Seven Principles of Synthetic Intelligence

Joscha Bach[1]
Institute for Cognitive Science, University of Osnabrück, Germany

Abstract. Understanding why the original project of Artificial Intelligence is widely regarded as a failure and has been abandoned even by most of contemporary AI research itself may prove crucial to achieving synthetic intelligence. Here, we take a brief look at some principles that we might consider to be lessons from the past five decades of AI. The author's own AI architecture – MicroPsi – attempts to contribute to that discussion.

Keywords. Artificial General Intelligence, AI, Synthetic Intelligence, Psi theory, MicroPsi, Cognitive Architectures

Introduction

When the Artificial Intelligence (AI) movement set off fifty years ago, it bristled with ideas and optimism, which have arguably both waned since. While AI as a method of engineering has continuously and successfully served as the pioneer battalion of computer science, AI's tenet as a method of understanding and superseding human intelligence and mind is widely considered a failure, and it is easy to imagine that a visit to one of today's AI conferences must be a sobering experience to the enthusiasts of the 1950es. The field has regressed into a multitude of relatively well insulated domains like logics, neural learning, case based reasoning, artificial life, robotics, agent technologies, semantic web, etc., each with their own goals and methodologies. The decline of the idea of studying *intelligence per se*, as opposed to designing systems that perform tasks that would require some measure of intelligence in humans, has progressed to such a degree that we must now rename the original AI idea into *Artificial General Intelligence*. And during that same period of fifty years, support for that very idea declined outside computer science as well: where the cybernetics movement influenced the social sciences, the philosophy of mind and psychology, the world around us is now a place much more hostile to AI than in the past. The philosophy of mind seems to be possessed and enamored by "explanatory gaps" and haunted by the ghosts of the mystical "first person perspective" [1] and "irreducible phenomenal experience" [2], and occasionally even radical substance dualism [3, 4]. Attempts in psychology at overarching theories of the mind have been all but shattered by the influence of behaviorism, and where cognitive psychology as sprung up in its tracks, it rarely acknowledges that there is something as "intelligence per se", as opposed to the individual performance of a group of subjects in an isolated set of experiments.

[1] Corresponding author: Joscha Bach, Prenzlauer Allee 40, 10405 Berlin, Germany. E-mail: joscha.bach@gmail.com

AI's gradual demotion from a science of the mind to the nerdy playpen of information processing engineering was accompanied not by utterances of disappointment, but by a chorus of glee, uniting those wary of human technological hubris with the same factions of society that used to oppose evolutionary theory or materialistic monism for reasons deeply ingrained into western cultural heritage.

Despite the strong cultural opposition that it always met, the advent of AI was no accident. Long ago, physics and other natural sciences had subscribed to the description of their domains (i.e. the regularities in the patterns as which the universe presents itself to us) using formal languages. In the words of information science, this means that theories in the natural sciences had become computational.[2] By the 1950es, information processing hardware, theory and culture had progressed so far that the nascence of a natural science of mind as a computational phenomenon was inevitable. And despite the cultural struggles and various technological dead-ends that AI has run into, despite its failure as a science and its disfiguring metamorphosis into an engineering discipline, the author believes that it already has managed to uncover most of the building blocks of its eventual success. I will try to hint at some of these lessons.

The second and final section of this paper will focus on an architecture implementing motivation in an AI system. MicroPsi is a cognitive model that represents the author's attempt to contribute to the discussion of Artificial General Intelligence (AGI), and here, I will give a very brief overview.

Principles of synthetic intelligence

Understanding the apparent failure of AI as a science involves naming some of the traps it fell into, and participating in the endeavor of AGI will require highlighting some of AI's original creeds. Naturally, my contribution to this ongoing discussion is going to be incomplete, slightly controversial and certainly error-prone.

