Mixing Cognitive Science Concepts with Computer Science Algorithms and Data Structures: An Integrative Approach to Strong AI

Moshe Looks
Dept. of Computer Science
Washington University in St. Louis
mlooks@cs.wustl.edu

Ben Goertzel
Applied Research Center
Virginia Tech, National Capitol Region
goertzel@vt.edu

Abstract

We posit that, given the current state of development of cognitive science, the greatest synergies between this field and artificial intelligence arise when one adopts a high level of abstraction. On the one hand, we suggest, cognitive science embodies some interesting, potentially general principles regarding cognition under limited resources, and AI systems that violate these principles should be treated with skepticism. But on the other hand, attempts to precisely emulate human cognition in silicon are hampered by both their ineffectiveness at exploiting the power of digital computers, and the current paucity of algorithm-level knowledge as to how human cognition takes place. We advocate a focus on *artificial general intelligence* design. This means building systems capturing the salient high-level features of human intelligence (e.g., goal-oriented behavior, sophisticated learning, self-reflection, etc...), yet with software architectures and algorithms specifically designed for effective performance on modern computing hardware. We give several illustrations of this broad principle drawn from our work, including the adaptation of estimation of distribution algorithms in evolutionary programming for complex procedure learning.

Level of Organization and Possible Mappings

In David Marr's seminal decomposition, any information-processing system may be understood at three nearly independent levels: (1) *computational theory*, a description of the problems the system attempts to solve; (2) *representations and algorithms*; and (3) *implementation*, the physical instantiation of the system's representations and algorithms (Marr 1982). Taking this as a springboard, we may characterize approaches to AI vis a vis human cognition by which levels they deem appropriate for mappings between human cognition and AI.

One extreme consists of AI approaches that don't draw mappings between natural and artificial cognition at any of these levels. The most notable examples of this, at present, are purely abstract mathematical theories of intelligence, such as the work of Schmidhuber (2005b; 2005a) and Hutter (2005). However, their approaches rest on the assumption of vast levels of computing power greater than anything that will be physically achievable in the foreseeable future. It is possible that such approaches will be augmented in the future by more plausible heuristics for dealing with limited computational power, but until this occurs they have little relevance to practical AI.

At the other end of the spectrum are claims such as Kurzweil's that the most effective path to strong AI is to "reverse-engineer" the human brain to as precise a level as is necessary to replicate its functioning (Kurzweil 2000). Others such as Jeff Hawkins have taken a slightly more moderate approach, arguing for understanding the brain fully and then creating loosely brain-like AI architectures inspired by what one has learned from neuroscience, without necessarily trying to imitate the details of human brain structure and function (Hawkins & Blakeslee 2004).

Both Kurzweil and Hawkins, in different ways, are suggesting a mapping between humans and AI's at Marr's level of representations and algorithms. Connectionist AI, generally speaking, is a different sort of attempt to create a mapping at this level, while moving a bit further up the hierarchy, and dealing with representations and algorithms that emulate brain structure and function on only a coarse level. For instance, while the backpropagation algorithm is not an accurate model of human brain dynamics, it has been claimed that subsymbolic approaches such as backpropagation neural nets can emulate brain function significantly better than AI algorithms such as genetic programming or backward-chaining inference (Rumelhart, McClelland, & PNP Research Group 1986; Mareschal & Johnson 2002).

Explicitly logic and expert rule based approaches to AI such as Cyc (Lenat & Guha 1990), SOAR (Newell 1990), and ACT-R (Anderson 1993) also attempt to map between humans and AIs at the level of representations and algorithms, but are based on an essentially different notion of what the important representations and algorithms for human cognition are. If one believes that the human brain fundamentally represents knowledge as logical propositions and understands the world and itself via logical inference, then this sort of AI system is the correct approach to take. To oversimplify a bit, we may say that advocates of these logic-based systems place the brain's neuronal network at Marr's implementation level, arguing that neurons are just the brain's particular way of implementing logic; whereas advocates of neural net AI place it at the level of representations and algorithms.

