Beyond Artificial Intelligence toward Engineered Psychology

Stevo Bozinovski[1] and Liljana Bozinovska[2]

[1] Mathematics and Computer Science Department
[2] Biological and Physical Sciences Department
South Carolina State University
300 College street, Orangeburg, SC, 29117
sbozinovski@scsu.edu, lbozinov@scsu.edu

Abstract. This paper addresses the field of Artificial Intelligence, road it went so far and possible road it should go. The paper was invited by the Conference of IT Revolutions 2008, and discusses some issues not emphasized in AI trajectory so far. The recommendations are that the main focus should be personalities rather than programs or agents, that genetic environment should be introduced in reasoning about personalities, and that limbic system should be studied and modeled. Engineered Psychology is proposed as a road to go. Need for basic principles in psychology are discussed and a mathematical equation is proposed as fundamental law of engineered and human psychology.

Keywords: Artificial Intelligence, Consequence-driven Systems theory, personality, Engineered Psychology, limbic system model, fundamental equation of engineered psychology, measuring units in psychology.

1 Introduction

Engineered Psychology is understood as building artificial creatures with human personality features. It considers behavior in a multidimensional *Aristotelian space*, containing at least coordinates of cognition, emotion, and willingness.

In the sequel first we give overview of the Artificial Intelligence development by decades of time starting with 1940-ies ending with current interest toward e-intelligence. Along we present a view on the work done within the ANW research group which is not very well known in the history of AI, in comparison to the work of the PDP group. The overview is then given in terms of phases the AI passed, concluding that now the interest is toward personalities rather than agents. Then we will present our Engineering Psychology approach, starting with the Consequence-driven Systems theory which is origin of the approach and presenting also a model of the brain limbic system. The first result of Engineered Psychology research is the mathematical equation which we call fundamental equation of engineered (and human) psychology. Finally we point out a need for measuring units in engineered and human psychology.

M. Ulieru, P. Palensky, and R. Doursat (Eds.): IT Revolutions 2008, LNICST 11, pp. 171–185, 2009.

2 Artificial Intelligence, So Far

Artificial Intelligence (AI) is contemporary well established area of Computer Science. Universities around the world regularly offer undergraduate course in AI within Computer Science programs. A course usually contains AI history and state of the art presentation of topics related to AI. Several textbooks provide overview of the field [25], [68], [63], [39], [48].

Many authors contributed toward establishing AI as a recognized discipline. The forerunners among them [10], [11], [24], [50], [35], [75], [71], [62], [53], [58], [59], [8], [66] contributed even before the term “Artificial Intelligence” was is use. After 1958 authors [69], [70], [78], [54], [2], [60], [72], [49], [32], [55], pointed toward various approaches toward AI. At the end of this period, due to the influential work [55], one of the approaches toward AI, the neural network approach was understood as restricted, potentially non-promising, which affected funding of that line of research. In 1970-ties, contributors [80], [28], [1], [26], [37], [81], [9], [76], [51], continued showing interest in various aspects of representing and manifesting AI. Particular belief was shown that special programming languages such as LISP and PROLOG should be used to foster development of AI programs.

The decade of 1980-ties is arguably the revolutionary step toward so called “new AI” or “embodied AI” or “behavior based AI” as distinction to the previous “disembodied AI”. The works such as [5], [14], [6], [33], [38], [36], [22], [23], [67], [56], [73], [74], [77], [31] contributed toward importance of agent-environment interaction. Parallel programming was introduced as a way of representing that interaction [13] as well as the need of sensors and actuators in an agent-environment interface [22]. In this period the neural network research was revived. We would emphasize the work of two organized group working in that direction: the Adaptive Networks Group (ANW) and the Parallel Distributed Processing (PDP) group. In stories about AI the work of PDP group is well known: they published an important volume on Parallel Distributed Processing (PDP) paradigm and explicitly rehabilitated the neural networks fields as very promising one within AI. The ANW group never published a volume together as a group so work as a group was not recognized within AI community. In the next section we will give a review of that work in the early 1980-ies.

During 1990-ties single important event overshadowed others: On May 11, 1997, computer program Big Blue beat the human chess champion! The first grand challenge of AI was no more! Consequently, a new grand challenge for AI was set up: Is it possible that a robot team in (European) football of 11 players would beat a corresponding human team [3]? And variation of that: Is it possible that a trained (rather than programmed) team [16] would achieve the same? Among other events in 1990-ies, it was pointed out that AI is not artificial by itself, it should be considered a natural intelligence in artificial creatures [4]. Increased interest has been shown toward emotions and its relation to intelligence ([41], [65], [42], [61], [16]).