1. Build whole functionalist architectures.

There are two aspects to that slogan: First, we are in need of *functionalist architectures*. That is, we need to make explicit what entities we are going to research, what constitutes these entities conceptually, and how we may capture these concepts. For instance, if we are going to research emotion, simply introducing a variable named "anger" or "pity" will not do. Rather, we will need to explain what exactly constitutes anger and pity within the system of a cognitive agent. We will – among other things – need to acknowledge that anger and pity have objects that require the perception and representation of (social) situations, and equip our model with these. We will have to capture that anger or pity have very different ways of affecting and modulating perception, learning, action selection and planning, memory and so on – and we have to depict these differences. To explicate concepts underlying intelligence and mind is to get away from *essentialist intuitions* (for instance the idea that emotion, personhood,

[2] In the sense that natural sciences assume that formal (i.e. computational) theories are adequate to capture their respective subject, the universe itself is a computational phenomenon. This is not a strong claim as it may seem to some, because it merely entails that the universe presents itself as information patterns to the systemic interface of the experimenter with his or her domain, and that these patterns are both *necessary* and *sufficient* for the experiment's measurement.

normative behavior, consciousness and so on just *are*, and *are done by some module* or *correspond to some parameter*), and to replace them by a *functional structure* that produces the set of phenomena that we associate with the respective concepts.

Second, we need *complete* and *integrated* systems. Isolated properties will not do, for perception is intrinsically related to deliberation, deliberation to emotion, emotion to motivation, motivation to learning and so on. The attempt to reduce the study of intelligence to a single aspect, such as reasoning or representation is like reducing the study of a car-engine to combustion, temperature fluctuations or rotational movement.

2. Avoid methodologism

When we grow up to be AI researchers, we are equipped with the beautiful tools our computer science departments have to offer, such as graph theory, binary, modal and fuzzy logic, description languages, statistical methods, learning paradigms, computational linguistics, and so on. As we discover the power of these tools, they tend to turn into the proverbial hammers that make everything look like a nail. Most AI researchers that abandoned the study of intelligence did not do so because they ran into difficulties along that course, but because they turned to some different (worthy) subject, like the study of graph-coloring, the improvement of databases, the design of programming languages, the optimization of internet agents, the definition of ontologies. However, there is currently no reason to think that understanding intelligence will be a by-product of proving the properties of our favorite description language, or the application of our favorite planner to a new domain of the funding agencies choosing. We will need to ask questions and find methods to answer them, instead of the other way around.

3. Aim for the big picture, not the individual experiment

Our understanding of intelligence will have to be based on the integration of research of the cognitive sciences, possibly in a similar vein as the medieval and renaissance map-makers had to draw on the data made available by travelers, tradesmen, geographers, geometers and explorers of their times. Just as these map-makers pieced together a map of the world from many sources of data, we will have to draw a map of cognition and the mind by integrating the knowledge of many disciplines. Our current world maps are not the result of choosing a small corner of a small village and improving the available measurements there, because these measurements are not going to add up into a unified picture of geography. (Before that happens, the landscape is likely going to change so much as to make our measurements meaningless for the big picture.) Our first proper maps were not patchworks of infinitesimally small measurements, but the product of gradual improvements of a *big picture*.

Disciplines that are concerned with individual measurements often sport methodologies that are incompatible with sketching big pictures. Note that Albert Einstein did not do a single experiment whilst designing the theory of relativity – instead, he noted and expressed the constraints presented by the data that was already available. Likewise, the study of AGI aims at a unified theory, and such a theory is going to be the product of integration rather than specialization.

This point is likely a controversial one to make, since it seems to insinuate that the exploration of specific topics in AI is futile or irrelevant, which of course it not the case – it is just unlikely to result in an understanding of *general* intelligence.

4. Build grounded systems, but do not get entangled in the Symbol Grounding Problem

Early AI systems tended to constrain themselves to micro-domains that could be sufficiently described using simple ontologies and binary predicate logics [5], or restricted themselves to hand-coded ontologies altogether. It turned out that these approaches did not scale to capturing richer and more heterogeneous domains, such as playing a game of soccer, navigating a crowded room, translating a novel and so on. This failure has opened many eyes to the *symbol grounding problem* [6], i.e. how to make symbols used by an AI system refer to the "proper meaning". Because of the infinitude and heterogeneity of content that an intelligent system must be capable of handling to satisfy a multitude of conflicting and evolving demands (after all, intelligence is the answer to that problem), AI systems will have to be equipped with methods of autonomously making sense of their world, of finding and exploiting structure in their environment. Currently, it seems clear that binary predicate logic reasoners are not well equipped for that task, and mental content will have to be expressed using hierarchical spreading activation networks of some kind. AI systems will probably have to be *perceptual symbol systems*, as opposed to *amodal symbol systems* (see Fig. 1) [7], that is, the components of their representations will have to be spelled out in a language that captures the richness, fluidity, heterogeneity and affordance orientation of perceptual and imaginary content.