Our own view is distinct from all of these. Like the designers of Cyc, SOAR, and theorists such as Pei Wang (1995) and Marcus Hutter (2005), our interest is in *artificial general intelligence* (AGI) - strong AI systems that confront the world autonomously, learn their own representations for real-world situations, reflect on themselves, and solve a variety of complex problems. Furthermore, we believe that the human mind/brain contains valuable clues for making AGI work under conditions of limited computational resources. However, we believe that, given the current highly incomplete state of neuroscience, the correct way to map between AI and human cognition is at Marr's level of computational theory.

We are neutral as to how directly the brain's neuronal network structure and dynamics relate to its cognitive representations and algorithms, and also as to how closely the brain's knowledge representation resembles formal logic and how closely its dynamics resemble logical inference. These are very important questions, but neuroscience has not yet progressed far enough to give us the answers. No one knows how an abstract proposition like "Every boy has a dog that every girl likes to call by a special name" is represented or manipulated or learned through experience in the human brain, and until we know this, we won't know the extent to which the conceptual premises of the most popular neural net or logic based approaches to AI are correct.

The best approach to strong AI at present, we suggest, is to learn what we can from the brain about what sort of high level architecture and general dynamics and representations are useful for achieving general intelligence under conditions of limited computational resources - and then fill in the algorithmic level with representations and algorithms that make sense in terms of the mathematics, computer science, and computer hardware and software that we know. This attitude leads into an integrative approach to AI, in which one takes a general architecture loosely inspired by human cognition, and then uses it to bind together components drawn from various areas of mathematics and computer science. Importantly however, it also leads to an approach that is different from classic "multi-agent" AI paradigms like the Society of Mind (Minsky 1986), because one of high-level lessons that we draw from contemporary neuroscience is that the brain is not a society but rather a tightly interlinked connection of components that are exquisitely tuned for real-time interaction with much more intense feedback and interplay than exists between the relatively separate individuals that form the parts of a society.

The Value of Current Cognitive Science to AI

We now offer a few more comments on the general value that we find present-day cognitive science offers AI.

Prior to the last few decades the traditional disciplines of psychology and neuroscience offered relatively little guidance to would-be AGI designers. The state of the art has improved massively since then, with the emergence of cognitive science and cognitive neuroscience. It is, however, not yet sufficient to give detailed prescriptions for the construction of AGI systems. Still, we believe, these bodies of knowledge can provide substantial inspiration for AGI design.

Cognitive science provides very clear advice on Marr's computational level, regarding what the overall conceptual architecture an AGI system should be like (in terms of the tasks solved by various subcomponents), if it is going to cognize in a manner even vaguely resembling that of human beings. We know what the major regions of the brain do, and also have a basic decomposition of human cognition into a list of interacting yet significantly distinct faculties. Great strides have also been taken on understanding the implementation level of individual neurons, extending to what might be termed the "lower representational level" (cf. Eliasmith & Anderson 2003).

On the other hand, while cognitive science also provides a variety of suggestions regarding concrete mechanisms for carrying out tasks such as perception, learning, and memory, the take-home message here is more clouded, for two reasons. First, there is often no general consensus on the correctness of these mechanisms. Second, it's not always clear that emulating human psychological or neural behavior is a practical approach to implementing intelligence on radically unbrainlike contemporary hardware.

Fundamentally, modern cognitive science has recognized that there is more to human cognition than its high-level architecture and low-level mechanisms. However, the cognitive sciences to date have had relatively little to say about the crucial *intermediate level* of intelligence, corresponding to the "algorithmic and upper representational level" in Marr's scheme. As Marvin Minsky recently proclaimed, "neuroscience has no theory for the middle level" (Hedberg 2005).

This is the main reason that there is, as yet, no extremely persuasive prescriptive guidance to AGI designers from cognitive science. We know which primary components a mind should have, and some low-level mechanisms that can facilitate them. At the same time little is known on how the different parts all work together, and how the low-level mechanisms coordinate to give rise to higher-level dynamics.