In the 2000-ies early workers of AI such as [27], [56] also devoted books related to emotion. Emotion became topic of graduate Computer Science seminars [82]. Motivation and emotion became modules of control architectures. But most important phenomenon is the spread of AI among high school through robotics competitions. Robotics and related AI issues become widespread challenge for talents all over the world. Also

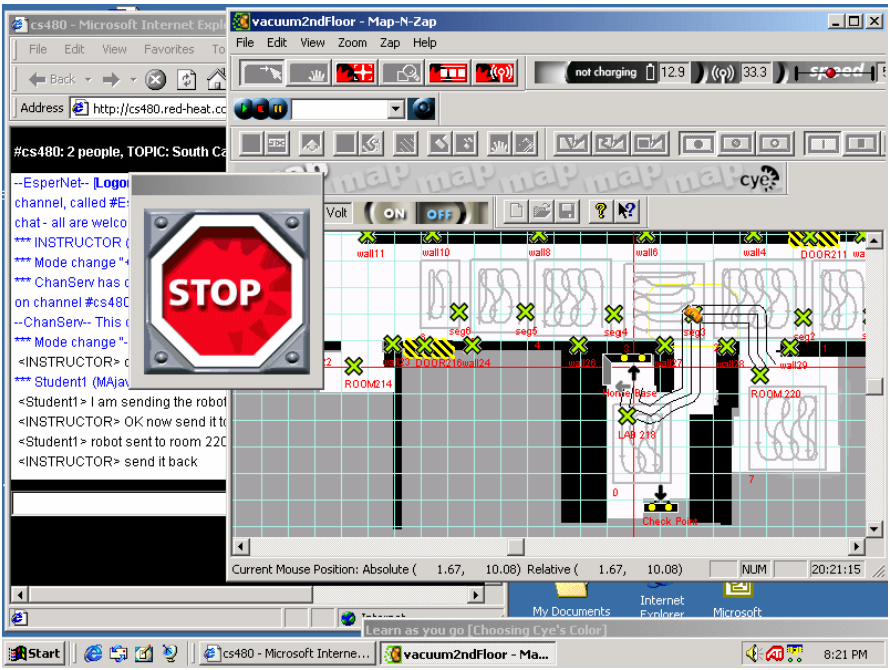

Fig. 1. A e-controlled mobile robot, cognitive model of its world

Internet and e-robotics (e-control of robots) is widely used. Fig. 1 shows our work with a commercially available e-controllable cye-type robot.

The robot is user programmable to have internal model of its environment, but is also able to learn new obstacles. The environment shown here is the AI/Robotics/Biocybernetis lab at South Carolina State University with corridor and other rooms. In particular Fig. 1 shows a session of our CS480 Introduction to Robotics class in distance learning using e-robotics [18]. Students are able using Internet to send the robot to a particular environment region to perform a vacuum cleaning task. In this particular e-robotics experiment the robot is controlled from a classroom in Europe. Fig. 1 shows how robot is executing a trajectory which it computes itself to avoid obstacles (dark areas) in order to reach its vacuum cleaning region. While the robot is moving the STOP sign is present so the robot can be stopped by click of a mouse. Left side of Fig. 1 shows the virtual classroom in which the instructor in USA communicates with students in Europe in real time. The experiment includes also a camera (not shows in Fig. 1) mounted on the robot so the students can scan the environment while the robot moves.

2.1 The ANW Group in Early 1980-ies

In early 1980-ties, possibly due to belief that artificial neural networks line of research is not promising one, in the USA there were not many organized groups working on the subject. In fact we know of only one: the Adaptive Networks (ANW) group, funded by Wright-Patterson Air Force Base (WPAFB) from Dayton, Ohio. The group was assembled by the Computer and Information Science (COINS) department of the University of Massachusetts at Amherst. Nico Spinelli was the group leader, who together with Michael Arbib led a Center for Systems Neuroscience within the department. During 1980-81 the group included the postdoc Andrew Barto, and the graduate students Richard Sutton, Charles Anderson, Jack Porterfield, Ted Selker, and Stevo Bozinovski. The main driving force was the WPAFB Program Officer, Harry Klopf, and his motto "*goal-seeking systems from goal-seeking components*" [40]. The neural networks paradigm well corresponds to that motto and the group pursued that direction. The group mostly oriented its research toward agent-environment type of research such as agents moving in an environment and agents moving objects in an environment. Maintaining some effort on classical conditioning and patters classification, the group focused on the concept of reinforcement and reinforcement learning ([53], [52], [64], [79]). The initial important result in this direction was the Associative Search Network (ASN), an agent that was able to learn to navigate toward a goal place using landmarks [5]. Next, the group took a challenge of *delayed reinforcement* learning, learning in cases where there is no immediate reinforcement (punish/reward) from the environment, rather it comes after several steps. That challenge led to introduction of emotions and genetics in neural nets and we present it here in more details.

Two instances of the problem were considered: the maze learning problem and the learning for inverted pendulum balancing problem. Two neural architectures were proposed for solving the problem: the Actor/Critic architecture, and the Crossbar Adaptive Array (CAA) architecture (Fig. 2). It is interesting to analyze the difference between the two architectures. The obvious difference is that A/C architecture needs two identical memory structures, V and W, to compute the internal reinforcement r^, and the action, while CAA architecture uses only one memory structure, W, the same size as one of the A/C memory structures. The most important difference however is in the *design philosophy*: In contrast to A/C architecture, *CAA architecture does not use any external reinforcement r*; it only uses the current situation X as input. A state value is used as secondary reinforcement concept, while genetically predefined state value, rather than an immediate reinforcement, is used as primary reinforcement. The CAA architecture *introduces the concept of state evaluation* and connects it to *the concept of feeling*. Although both architectures were designed to solve the same basic problem, the maze learning problem, the challenges of the design were different. The A/C architecture effort was challenged by the mazes from animal learning experiments, where there are many states and there is only one rewarding state (food). The CAA architecture effort from the start was challenged by the mazes defined in the computer game of Dungeons and Dragons, where there is one goal state, but many reward and punishment states along the way. So, from the very beginning CAA effort adopted the concept of dealing with *pleasant and unpleasant states*, feelings and emotions.