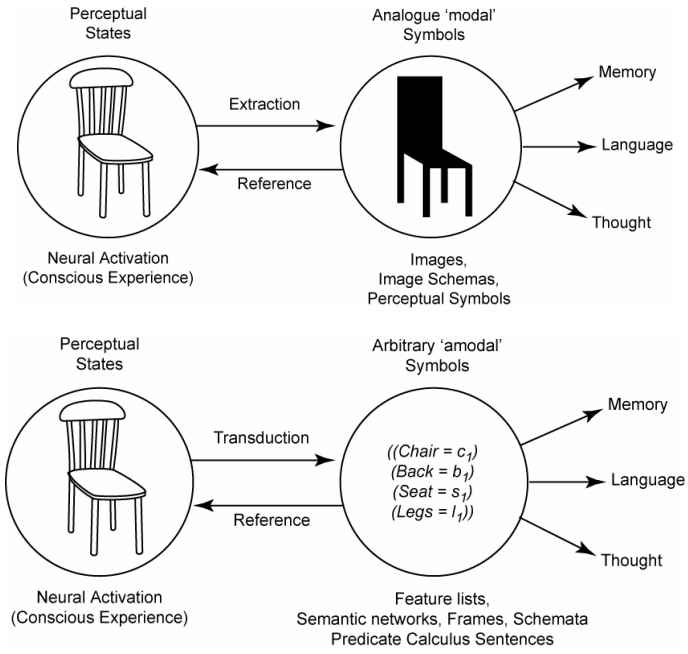

Figure 1: Modal representations, as opposed to amodal representations [7]

There is a different, stronger reading of the symbol grounding problem that has begun to haunt AI ever since Brooks' early approach of building simple physically embodied machinery [7], and which is well exemplified in John Searle's famous "Chinese room"

metaphor [8]. This reading expresses the intuition that "mere" symbol manipulation or information processing would never be able to capture the "true meaning" of things "in the real world". The symbol grounding problem has lead to the apostasy of those factions within the *"Nouvelle AI"* movement that came to believe that "a software agent can never be intelligent" [10, 11], as if only the divine touch of the "real reality" could ever infect a system with the mystical spark of knowing "true meaning". As a consequence, the protagonists of "Nouvelle AI" have abandoned the study of language, planning, mental representation in favor of pure, "embodied systems", such as passive walkers and insectoid robots.

5. Do not wait for the rapture of robotic embodiment

Even to the hardened eye of this author, it is fascinating to see a little robot stretching its legs. Eventually, though, the level of intelligence of a critter is not measured by the number of its extremities, but by its capabilities for representing, anticipating and acting on its environment, in other words, not by its brawns but by its brains. Insects may continue to rule the planet long after humankind has vanished, but that does not make them smarter than us. There may be practical questions to build robots instead of virtual agents, but the robotics debate in AI is usually not about practicality:

Unfortunately, a lot of research into AI robots is fueled by the *strong sense* of "meaning" originating in a Searle style conception of the *Symbol Grounding problem*. This sense of meaning, however, can itself not be grounded! For any intelligent system, whether a virtual software agent or a physically embodied robot (including us humans), the environment presents itself as a set of *dynamic patterns* at the systemic interface (for instance, the sensory[3] nerves). For all practical purposes, the universe is a pattern generator, and the mind "makes sense" of these patterns by encoding them according to the regularities it can find. Thus, the representation of a concept in an intelligent system is not a pointer to a "thing in reality", but a set of hierarchical constraints over (for instance perceptual) data. The encoding of patterns that is represented in an intelligent system can not be described as "capturing true meaning" without the recourse of epistemologically abject realist notions; the quality of a world model eventually does not amount to how "truly" if depicts "reality", but how adequately it encodes the (sensory) patterns.[4]

Even though the advocates of *Strong Symbol Grounding* are mistaken, and there is no epistemological reason why the set of patterns we associate with our concept of a physical universe (i.e. "real things") and that we feed into our AI model should not originate in an artificial pattern generator (such as a virtual world), there are practical difficulties with purely virtual agents: Virtual environments tend to lack richness of presentation, and richness of internal structure.

