the inequality $\kappa > 10$ (i. e., when the mean energy loss in the absorber is greater than the maximum energy transfer in a single collision).

2. Particles traversing thin counters and incident electrons under any conditions.

   The relevant inequalities and distributions are $0.01 < \kappa < 10$, Vavilov distribution, and $\kappa < 0.01$, Landau distribution.

4.2 VARIOUS MATHEMATICAL EXAMPLES

If $n > 2$, the identity

$$t[u_1, \ldots, u_n] = t\big[t[u_1, \ldots, u_{n_1}], t[u_2, \ldots, u_n]\big]$$

defines $t[u_1, \ldots, u_n]$ recursively, and it can be shown that the alternative definition

$$t[u_1, \ldots, u_n] = t\big[t[u_1, u_2], \ldots, t[u_{n-1}, u_n]\big]$$

gives the same result.

Part VI

CONCLUSIONS AND FUTURE DIRECTIONS

Part VII

APPENDIX

A

RELATED PHILOSOPHY

A.0.1 *The Objective Modelling Assumption*

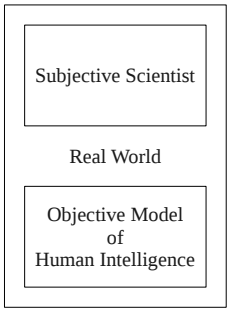


Figure 4: The objective-subjective modelling assumption.

We assume that the phenomenon that we are trying to model, namely human intelligence, is an objective process that we can describe. This is the objective-subjective philosophical assumption that is inherent in any objective scientific hypothesis. We make this assumption in order to avoid logical problems of circular causality that occur when trying to find a non-objective description of reflective thinking. Figure 4 shows how, given the objective assumption, the subjective scientist is part of the real world, while she is studying an objective phenomenon. We refer the reader to Appendix A for a further discussion of the problems of mistaking an objective model for reality itself. We will now continue to describe our model of human intelligence in an objective way, aware of the utility and artificiality of our objective assumption.

A.1 BEING, TIME, AND THE VERB-GERUND RELATIONSHIP

A.2 THE INTENSIONAL STANCE

A.3 REFLECTIVE REPRESENTATIONS

[4]

B

RELATED RESEARCH IN PSYCHOLOGY

Between the ages of 1-3 years old, children display primary emotions, such as joy, disappointment, and surprise. These emotional processes have been hypothesized to be related to the process of failing or succeeding to accomplish a goal. Around age 4, children begin to display emotions that involve the self, such as guilt and shame. It has been hypothesized that these emotions relate to another person's evaluation of the child's goals as good or bad.

We approach modelling this developmental process by applying Marvin Minsky's theory of the child-imprimer relationship. According to Minsky's theory, at a young age, a human child becomes attached to a person that functions as a teacher. The imprimer could be a parent or a caregiver or another person in the child's life, but the function of the imprimer is to provide feedback to the child in terms of what goals are good or bad for the child to pursue.

B.1 SIMULATION THEORY OF MIND VERSUS THEORY THEORY OF MIND

B.2 EMOTION OR AFFECT VERSUS GOAL-ORIENTED COGNITION

B.3 EMBARRASSMENT, GUILT, AND SHAME

C

RELATED RESEARCH IN NEUROSCIENCE

C.1 NEURAL CORRELATES OF CONSCIOUSNESS

C.2 LEARNING BY POSITIVE AND NEGATIVE REINFORCEMENT

D

RELATED RESEARCH IN ARTIFICIAL INTELLIGENCE

D.1 INFERENCE, PLANNING, AND MACHINE LEARNING ARE ALL PARTIAL SOLUTIONS

"These systems use multiple representations including semantic networks, propositional and first-order probabilistic graphical models, case bases of story scripts, rule based systems, logical axioms, shape descriptions, and even English sentences." — Push Singh's webpage

D.2 COMPARABLE COGNITIVE ARCHITECTURES

D.2.1 *EM-ONE*

D.2.2 *Icarus*

D.2.3 *ACT-r*

D.2.4 *Soar*

E

RELATED RESEARCH IN COMPUTER SCIENCE

F

FUTURE APPLICATIONS TO MENTAL HEALTH

G

FUTURE APPLICATIONS TO EDUCATION

BIBLIOGRAPHY

[1] Robert Bringhurst. *The Elements of Typographic Style*. Version 2.5. Hartley & Marks, Publishers, Point Roberts, WA, USA, 2002.

[2] David Hume. *Enquiries concerning the human understanding: and concerning the principles of morals*, volume 921. Clarendon Press, 1902.

[3] Marvin Minsky. Logical vs. analogical or symbolic vs. connectionist or neat vs. scruffy. In *Artificial intelligence at MIT expanding frontiers*, pages 218–243. MIT press, 1991.

[4] Josef Perner. *Understanding the representational mind.* the MIT press, 1991.

COLOPHON

This thesis was typeset with LaTeX 2ε using Hermann Zapf's *Palatino* and *Euler* type faces (Type 1 PostScript fonts *URW Palladio L* and *FPL* were used). The listings are typeset in *Bera Mono*, originally developed by Bitstream, Inc. as "Bitstream Vera". (Type 1 PostScript fonts were made available by Malte Rosenau and Ulrich Dirr.)

The typographic style was inspired by Bringhurst's genius as presented in *The Elements of Typographic Style* [1]. It is available for LaTeX via CTAN as "`classicthesis`".

NOTE: The custom size of the textblock was calculated using the directions given by Mr. Bringhurst (pages 26–29 and 175/176). 10 pt Palatino needs 133.21 pt for the string "abcdefghijklmnopqrstuvwxyz". This yields a good line length between 24–26 pc (288–312 pt). Using a "*double square textblock*" with a 1:2 ratio this results in a textblock of 312:624 pt (which includes the headline in this design). A good alternative would be the "*golden section textblock*" with a ratio of 1:1.62, here 312:505.44 pt. For comparison, `DIV9` of the `typearea` package results in a line length of 389 pt (32.4 pc), which is by far too long. However, this information will only be of interest for hardcore pseudo-typographers like me.

To make your own calculations, use the following commands and look up the corresponding lengths in the book:

```
\settowidth{\abcd}{abcdefghijklmnopqrstuvwxyz}
\the\abcd\ % prints the value of the length
```

Please see the file `classicthesis.sty` for some precalculated values for Palatino and Minion.

145.86469pt

DECLARATION

Put your declaration here.

Cambridge, August 2011

Bo Morgan