The CAA approach proved more efficient: it was the only architecture that presented solutions to both instances of the delayed reinforcement learning problem in 1981 ([12], [13], [14], [6], [15]). The original CAA idea of having *one memory structure for crossbar computation of both state evaluations and action evaluations* was later also implemented in reinforcement learning architectures such as Q-learning system [7], [77] and its successor architectures. The CAA memory was named Q-table.

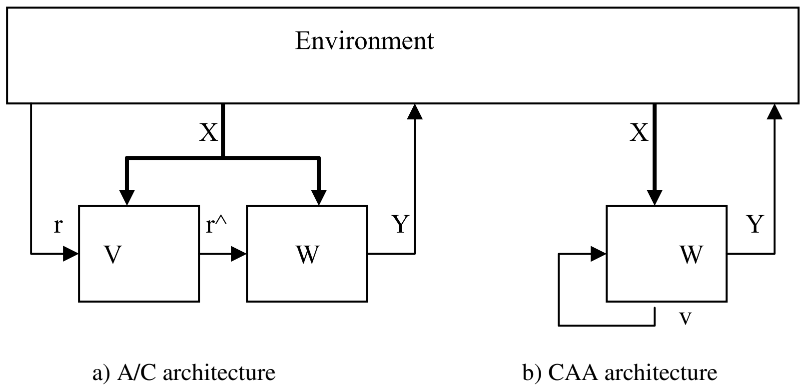

Fig. 2. ANW neural architectures for solving delayed reinforcement learning problem

The CAA approach introduced neural learning systems that can *learn without external reinforcement*. It proposes a *paradigm shift* in learning theories, from the concept of reinforcement (reward, punishment, payoff, …) to the concept of *state as a consequence* and *feeling as a state evaluation*. Including concepts like feelings and emotions, the CAA approach introduced emotive abilities in cognitive neural based agents.

2.2 Phases of AI Development: Programs, Agents, Personalities

Reviewing trajectory of the AI research so far we can distinguish three phases. The initial, zero-phase would be the cybernetics phase, where researchers were mostly interested in control structures, adaptation, and pattern recognition.

The first phase is so called representational (or disembodied) phase where researchers were mostly interested in representations of environments in internal knowledge structures of agents. The principal metaphor was information processing and AI programs, including special programming languages. Fig. 3 shows the principal postulates of AI understanding:

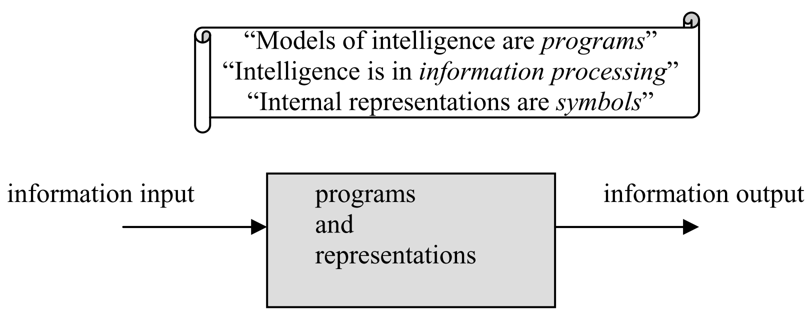

Fig. 3. Understanding artificial intelligence, representational phase

Second phase emphasizes behavior as representation of intelligence. Fig. 4 shows the behavioral metaphor and its main postulates of AI understanding. This phase is also known the embodied phase of AI. In order to sense the reality an AI agent needs sensors, and to wear sensors an agent needs a body.