Where experimenters specify virtual environments, they usually encode structures and details with certain pre-specified tasks and ontologies in mind, thereby restricting the AI agent situated in such an environment to the re-discovery of these tasks and

[3] Note that the perceptual input of a system is completely made up of sensory input, for it can perceive its output only insofar it is provided by additional sensors. So, without loss of generality, sensory data stemming from sensor-actor coupling of a system are just a specific sub-class of sensory data in general. This is by no means a statement on how an AI system should treat sensor-actor coupling, however.

[4] The adequacy of an encoding over the patterns that represent an environment can be measured in terms such as completeness, consistency, stability, sparseness, relevance to a motivational sub-system and computational cost of acquisition.

limited ontologies and depriving it of opportunities for discovery and invention. Hand-crafted virtual environments (such as virtual soccer [12] or role-playing game worlds) are probably much too simplistic to act as a benchmark problem for AGI. Limited real-world problems, such as robotic soccer or the navigation of a car through a desert, suffer from the same shortcoming. If we take our agents from the confines of a virtual micro-world into the confines of a physical micro-world, the presented environment still falls short on establishing a benchmark that requires AGI.

On the other hand, there are virtual environments in existence that sport both structural and presentational richness to a degree comparable to the physical and social world, first among them the World Wide Web. Even the ocean of digitized literature might be sufficient: Humankind's electronic libraries are spanning orders of magnitude more bits of information than what an individual human being is confronted with during their lifetime, and the semantics of the world conceptualized in novels and textbooks inherits its complexity from the physical and social environment of their authors. If it is possible for an intelligence system to extract and encode this complexity, it should be able to establish similar constraints, similar conceptual ontologies, as it would have while residing in a socially and physically embedded robotic body.

Robots are therefore not going to be the singular route to achieving AGI, and successfully building robots that are performing well in a physical environment does not necessarily engender the solution of the problems of AGI. Whether robotics or virtual agents will be first to succeed in the quest of achieving AGI remains an open question.

6. Build autonomous systems

As important as it is to integrate perception, memory, reasoning and all the other faculties that an intelligent system employs to reach its goals is integration of goal-setting itself. General intelligence is not only the ability to reach a given goal (and usually, there is some very specialized, but non-intelligent way to reach a singular fixed goal, such as winning a game of chess), but includes the setting of novel goals, and most important of all, about exploration. Human intelligence is the answer to living in a world that has to be negotiated to serve a multitude of conflicting demands. This makes it a good reason to believe that an environment with fixed tasks, scaled by an agent with pre-defined goals is not going to make a good benchmark problem for AGI.

The motivation to perform any action, such as eating, avoiding pain, exploring, planning, communicating, striving for power, does not arise from intelligence itself, but from a motivational system underlying all directed behavior. In specifying a motivational system, for instance as a set of conflicting drives, we have to make sure that every purposeful action of the system corresponds to one of its demands; there is no reason that could let us take behavioral tendencies such as self-preservation, energy conservation, altruistic behavior for granted – they will have somehow to be designed into the system (whereby 'somehow' includes evolutionary methods, of course).

7. The emergence of intelligence is not going to happen all by itself

While the proposal of AGI or *synthetic intelligence* is based on a computational monism, [5] dualist intuitions are still widespread in western culture and in the contemporary philosophy of mind, and they are not going to give in without a fight. Because a naked ontological dualism between mind and body/world is notoriously hard to defend, it is sometimes covered up by wedging the popular notion of *emergence* into the "explanatory gap" [13]. Despite the steady progress of neuroscience and computational models of neural activity, there is an emergentist proposal that assumes so-called *"strong emergence"*, which proposes that the intelligent mind, possibly including human specifics such as social personhood, motivation, self-conceptualization and phenomenal experience, are the result of non-decomposable intrinsic properties of interacting biological neurons, or of some equally non-decomposable resonance process between brains and the physical world. Thus, "strong emergence" is basically an anti-AI proposal.

Conversely, *"weak emergence"* is what characterizes the relationship between a state of a computer program and the electrical patterns in the circuits of the same computer, i.e. just the relationship between two modes of description. In that sense, emergent processes are not going to "make intelligence appear" in an information processing system of sufficient complexity. We will still need to somehow (on some level of description) implement the functionality that amounts to AGI into our models.