As an example of the extreme paucity of intermediate-level explanations, consider the rather critical notion of the *self*. Thomas Metzinger (Metzinger 2004) has recently given a masterful treatment of the philosophy and neuropsychology of self. But as Metzinger points out, while contemporary cognitive neuroscience tells us a lot about various dysfunctions of self-construction and self-awareness, it is virtually silent on how selves are constructed by brains in the first place. We know the self-building process involves the coordinated activity of a number of different brain regions, and we can list these; and we know some basic neural mechanisms that assist with the process, and can be diverted from their normal behavior via disturbances in the levels of various chemicals in the brain. But what does this "coordinated activity" consist of? How are chemical and neuron-level processes orchestrated across various brain regions to allow the human brain to model the mind that emerges from it in an approximate yet contextually-very-useful way? On issues like this, the cognitive sciences are basically silent.

As a much simpler example illustrating the same point, consider the issue of the relevance of temporal pattern-recognition to visual object-recognition. Hawkins has suggested that object recognition is based on hierarchical recognition of time series of visual sensations. The more traditional view argues that, while temporal analysis is relevant, it is not critical to the ordinary object recognition process (Hawkins & Blakeslee 2004). Although he presents a list of falsifiable predictions associated with his position, there is still no compelling evidence in favor of either side; both approaches are compatible with the known architecture of the visual cortex, as well as the basic neurocognitive dynamics of Hebbian learning and neuronal activation spreading. This again illustrates the type of explanation that the cognitive sciences currently provide only infrequently.

Given the current state of the cognitive sciences, there are three basic rational positions one can take vis a vis the creation of artificial general intelligence. The first is to simply wait until the cognitive sciences advance further, and can provide more constructive design constraints and guidance. The opposite position is to ignore the cognitive sciences and attempt to design an AGI on other grounds e.g. based on the mathematics of reasoning, or general considerations regarding the dynamics of complex self-organizing systems. Of course it is worth reflecting that many of these "other grounds" were originally conceived as qualitative models of human thought. The middle path is to seek the creation of an AGI design consistent with the computational theories provided by cognitive science, introducing additional ideas to fill in the gaps. Of course, this approach has a great deal of flexibility. Where design is concerned, it must be characterized more as an art than a science; however, practical experimentation with designs created in this way is just as scientific as any other kind of AI work.

Design Considerations for AGI

As one illustration of how AGI design may proceed, motivated by the above points, we now review some of the general considerations entering into the design of an integrative AI system on which we have collaborated during recent years, the Novamente AI Architecture; see (Looks, Goertzel, & Pennachin 2004; Goertzel & Pennachin 2005) for more details. What we will focus on here is a couple of ways in which we have defined the overall architecture of the system based on high-level analogies with human cognition, then filled in the details with algorithms that are not at all brain-like but are rather motivated by the exigencies of modern computer hardware and the knowledge base of modern computer science.

As described above, our general approach is to attempt to preserve the spirit of human information-processing as understood by modern cognitive science, while adapting known particularities to allow for computational efficiency on modern hardware. Briefly, the brain contains a massive number of fairly slow and noisy processing units, and has an architecture in which memory and processing are largely overlapping concepts; on the other hand, modern computer networks have a small number of very fast and accurate processors, and an architecture in which memory and processing are distinct (see Eliasmith & Anderson 2003, for example, for more details).

These differences mean that the evolved mechanisms of human cognition, which are effective when realized physically in neural wetware, are not very well-suited to efficient implementation *in silico*. See (Moravec 1998), for example, for an estimate of the computational gap between human brains and modern computers. Thus, from a near-term AGI point of view, the most sensible design strategy seems to be to try to understand the essence of human cognitive functionality as well as possible, then finding roughly equivalent algorithms and representations friendly to modern computing hardware – without worrying about exactly how these functions are implemented in the brain via neurons, neurotransmitters, extracellular charge diffusion and the like. A detailed example illustrating this sort of strategy in action is presented later.