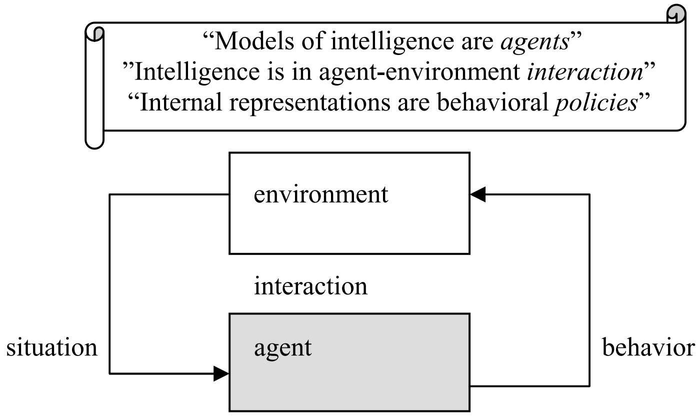

Fig. 4. Understanding artificial intelligence, behavioral (embodied) phase

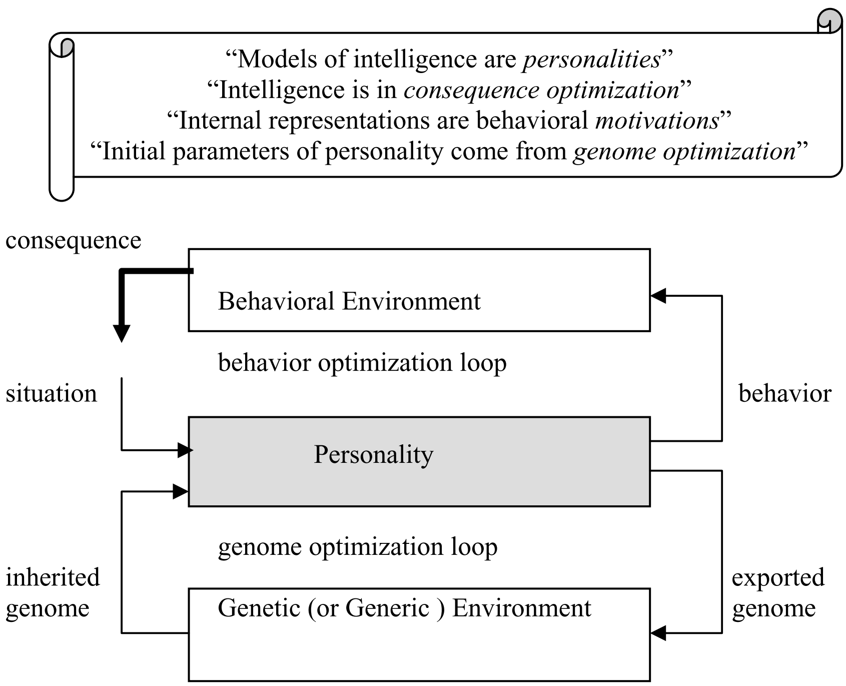

Fig. 5. Understanding Artificial Intelligence, personalities phase

The third phase we propose is the personality phase. Fig. 5 shows the personality metaphor and its principal postulates. The concept of personality assumes comparison of two behaviors and features that distinct one personality from another one. It is assumed that a personality cannot be considered separated neither from its behavioral environment nor from its genetic (generic) environment. Genome optimization loop (evolution) ensures optimization of the initial parameters of personality and its connection to the changing behavioral environment.

3 Engineered Psychology Rather Than Artificial Intelligence

Personality can be understood as *expectancy of a consistent pattern of attitudes and behavior.* All people exhibit recognizable individual actions that serve to identify them. Classical problems in personality include: Where do those individual characteristics come from? Are they truly unique or just particular combination of characteristics all people possess? Are they learned, inherited, or both? Can personality be altered, and if so, how? [46]

For many years Artificial Intelligence research attention was primarily given to the *concept of intelligence,* not necessarily an embodied one. In recent years, growing interest has been in building agents that will interact with humans, and will exhibit features of personality. The effort can be described as understanding the *concept of personality* and building an *artificial personality*, with its features like emotions and motivations, among others. In the sequel we address some of the challenges of artificial personality research within our theory of Consequence-driven Systems.

3.1 Consequence-Driven Systems Theory

Consequence-driven Systems theory [14], [17], [20], [21] is an attempt to understand and build an *agent (animat) personality*. It tries to *find an architecture* that will ground the notions such as motivation, emotion, learning, disposition, anticipation, curiosity, confidence, and behavior, among other notions usually present in a discussion about an agent personality. Searching through the literature we found relevant works that can be considered as grounds on which our work is a continuation. In particular we found a highly relevant connection with the works [43], [44], [45], [47]. Among several related efforts in contemporary research on emotion-based architectures, we distinguish the work [29], [30] by its basic concepts, architecture engineering, and realization aspects that resulted in developed an emotion learning architecture that has been implemented successfully in simulated as well as real robots. Starting from concepts such as feelings and hormonal system, the approach has similarity with our approach in issues like having innate emotions that will define goals and learning emotional associations between states and actions that will determine the agent's future decisions.

In the sequel we will briefly review the Consequence Driven Systems theory and its features.

3.1.1 Main Concepts of the Theory

The grounding postulates of the Consequence Driven Systems theory are: 1) there are three environments, 2) there are three tenses, 3) the neural system computes simultaneously both behaviors and emotions, from its memory 4) emotions are computed as state evaluations, 5) motivations are learned as behavior evaluations and 6) what is obtained should correlate to the brain limbic system.

Three environments. The theory assumes that an agent should always be considered as a three-environment system. The agent expresses itself in its *behavioral environment,* where it behaves. It has its own *internal environment* where it synthesizes its behavior. It also has access to a *genetic environment* where from it receives its initial conditions for existing in its behavioral environment. The genetic and the behavioral environment are related: The initial knowledge transferred through the imported *species (or personality) genome* properly reflects, *in the value system of the species*, the dangerous situations for the species in the behavioral environment. It is assumed that all the agents import some genomes at the time of their creation. Some of the agents will be able to export their genome.

Three tenses. The theory emphasizes that *an agent should be able to understand the temporal concepts,* like past, present*, and the future*, in order to be able to self-organize in an environment. The past tense is associated with the evaluation of its previous performance, the present tense is associated with the concept of emotion that the agent computes toward the current situation (or the current state), and the *future* with the *moral (self-advice)* about a behavior that in future should be associated with a situation. A *Generic Architecture for Learning Agents* (GALA architecture) is proposed within the theory (Fig. 6).