This brief summary of principles of synthetic intelligence does not answer the main question, of course: How do we capture the functionality of Artificial General Intelligence? – In cognitive science, we currently have two major families of architectures, which seem to be hard to reconcile. One, the classical school, could be characterized as *Fodorian Architectures*, as they perceive thinking as the manipulation of a language of thought [14], usually expressed as a set of rules and capable of recursion. Examples, such as ACT [15] and Soar [16] are *built* incrementally by adding more and more functionality, in order to eventually achieve the powers inherent to general intelligence. The other family favors distributed approaches [17, 18] and *constrains* a dynamic system with potentially astronomically many degrees of freedom until the behaviors tantamount to general intelligence are left. This may seem more "natural" and well-tuned to the "brain-level" of description, because brains are essentially huge dynamical systems with a number of local and global constraints imposed on them, and the evolution of brains from mice-sized early mammalians to *homo sapiens* has apparently not been a series of incremental functional extensions, but primarily a matter of scaling and local tuning. Yet many functional aspects of intelligence, such as planning and language, are currently much harder to depict using the dynamical systems approach.

The recent decade has seen the advent of several new architectures in AI, which try to combine both approaches in a *neuro-symbolic* fashion, such as Clarion [19], LIDA [20], the MirrorBot [21] and the author's own MicroPsi [22], which will briefly be introduced on the remaining pages.

[5] Computational monism itself amounts just to the subscription of contemporary materialism. Cartesian matter ('res extensa') sports unnecessary intrinsic properties, such as locality and persistence, which get in the way when doing contemporary physics. Today's matter is not the same wholesome solid as it used to be in Laplace's time and day; now it is just a shifty concept that we apply to encode the basic regularities in patterns presented to us by our environment.

The Psi theory and the MicroPsi architecture

MicroPsi [22, 23] is an implementation of Dietrich Dörner's *Psi theory* of mental representation, information processing, perception, action control and emotion [24] as an AI architecture. MicroPsi is an attempt to embody the principles discussed above:

1. MicroPsi aims at explaining intelligence by a **minimal orthogonal set of mechanisms** that together facilitate perception, representation, memory, attention, motivation, emotion, decision-making, planning, reflection, language. These features are not explained as parameters or modular components, but in terms of the **function of the whole system**; for instance, emotions are explained as specific *configurations* of cognitive processing rather than as given parameters; behavior is not the result of pre-defined goal directed routines, but of a demand-driven motivational system and so on.

2. An integrated cognitive architecture will require the recourse to **methods from many disciplines**; MicroPsi originates in theories of problem solving in psychology and integrates ideas from gestalt theory, motivational psychology and experiments in the study of emotion. Also, it has learned a lot from cognitive modeling paradigms and from representational strategies and learning methods in AI.

3. The facets of cognition are not seen as separate modules that could be understood, tested and implemented one by one – rather, they are aspects of a broad architecture. The model **combines a neuro-symbolic theory of representation** with a **top-down/bottom-up theory of perception**, a **hierarchical spreading activation theory of memory** with a **modulation model of emotion**, a **demand/drive based theory** of dynamic physiological, cognitive and social **motivation** with a model of the **execution** and regulation **of behaviors**.

4. Representations in MicroPsi are always **grounded in environmental interaction** or abstractions thereof. It does not matter, however, if the environment is simulated or physical.

5. The difficulty of providing a rich pre-programmed environment vs. the limitations that come with robot engineering have lead to ***both* simulation** worlds and **robotic** experiments for MicroPsi agents. At the current stage of development, we seem to learn much more from simulations, though.

6. MicroPsi agents are **autonomous**, their behavior is governed by a **set of primary urges** which determine motives which in turn give rise to intentions. All behavior, including cognitive exploration and social interaction, can be traced back to one or more primary urges.

7. There are many aspects of intelligence that MicroPsi does not address well yet. However, we do not believe that these will be automagically spring into existence with gradual improvements of the learning, representation or interaction modes of the model. We think that the **deficits of MicroPsi** highlight specific mechanisms, for instance for perceptual integration, language and self-monitoring, that we have not sufficiently understood to implement them. MicroPsi might propose some interesting answers, but more importantly, it helps to detail a lot of **useful questions**.