Organizational Principles

We consider it valuable to think about the overall organizational principles of the brain and then map these into structures within AI systems. For instance, hierarchical structure pervades the brain– the cortex is hierarchical, visual perception has been shown to use this physical hierarchy for (conceptually) hierarchical pattern recognition, and a number of theorists have proposed that this principle applies more generally (Hawkins & Blakeslee 2004). Correspondingly, the prevalence of associative structures in the brain has been well-known since the time of Peirce and William James, and has been validated via numerous experiments on the structure of memory (Baddeley 1999). This leads to a "dual network" (Goertzel 1993) structure as a model for the brain's statistics of concepts, perception, and action; an hypothesis that is not convincingly proven but is highly plausible. The way in which self-structures may emerge from dual networks based on experiential learning has been discussed in depth in (Goertzel 1997), with many connections drawn to the literature on personality psychology, e.g. Rowan's (1990) theory of subpersonalities.

Another key aspect of human cognition that we believe is aptly generalizable to AI systems is the central role of pattern recognition and creation in the human mind. The "patternist" perspective on intelligence has deep roots in philosophy; Peirce (1892) and James (1890) established this perspective long ago (though using the language of "habits"rather than "patterns"), and it has more recently reared its head in the form of algorithmic-information-theoretic models of the mind as "compact programs for computing data" (Solomonoff 1964; Baum 2004). The mathematical theory of pattern given in (Goertzel 1997) shows how algorithmic information theoretic models are substantially the same as models based on concepts such as pattern and habit. Neurobiological theorists such as Gregory Bateson (1980) and Daniel Dennett (1978) have connected this patternist view with contemporary biology and psychology.

Combining these two general principles (mind-as-habit-system and dual-network structure), shown to be harmonious with contemporary cognitive science, one obtains a general model of cognition, the Psynet Model of Mind, articulated in some depth in our prior publications (Goertzel 1993; 1994; 1997).

A Cognitive Science Inspired Checklist for AGI

By integrating various principles from and hypotheses regarding human cognition, one can formulate a kind of "checklist" for AGI, consisting of principles that seem very likely to be fulfilled by human cognition and that can plausibly be fulfilled by an appropriately constructed AI system running efficiently on contemporary or near-future hardware. First of all, an AI system must be a dynamical system, consisting of entities (processes) which are able to act on each other (transform each other) in a variety of ways, and some of which are able to identify patterns. Furthermore, it must:

- be sufficiently flexible to enable the crystallization of a dual network structure, with emergent, synergetic hierarchical and heterarchical subnets,
- contain a mechanism for the spreading of attention in directions of shared relationship,
- have access to a rich stream of perceptual data, so as to be able to build up a decent-sized pool of grounded patterns, leading ultimately to the recognition of the self,
- be able to carry out actions in some external domain (e.g. a physical or simulation world), in a way supporting rich feedback between perception, cognition and action,
- contain entities that can reason based on uncertain information (so that, among other things, it can transfer information from grounded to ungrounded patterns),
- contain entities that can manipulate categories (hierarchical subnets) and transformations involving categories in a sophisticated way, so as to enable syntax and semantics,
- recognize symmetric, asymmetric and emergent meaning sharing, and build meanings using temporal and spatial relatedness, as well as relatedness of internal structure, and relatedness in the context of the system as a whole,
- have a specific mechanism for paying extra attention to recently perceived data ("short-term memory"),
- be embedded in a community of similar dynamical systems, so as to be able to properly understand itself.

This is by no means a full list; each element of this list conceals a complex story, and we don't have space to tell these stories here. The point, for the present, is merely to illustrate the sorts of principles that we believe are meaningfully derivable from cognitive science and applicable to AI. These principles are far from vacuous, but they are also far from prescriptive at the algorithm and data structure level. They give guidance as to system architecture and as to what the requirements are for the various algorithms and data structures involved in an AI system. Cognitive science also gives various indications as to how humans instantiate these general principles; but our suggestion is that these more detailed indications may not be that applicable to digital AI systems, because they in large part represent adaptations to the specific physical substrate from which human intelligence emerges.

One interesting observation is that none of the currently popular AI approaches incorporate all of the principles listed above. One of our goals with the Novamente project has been to craft an AGI approach that is consonant with the high-level principles of cognition that we see as being meaningfully abstractable from contemporary cognitive science.