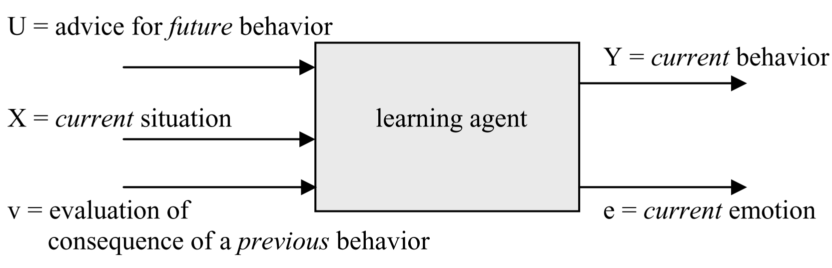

Fig. 6. The GALA architecture

Note that GALA is a genuine generic, black-box-only architecture: only the inputs and outputs are specified. Yet it is very specific the way inputs and outputs are defined. The GALA architecture is a reconfigurable architecture: The derivation of various types of learning architectures from the GALA architecture is described elsewhere [14], [19].

Crossbar-adaptive array architecture. The CAA architecture (Fig. 7) is derived as an emotion learning agent from the GALA architecture.

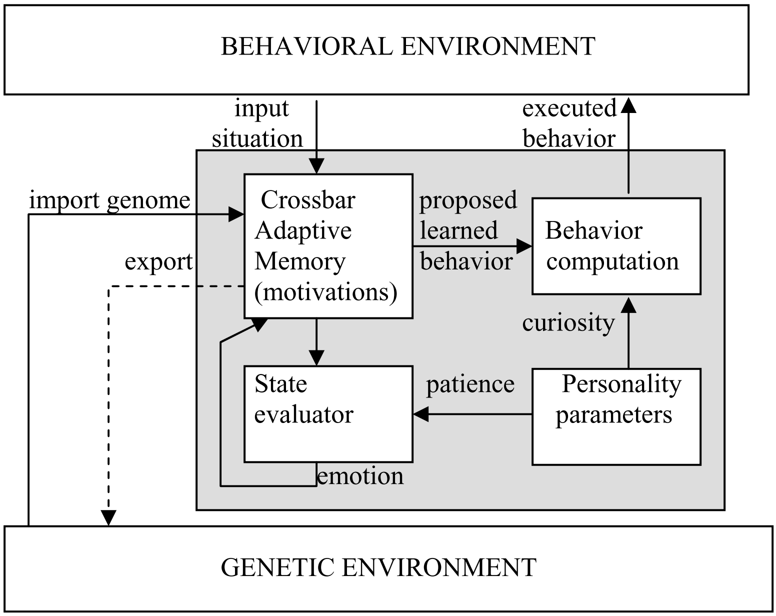

Fig. 7. Crossbar-adaptive Array (CAA) architecture

The CAA architecture is a personality architecture capable of learning using *backpropagated emotions*. In a crossbar fashion, the CAA architecture computes both state evaluations (emotions) and behavior evaluations (motivations). It contains three basic modules: crossbar-learning memory, state evaluator, and behavior selector. In its basic behavioral cycle, the CAA architecture firstly computes the emotion of being in the current state. Then, using a feedback loop, it computes the possibility of choosing again, in a next time, the behavior to which the current situation is the consequence. The state evaluation module computes the global emotional state of the agent and *broadcasts it* (e.g. by way of a neuro-hormonal signal) to the crossbar-learning memory. The behavior computation module using some kind of behavior-algebra initially performs a *curiosity driven*, default behavior, but gradually that behavior is replaced by a learned behavior. The forth module defines specific *personality parameters* of a particular agent, such as *curiosity*, *tolerance* threshold (*patience*), etc. Indeed CAA includes models of those personality paramets.

Correlation to the limbic system. A possible model of the limbic system that correlates to the CAA architecture is given in Fig. 8.

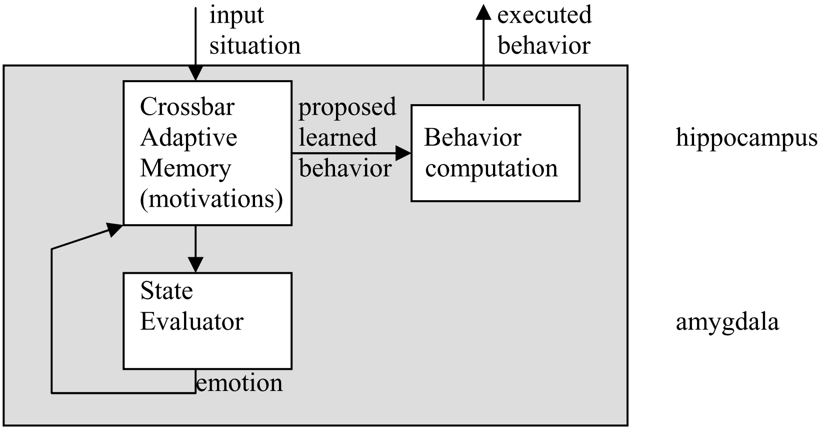

Fig. 8. Crossbar Adaptive Array architecture as model of the limbic system

The Crossbar Adaptive Array as model of the limbic system contains both the hippocampus level and amygdale level of processing.

Emotion. Is evaluation of a situation (or state). Usually that situation is observed as a consequence of a previous behavior.