Representation in the Psi theory

The most basic elements in Dörner's representations are threshold elements, which are arranged into groups, called *quads*. These are made up of a central neuron, surrounded by four auxiliary neurons acting as gates for the spreading of activation through the

network. A network of quads amouts to a semantic network with four link types, called *SUB*, *SUR*, *POR* and *RET*. *SUB*: stands for "has-part". If an element *a* has a SUB-link to an element *b*, it means that *a* has the part (or sometimes the property) *b*. *SUR* is the inverse relation to SUB and means "is-part". If *a* is SUR-linked to *b*, then *a* is a part (or sometimes a property) of *b*. *POR* (from latin *porro*) is used as a causal (subjunctive), temporal or structural ordering relation between adjacent elements. If *a* has a POR-link to *b*, then *a* precedes (and sometimes leads to or even causes) *b*. *RET* (from latin *retro*) is the inverse relation to POR. If there is a RET-link between *a* and *b*, then *a* succeeds (and sometimes is caused by *b*.

Quads make up perceptual schemas and frames: Individual quads stand for concepts. If they are SUB/SUR linked, a partonomic (has-part/is-part) relationship is expressed. The lowest level of such a partonomic tree is made up by *perceptual* neurons and *motor* neurons. They provide the grounding of the system's representations, since they are directly linked to the system's environment.

In MicroPsi, quads are extended to cover more basic relations and are expressed by *concept nodes*. In addition to the basic POR/RET and SUB/SUR links, they also offer link types for taxonomic and labeling relationships.

Object schemas are organized as parts of situational frames. The world model of the system is established, reinforced or changed by *hypothesis based perception* ("hypercept"). Hypercept is a paradigm that might be described as follows:

- Situations and objects are always represented as hierarchical schemas that bottom out in references to sensory input.
- Low-level stimuli trigger (bottom-up) those schema hypotheses they have part in.
- The hypotheses thus activated heed their already confirmed elements and attempt (top-down) to get their additional elements verified which leads to the confirmation of further SUB-hypotheses, or to the rejection of the current hypothesis.
- The result of hypercept is the strongest activated (matching) hypothesis.

At any time, the system *pre-activates* and *inhibits* a number of hierarchical schema hypotheses, based on context, previous learning, current low-level input and additional cognitive (for instance motivational) processes. This pre-activation speeds up the recognition by limiting the search space.

Hypercept is not only used on visual images, but also on inner imagery, memory content, auditory data and symbolic language.

The current situational frame is stored as the head element of a growing protocol chain, which is formed by decay and re-arrangement into long-term memory.

Behavior Execution

A neuron that is not part of the currently regarded cortex field is called a *register*. Neural programs are chains of registers that call associators, dissociators, activators and inhibitors. (These "calls" are just activations of the respective elements.) In the course of neural execution, elements in the cortex field are summarily linked to specific registers which are part of the executed chain of neurons. Then, operations are performed on them, before they are unlinked again.

Dörner describes a variety of cognitive behaviors for orientation, anticipation, planning, memory retrieval and so on, often along with possible implementations. [24, 25]

During action execution, the system establishes motives based on a set of primary urges which are hard-wired. Currently, these urges consist of demands for *fuel*, *water*, *intactness*, *affiliation*, *competence* and *certainty*.