Evolutionary Learning in Brains and Computers

We now turn to a more specific and delimited example of our general point regarding the optimal level at which to interrelate cognitive science and AGI. While this point is relevant at the level of overall AGI system design, we also feel it is relevant more granularly, at the level of the design of specific AI algorithms solving specific problems. We now give an example of how cognitive and computer science inspirations may be blended in the area of evolutionary learning.

Major subfields of both cognitive science and AI are concerned with evolutionary learning processes. We focus here on attempts in both fields to augment purely local techniques such as Hebbian (associative) learning with more global methods, which attempt to make large leaps to find answers far removed from existing knowledge. This is a form of "evolutionary learning" (Holland 1975), which Edelman (1988) has presented as "Neural Darwinism", and others (cf. Calvin & Bickerton 2000) as the notion of mind as a "Darwin Machine".

Deacon articulates the need for such evolutionary methods in the context of designing a system replicating iconic and indexical relationship learning (prerequisites for high-level symbolic communication and reasoning):

> The first requirement ... is that it must be actively and spontaneously adaptive. It must continuously evolve new means of fitting with and anticipating its environment, even if this environment is constrained ... And I mean evolve, not in a genetic or phylogenetic sense, but through a moment-to-moment natural selection of patterned information. It must be capable of generating new patterned information where none previously existed. This is very different from the sort of data architecture comprising a massive list of templates to which an input is compared and matched. (Deacon 1997)

Neurodevelopment and Darwin Machines

It is known that the immune system adapts to via a form of evolutionary learning, and (Edelman 1988) has proposed that the brain does so as well, evolving new "neuronal maps", patterns of neural connection and activity spanning numerous neuronal clusters that are highly "fit" in the sense of contributing usefully to system goals. He and his colleagues have run computer simulations showing that Hebbian-like neuronal dynamics, if properly tuned, can give rise to evolution-like dynamics on the neuronal map level ("neuronal group selection").

More recently, (Deacon 1997) has articulated ways in which, during neurodevelopment, difference computations are in competition with each other (e.g., to determine which brain regions are responsible for motor control). More generally, he posits that there will be a kind of continuous flux as control shifts between competing brain regions, again, based on high-level "cognitive demand" (p. 457). Similarly, (Calvin & Bickerton 2000) has given plausible neural mechanisms ("Darwin Machines") for synthesizing short "programs", for tasks such as throwing a rock or parsing a sentence, which are represented as coherent firing patterns in the cerebral cortex. A population of such patterns, competing for neurocomputational territory, replicates with variations, under selection pressure to conform to background knowledge and constraints.

Computational Methods

There is a long history in AI of applying methods derived from evolution to practical problem-solving; the original genetic algorithm of (Holland 1975), initially a theoretical model, has been adapted successfully to a wide variety of applications (cf. Mitchell 1996). Briefly, the methodology applied, similar to the Darwin Machines described above, is: (1) generate a random population of solutions to a problem; (2) evaluate the solutions in the population using a predefined *fitness function*; (3) select solutions from the population proportionate to fitness, and recombine/mutate them to generate a new population; (4) goto step (2). Koza (1992) has adapted Holland's paradigm from the case of fixed-length bit-strings to evolve variable-sized and shaped trees (typically Lisp S-expressions), which can in principle represent arbitrary computer programs.

Bridging the Gap

Now, consider applying the approach we have proposed herein to the design of a family of algorithms consonant with the computational theory and goals of evolutionary learning in human cognition, adapted to modern-day hardware. To begin with, one need not be constrained by the distributed nature of biological and most neural computation; perhaps a more effective system may be obtained via a centralized data repository maintaining precise statistics on the effectiveness of different solutions?

Indeed, within the field of evolutionary computation, a family of biologically and neurologically implausible algorithms have been developed known as *estimation of distribution algorithms* (cf. Pelikan, Goldberg, & Lobo 1999), which outperform genetic algorithms and related techniques across a range of problems, by maintaining centralized probabilistic models of the population learned with sophisticated datamining techniques. One of the most powerful of these methods is the *Bayesian Optimization Algorithm* (BOA) (Pelikan, Goldberg, & Cantú-Paz 2002). Recently, we have adapted the BOA to learn program trees, creating a method known as BOA Programming (Looks, Goertzel, & Pennachin 2005).