Motivation. The motivation for a behavior is the anticipation of a future consequence of that behavior; the emotion is an evaluation system for computing the value of that consequence. The motivational system can be understood as a priority system for executing behaviors. Motivations are computed from the values of their (anticipated) consequence states. The Consequence-driven System theory introduced the *principle of learning and remembering only motivations (behavior evaluations), not the state evaluations.* There is no need to store the state values, since they can be computed from the (current) behavior values. The approach of storing action values rather than state values was also emphasized by Watkins [77], who discovered the relationship between Reinforcement Learning and Dynamic Programming [8]. There are different approaches of what an agent learns in an environment. It can learn the whole graph, like a *cognitive map*, or it can learn only a *policy* (the set of states and behaviors associated to those states). In case of policy learning the actual map (interconnection network between the states) is provided by the environment and is not stored within the agent.

3.2 Fundamental Law of Engineered (and Human) Psychology

Since beginning of philosophy and psychology, behavior, motivation for a behavior, and emotions related to a situation have been of interest for science. Mathematical models were proposed for those concepts. However it was only recently [19], [20] when the first mathematical model was proposed that *explicitly related* motivation, emotion, behavior and personality. Fig. 9 shows the approach toward the modeling. The approach is general and applies to both human and engineered psychology.

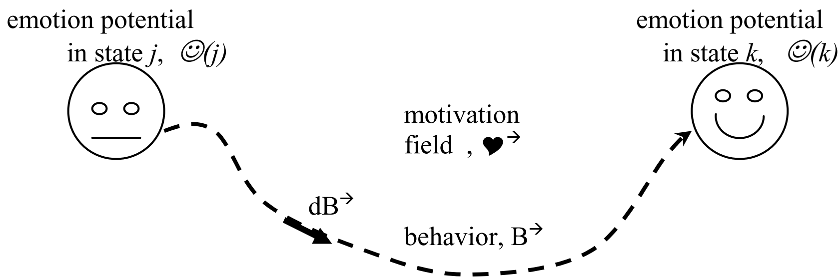

Fig. 9. Emotional space and motivational gradient

As Figure 9 shows, it is assumed that each state *j* of a system (e.g. animat) is assigned an *emotion potential ☺(j).* The emotion potential space defines by itself a vector field denoted as *motivational field*, ♥→. In such a way each point in the space is assigned a *motivation gradient* toward the highest emotional value (ultimate goal) at that region of the space.

$$\text{☺} = p \int_{j}^{k} \heartsuit^{\rightarrow} dB^{\rightarrow} \quad . \qquad (1)$$

In such a setup, a behavior $B^{\rightarrow}$ (that is motivated by the motivation field ♥→ induced by the emotional potential ☺) obeys the relation (1). Equation (1) connects the emotion potential in a current state *j*, an *anticipated emotion potential* in a next state *k*, the motivation of moving toward state *k*, and behavior toward the state *k*. Here $dB^{\rightarrow}$ is a differential segment of a behavior, an *action,* which sometimes might not be in the same direction as the motivational field, for example due to some obstacles (resistances) along the way. The constant *p* allows for some personality features of a particular agent. In other words the equation (1) is a mathematical way of saying that the *happiness is to be earned.* We propose this equation as a *fundamental equation of engineered (and human) psychology*.

Related to the equation (1) is an algebraic *motivation learning function* which can be written as

$$\heartsuit(b|s) \mathrel{+}= (\text{☺}(s') - \text{🕓}(b|s)) \qquad (2)$$

where += is the update operator we borrowed from object oriented languages for use in the simulation experiment described further in this text. The equation (2) simply states that *the motivation update is proportional to the satisfaction expected.* In more detail, it states that the motivation ♥(*b*|*s*) of performing behavior *b* in *s* depends on the personally anticipated emotion ☺(*s'*) in the *consequent situation s'* but also on the personally *anticipated effort* 🕓(*b*|*s*) needed for executing behavior *b* in situation *s*. We introduce the concept of *willingness to perform a behavior* as a function of (☺(*s'*) - 🕓(*b*|*s*)).

Simulation work that shows the validity of this theory is given in [20].

3.3 Basic Principles and Measuring Units

Comparing to physics, psychology is still just an empirical science. There are no basic principle equations like F = ma, and U = IR, and there is no system of units in for emotion and other concepts, in comparison to physics with units like kilogram for mass and Newton for force. An attempt for establishing basic principles might be indeed the equation (1). It relates behavior emotion and motivation, which means that at least one of those psychological concepts does not need a special measuring unit, it can be inferred from the other two. In [20] behavior is left to be such a concept, while emotion is measured in (… 2☹, ☹, 😐, ☺, 2☺, …) units and motivation in (…,-2♥, -♥, 0♥, ♥, 2♥, …) units. Any names can be given to those units and for now we call them Dze' and Daron respectively (names are from old macedonian mythology). So in engineered psychology we may talk in terms of motivation of 5 Darons and emotion of -3 Dze's. Let us note that Smiley is widely used name for the ☺-emoticon. Other names might also be proposed.

4 Conclusion

This work suggests that it is time review the undergraduate courses in AI, with direction toward Engineered Psychology. Maybe is time to replace the topic about search techniques which are now standard part of theory of algorithms, and go toward topics beyond intelligence and cognition, by including other features of a personality.