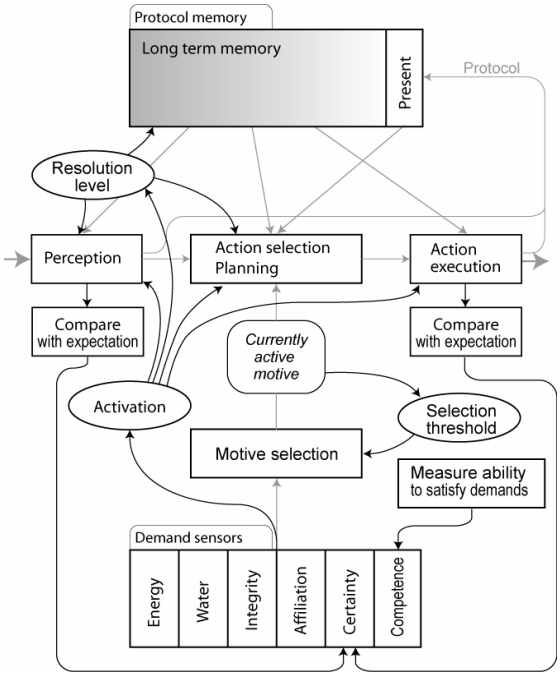

Figure 2: Psi architecture

Fuel, water and intactness are examples of physiological needs. Affiliation is a *social* urge – to satisfy it, the system needs *affiliation signals* from other agents. Thus, Psi agents may reward each other. *Competence* measures the ability to reach a given goal and the ability to satisfy demands in general (coping potential). The urge for *certainty* is satisfied by successful exploration of the environment and the consequences of possible actions, and it is increased by violations of expectations. Competence and certainty are *cognitive urges*. Together they govern explorative strategies.

Every increase of an urge creates a negative reinforcement signal, called *displeasure signal*; conversely, a decrease of an urge results in a *pleasure signal*. These signals are used to strengthen links in the current protocol and thus enable reinforcement learning of behavior. At any time, the system evaluates the urge strengths and, based on an estimate of the competence for reducing individual urges, determines a *currently active motive*. This motive pre-activates memory content and behavior strategies and is used in determining and executing a plan to achieve it.

To adapt the cognitive resources to the situation at hand, the system's activity is influenced by a set of *modulators*. Within the Psi theory, a configuration of modulator settings (together with appraisals of certainty, competence and current pleasure/displeasure status) is interpreted as an emotional state [25].

The MicroPsi Framework

MicroPsi has been developed in an AI context and offers executable neural representations, multi-agent capabilities and visualization tools. The author's group has used it to model perceptual learning [26], the evolution of motivational parameters in an artificial life setting and as an architecture for controlling robots.

The framework consists of the following components: a graphical editor for designing executable spreading activation networks (which make up the Psi agent's control structures and representations), a network simulator (integrated with the editor and monitoring tools to log experiments), an editor and simulator for the agent's environment and a 3D viewer which interfaces with the simulation of the agent world.

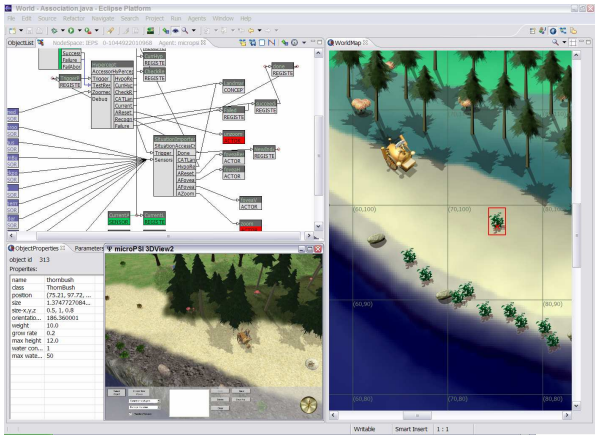

Figure 3: MicroPsi network editor and agent environment

MicroPsi has matured into a runtime environment for various cognitive modeling tasks. Among other things, we are using it for

- Building agents according to the Psi theory. These agents are autonomous systems with a set of innate urges, situated in a simulated environment, where they perceive objects using a simplified *hypercept* approach. Perceptual content is represented as partonomic schema descriptions and arranged into protocols, which are later retrieved to generate plans to satisfy the agent's urges.

- Performing neural learning using hybrid representations. To connect higher, gradually more abstract layers within MicroPsi network representations to real-world sensory input, we are setting up a matrix of foveal sensor nodes, which correspond to pixels in the camera of a robot. By moving the foveal field through the camera image, the image is scanned for salient features. Using backpropagation learning, we are training it to identify edge segments in the camera image, which in turn make up the lowest layer in an instantiation of *hypercept*.

- Evolving motivational parameter settings in an artificial life environment. Here, groups of MicroPsi agents jointly explore their territory and cooperate to find and defend resources. Suitable cooperative behaviors are evolved based on mutations over parameters for each urge and modulator in accordance to the given environment.

- Implementing a robotic control architecture using MicroPsi. A simplified neurobiological model of behavior and perception of mice in a labyrinth is mimicked using Khepera robots that are embodied MicroPsi agents.