The basic steps of BOA are: (1) generate a random population of solutions to a problem; (2) evaluate the solutions in the population using a predefined *fitness function*; (3) from the promising solutions in the population, learn a generative model; (4) create new solutions using the model, and merge them into the existing population; (4) goto step (2).

The neurological implausibility of this sort of algorithm is readily apparent. One of the major lessons of cognitive neuroscience has been the elegance with which the brain mixes localized and distributed processes (see above). This sort of process is radically different from the BOA, where there is no spontaneous growth, adaptation or combination of the individual population elements, but rather the population elements are studied by an external process and then autocratically modified. A whole BOA population is self-organizing in a certain sense, but this does not come out of the coupling of largely-independently-acting components.

But the question is, why is it desirable for a digital computer based implementation of evolutionary programming to emulate the more fully self-organizing nature of evolution in the human brain? Some aspects of self-organization, we conjecture, are key to intelligence under limited resources and must be retained in any digital AI system that aims at strong AI; see above and (Goertzel 1994) for more details. On the other hand, the decentralization of evolutionary learning does not seem to us to embody any profound aspect of general cognition; rather it seems to us a consequence of the particularities of biological systems. Our conjecture is that, from an AI design point of view, it makes more sense to derive from human cognition the maxim that "evolutionary learning is a valuable part of cognition" but then replace biologically-plausible evolutionary learning methods with others that are more effective given the digital computing substrate.

We have created an extended version of BOA, which is an integral component of the Novamente AI Architecture mentioned above, where it is applied in conjunction with probabilistic inference in a variety of configurations. For example, we are deploying Novamente within a simulated world to model aspects of infant development and cognition; this is related to, but distinct from, development robotics approaches such as (Blank, Kumar, & Meeden 2002). In this context, configurations include:

- Temporal prediction based on past sequence observations.
- Learning a compact predicate to describe a set of observations based on salient features.
- Learning predicates to discriminate between different known object classes (i.e., supervised categorization).
- Learning sets of items which, when given as inputs to an inference configuration, lead to valuable conclusions.

The last these items touches briefly on the importance of integration between the components of a system, another computational-level goal of human cognition which we attempt to realize quite differently. Inference is controlled by evolutionary learning, which in turn is driven by inference.

Conclusions

In our view, the knowledge generated by the cognitive sciences so far is too primitive and too haphazard to be used as a blueprint for AI. However, this doesn't mean it can't be used as an inspiration. By taking ideas from human cognition at Marr's computational level and using these to inspire AGI architectures, one can vastly narrow down the space of possible AGI designs. But to fill in the details one cannot viably draw on knowledge about human brains or minds, because its paucity and uncertainty; so if one wishes to proceed with

AGI design, the only choice is to use the best of contemporary computer science to fill in appropriate representations and algorithms. This leads to an AGI design approach in which the primary challenge is fitting together computer science ideas in a manner that leads to a whole system compatible with the high-level nature of human intelligence. Such an approach is not easy and is not always elegant; but in our view, this is the only practical way to proceed if one wants to seriously approach strong AI without waiting for revolutionary improvements in neuroscience or mathematical theories of general intelligence.

References

Anderson, J. R. 1993. *Rules of the Mind*. Erlbaum.

Baddeley, A. 1999. *Essentials of Human Memory*. Taylor and Francis Group.

Bateson, G. 1980. *Mind and Nature: a Necessary Unity*. Fontana.

Baum, E. B. 2004. *What is Thought?* MIT Press.

Blank, D. S.; Kumar, D.; and Meeden, L. 2002. A developmental approach to intelligence. In *Proceedings of the Thirteenth Annual Midwest Artificial Intelligence and Cognitive Science Society Conference*.

Calvin, W. H., and Bickerton, D. 2000. *Lingua ex Machina*. MIT Press.

Deacon, T. 1997. *The Symbolic Species*. Norton.