A result of Engineered Psychology approach is the first mathematical model that explicitly relates motivation, emotion, behavior, and personality. To the best of our knowledge such a model has not been proposed nether in psychology nor in Artificial Intelligence research. By building psychology in artificial creatures we believe we will better understand psychology in general. A step has been made toward the search for basic principles and measuring units in both engineered and human psychology.

References

1. Anderson, J., Bower, G.: Human Associative Memory. Erlbaum, Mahwah (1973)
2. Arbib, M.A.: Brains, Machines, and Mathematics. McGraw Hill, New York (1964)
3. Asada, M., Kitano, H., Noda, I., Veloso, M.: Robocup: Today and Tomorrow – What we Learned. Artificial Intelligence 110, 193–214 (1999)
4. Balkenius, C.: Natural Intelligence in Artificial Creatures. Lund University Press (1995)
5. Barto, A., Sutton, R.: Landmark Learning: An Illustration of Associative Search. Biological Cybernetics 42, 1–8 (1981)
6. Barto, A., Sutton, R., Anderson, C.: Neuronlike Elements that can Solve Difficult Learning Control Problems. IEEE Transactions on Systems, Man, and Cybernetics 13, 834–846 (1983)
7. Barto, A., Sutton, R., Watkins, C.: Learning with Sequential Decision Making. In: Gabriel, M., Moore, J. (eds.) Learning and Computational Neuroscience: Fundamentals of Adaptive Networks, pp. 539–602. MIT Press, Cambridge (1990)
8. Bellman, R.: Dynamic Programming. Princeton University Press, Princeton (1957)
9. Bledsoe, W.: Non-resolution Theorem Proving. Artificial Intelligence 9 (1977)

10. Boole, J.: An Investigation of the Laws of Thought. Walton and Maberly (1854)
11. Bouton, A.: A Game with a Complete Mathematical Theory. Annals of Mathematics Princeton 3, 35–39 (1901/1902)
12. Bozinovski, S.: A Self-learning System using Secondary Reinforcement. ANW Report, November 25, 1981. COINS. University of Massachusetts at Amherst (1981)
13. Bozinovski, S.: Inverted Pendulum Learning Control. ANW Memo. December 10, 1981. COINS. University of Massachusetts, Amherst (1981)
14. Bozinovski, S.: A Self-learning System Using Secondary Reinforcement. In: Trappl, R. (ed.) Cybernetics and Systems Research, pp. 397–402. North-Holland, Amsterdam (1982)
15. Bozinovski, S.: Crossbar Adaptive Array: The First Connectionist Network that Solved the Delayed Reinforcement Learning Problem. In: Dobnikar, A., Steele, N., Pearson, D., Alberts, R. (eds.) Artificial Neural Networks and Genetic Algorithms, pp. 320–325. Springer, Heidelberg (1999)
16. Bozinovski, S., Jaeger, H., Schoell, P.: Engineering Goalkeeper Behavior using Emotion Learning Method. In: Sablatnoeg, S., Enderle, S. (eds.) Proc. Workshop on RoboCup, KI 1999, pp. 48–56. Deutsche Jarhestagung fuer Kuenstliche Intelligenz, Bonn (1999)
17. Bozinovski, S., Bozinovska, L.: Self-learning Agents: A Connectionist Theory of Emotion Based on Crossbar Value Judgement. Journal of Cybernetics and Systems 32, 637–669 (2001)
18. Bozinovski, S., Jovancevski, G., Ackovska, N.: Distributed Interactive Robotics Classroom. Optoelectronics Information-Power Technologies 2(4), 5–9 (2002)
19. Bozinovski, S.: Motivation and Emotion in Anticipatory Behavior of Consequence Driven Systems. In: Butz, M., Sigaud, O., Gerard, P. (eds.) Proc. Workshop on Adaptive Behavior in Anticipatory Learning Systems, Edinburgh, pp. 100–119 (2002)
20. Bozinovski, S.: Anticipation Driven Artificial Personality. Building on Lewin and Loehlin. In: Butz, M.V., Sigaud, O., Gérard, P. (eds.) Anticipatory Behavior in Adaptive Learning Systems. LNCS, vol. 2684, pp. 133–150. Springer, Heidelberg (2003)
21. Bozinovski, S., Bozinovska, L.: Evolution of a Cognitive Architecture for Emotional Learning from a Modulon-structured Genome. Journal of Mind and Behavior 29(1-2), 195–216 (2008)
22. Braitenberg, V.: Vehicles: Experiments in Synthetic Psychology. MIT Press, Cambridge (1986)
23. Brooks, R.: A Robust Layered Control for a Mobile Robot. IEEE Journal of Robotics and Automation 2, 14–23 (1986)
24. Church, A.: The Calculi of Lambda Conversion. Annals of Mathematical Studies, vol. 6. Princeton Univesity Press, Princeton (1941)
25. Crevier, D.: AI: The Tumultuous History of the Search for Artificial Intelligence. Basic Books, New York (1993)
26. Duda, R., Hart, P.: Pattern Classification and Scene Analysis. John Wiley, Chichester (1973)
27. Fellous, J.-M., Arbib, M. (eds.): Who Needs Emotions? Oxford University Press, Oxford (2005)
28. Fikes, R., Hart, P., Nilsson, N.: Learning and Executing Generalized Robot Plans. Artificial Intelligence 3(4), 251–288 (1972)
29. Gadanho, S.: Reinforcement Learning in Autonomous Robots: An Empirical Investigation of the Role of Emotions. PhD Thesis. University of Edinburgh (1999)
30. Gadanho, S.: Learning Behavior-selection by Emotions and Cognition in a Multi-goal Robot Task. Journal of Machine Learning Research 4, 385–412 (2003)