Since its beginnings in 2001, the MicroPsi framework and the associated cognitive architecture have come a long way. Even though MicroPsi is far from its goal – being a broad and functional model of human cognition and emotion – it fosters our understanding and serves as a valuable tool for our research.

Acknowledgments

This work has been supported by the University of Osnabrück. I am indebted to Ronnie Vuine, who vitally contributed to the technical framework and the MicroPsi agents, to Matthias Füssel and Daniel Küstner, who built the environment simulator, to David Salz who is responsible for 3D visualization components, to Markus Dietzsch, who is performing the artificial life experiments, and to Daniel Weiller and Leonhardt Laer, who are in charge of robotic experiments. Also, I want to thank two anonymous reviewers for their constructive criticism.

References

[1] Jackson, F. (1982). Epiphenomenal qualia. Philosophical Quaterly 32: 27-36
[2] Block, N. (2001). Paradox and Cross Purposes in Recent Work on Consciousness. In Dehaene and Naccache: Special Issue of *Cognition*, Vol. 79, The Cognitive Neuroscience of Consciousness, 197-219
[3] Chalmers, D. (1996): The Conscious Mind. New York: Oxford University Press
[4] Searle, J. R. (1992): The Rediscovery of the Mind, MIT Press, Cambridge
[5] Dreyfus, H. L. (1992): What Computers still can't do. A Critique of Artificial Reason. MIT Press
[6] Harnad, S. (1990). The symbol grounding problem. Physica D, 42, 335-346.
[7] Barsalou, L.W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22, 577-609.
[8] Brooks, R. A. (1991): Intelligence Without Reason, IJCAI-91
[9] Searle, J. R. (1980): Minds, brains, and programs. Behavioral and Brain Sciences 3 (3): 417-45
[10] Pfeifer, R., Stieler, W. (2007): „Es geht um den physikalischen Prozess." Technology review, online at http://www.heise.de/tr/artikel/95539, Oct 1st 2007
[11] Pfeifer, R., Bongard, J. (2006): How the body shapes the way we think. MIT Press
[12] Noda, I (1995). Soccer Server: A Simulator of Robocup. In Proceedings of AI symposium 1995, Japanese Society for Artificial Intelligence.
[13] Stephan, A. (1999). *Emergenz: Von der Unvorhersagbarkeit zur Selbstorganisation.* Dresden Univ. Press
[14] Fodor, J. A. (1975). *The Language of Thought*, Cambridge, Massachusetts: Harvard University Press.
[15] Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.
[16] Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, *33*(1), 1-64.
[17] Rumelhart, D. E., McClelland, J. L. and the PDP Research Group (1986): Parallel Distributed Processing, (Vols. 1&2), Cambridge, Massachusetts: MIT Press
[18] Smolensky, P., Legendre, G. (2005): The Harmonic Mind. From Neural Computation to Optimality-theoretic Grammar, Vol. 1: Cognitive Architecture, MIT Press
[19] Sun, R. (2003) A Detailed Specification of CLARION 5.0. Technical report.
[20] Franklin, S. (2007). A foundational architecture for Artificial General Intelligence. In Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms, Proceedings of the AGI workshop 2006, ed. Ben Goertzel and Pei Wang: 36-54. Amsterdam: IOS Press.
[21] Borst, M., Palm, G. (2003): Periodicity Pitch Detection and Pattern Separation using Biologically Motivated Neural Networks. In: Proc. 6. Tübinger Wahrnehmungskonferenz, 52
[22] Bach, J. (2003). The MicroPsi Agent Architecture. Proceedings of ICCM-5, International Conference on Cognitive Modeling, Bamberg, Germany, 15-20
[23] Bach, J. (2008): PSI – An architecture of motivated cognition. Oxford University Press.
[24] Dörner, D. (1999). *Bauplan für eine Seele*. Reinbeck
[25] Dörner, D., Bartl, C., Detje, F., Gerdes, J., Halcour, (2002). *Die Mechanik des Seelenwagens. Handlungsregulation*. Verlag Hans Huber, Bern
[26] Bauer, C. (2005): Kategorielernen im MicroPsi-Agenten. Diploma Thesis, Technische Univ. Berlin