Dennett, D. 1978. *Brainstorms: Philosophical Essays on Mind and Psychology*. Bradford Books.

Edelman, G. M. 1988. *Neural Darwinism: The Theory of Neuronal Group Selection*. Basic Books.

Eliasmith, C., and Anderson, C. 2003. *Neural Engineering*. MIT Press.

Goertzel, B., and Pennachin, C. 2005. The Novamente artificial intelligence engine. In Goertzel, B., and Pennachin, C., eds., *Artificial General Intelligence*. Springer-Verlag. 63–130.

Goertzel, B. 1993. *The Evolving Mind*. Gordon and Breach.

Goertzel, B. 1994. *Chaotic Logic*. Plenum.

Goertzel, B. 1997. *From Complexity to Creativity*. Plenum.

Hawkins, J., and Blakeslee, S. 2004. *On Intelligence*. Times Books.

Hedberg, S. R. 2005. Celebrating twenty-five years of AAAI: Notes from the AAAI-05 and IAAI-05 conferences. *AI Magazine* 26:3:12–16.

Holland, J. H. 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

Hutter, M. 2005. Universal algorithmic intelligence: A mathematical top-down approach. In Goertzel, B., and Pennachin, C., eds., *Artificial General Intelligence*. Springer-Verlag. 200–227.

James, W. 1890. *The Principles of Psychology*. Holt.

Koza, J. R. 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press.

Kurzweil, R. 2000. *The Age of Spiritual Machines*. Penguin.

Lenat, D. B., and Guha, R. V. 1990. *Building Large Knowledge-Based Systems*. Addison-Wesley.

Looks, M.; Goertzel, B.; and Pennachin, C. 2004. Novamente: An integrative architecture for artificial general intelligence. In *Proceedings of the AAAI Fall Symposium Series*. AAAI Press.

Looks, M.; Goertzel, B.; and Pennachin, C. 2005. Learning computer programs with the Bayesian optimization algorithm. In *Proceedings of the Genetic and Evolutionary Computation Conference*. Springer-Verlag.

Mareschal, D., and Johnson, S. P. 2002. Learning to perceive object unity. *Developmental Science* 5:2:151–185.

Marr, D. 1982. *Vision: a computational investigation into the human representation and processing of visual information*. W. H. Freeman.

Metzinger, T. 2004. *Being No One*. MIT Press.

Minsky, M. 1986. *The Society of Mind*. Simon and Schuster.

Mitchell, M. 1996. *An Introduction to the Genetic Algorithm*. Boston, MA: MIT Press.

Moravec, H. 1998. When will computer hardware match the human brain? *Journal of Evolution and Technology* 1.

Newell, A. 1990. *Unified Theories of Cognition*. Harvard University Press.

Peirce, C. S. 1892. The law of mind. In *The Collected Papers of Charles Saunders Peirce (1980)*. Harvard University Press.

Pelikan, M.; Goldberg, D. E.; and Cantú-Paz, E. 2002. Linkage problem, distribution estimation, and Bayesian networks. *Evolutionary Computation* 8(3):311–341.

Pelikan, M.; Goldberg, D. E.; and Lobo, F. G. 2002. A survey of optimization by building and using probabilistic models. *Comput. Optim. Appl.* 21(1).

Rowan, J. 1990. *Subpersonalities: The People Inside Us*. Routledge.

Rumelhart, D.; McClelland, J.; and PNP Research Group. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press.

Schmidhuber, J. 2005a. Gödel machines: Fully self-referential optimal universal self-improvers. In Goertzel, B., and Pennachin, C., eds., *Artificial General Intelligence*. Springer-Verlag. 228–291.

Schmidhuber, J. 2005b. The new AI: General & sound & relevant for physics. In Goertzel, B., and Pennachin, C., eds., *Artificial General Intelligence*. Springer-Verlag. 176–199.

Solomonoff, R. 1964. A formal theory of inductive inference. *Information and Control* 7:1:1–22.

Wang, P. 1995. *Non-Axiomatic Reasoning System: Exploring the Essence of Intelligence*. Ph.D. Dissertation, University of Indiana.