31. Goldberg, D.: Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading (1989)
32. Green, C.: Theorem-proving by Resolution as a Basis for Question-answering System. In: Michie, D., Meltzer, B. (eds.) Machine Intelligence, vol. 4, pp. 183–205 (1969)
33. Grossberg, S.: Studies of Mind and Brain: Neural Processing of Learning, Perception, Development, and Motor Control. Reidel Press, Boston (1982)
34. Haugeland, J.: Artificial Intelligence: The very Idea. MIT Press, Cambridge (1985)
35. Hebb, D.: The Organization of Behavior. Wiley, Chichester (1949)
36. Hillis, D.: The Connection Machine. MIT Press, Cambridge (1985)
37. Holland, J.: Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor (1975)
38. Hopfield, J.: Neural Networks and Physical Systems with Emergent Collective Computation Abilities. In: Proc. National Academy of Science, vol. 79, pp. 2554–2558 (1984)
39. Jones, T.: Artificial Intelligence: A Systems Approach. Infinity Science Press (2008)
40. Klopf, H.: Goal-seeking Systems from Goal-seeking Components: Implementation to AI. The Cognition and Brain Theory Newsletter 3 (1979)
41. LeDoux, J.: Emotion and the Limbic System Concept. Concepts in Neuroscience 2, 169–199 (1991)
42. LeDoux, J.: The Emotional Brain. Simon and Schuster (1996)
43. Lewin, K.: Dynamic Theory of Personality. McGraw Hill, New York (1935)
44. Lewin, K.: Principles of Topological Psychology. McGraw Hill, New York (1936)
45. Lewin, K.: Field Theory in Social Science. Harper and Row (1951)
46. Liebert, R., Spiegler, M.: Personality. The Dorsey Press (1974)
47. Loehlin, J.: Computer Models of Personality. Random House (1968)
48. Luger, G.: Artificial Intelligence: Structures and Strategies for Complex Problem Solving. Addison Wesley, Reading (2009)
49. McCarthy, J.: Programs with Common Sense. In: Minsky (ed.) Semantic Information Processing, pp. 403–418 (1968)
50. McCulloch, W., Pitts, W.: A Logical Calculus of the Ideas Immanent in Nervous Activity. Bulletin of Mathematical Biophysics 5, 115–133 (1943)
51. McDermot, D.: Planning and Acting. Cognitive Science 2, 71–109 (1978)
52. Mendel, J.: Reinforcement Learning Control and Pattern Recognition Systems. In: Mendel, J., Fu, K. (eds.) Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications, pp. 287–318. Academic Press, London (1970)
53. Minsky, M.: Theory of Neural-analog Reinforcement Systems and its Application to the Brain-model System. Ph.D. Thesis, Princeton University (1954)
54. Minsky, M.: Steps toward Artificial Intelligence. In: Proceedings IRE, vol. 49, pp. 8–30 (1961)
55. Minsky, M., Papert, S.: Perceptrons: An Introduction to Computational Geometry. MIT Press, Cambridge (1969)
56. Minsky, M.: The Society of Mind. Simon and Schuster (1986)
57. Minsky, M.: The Emotion Machine. Simon and Schuster (2006)
58. Newel., A.: The Chess Machine: An Example of Dealing with Complex Tasks of Adaptation. In: Proc. Western Joint Computer Conference, pp. 101–108 (1955)
59. Newel, A., Shaw, J., Simon, H.: Empirical Explorations of the Logic Theory Machine: A Case Study of Heuristics. In: Proceedings Western Joint Computer Conference, vol. 15, pp. 218–239 (1957)
60. Nilsson, N.: Learning Machines. McGraw-Hill, New York (1965)

61. Petta, P., Trappl, R.: Personalities for Synthetic Actors: Current Issues and Some Perspectives. In: Petta, P., Trappl, R. (eds.) Creating Personalities for Synthetic Actors. LNCS, vol. 1195, pp. 209–218. Springer, Heidelberg (1997)
62. Piaget, J.: The Construction of Reality in the Child. Basic Books, New York (1954)
63. Pfeifer, R., Scheier, C.: Understanding Intelligence. The MIT Press, Cambridge (2000)
64. Rescorla, R., Wagner, A.: A Theory of Pavlovean Conditioning: Variations in the Effectiveness of Reinforcement and Non-reinforcement. In: Black, A., Procasy, W. (eds.) Classical Conditioning II. Current Research and Theory, pp. 64–99. Appleton-Century-Croffts (1972)
65. Rolls, E.: A Theory of Emotion and Consciousness, and its Application in Understanding the Neural Basis of Emotion. In: Gazzaniga, M. (ed.) The Cognitive Neuroscience, pp. 1091–1106. The MIT Press, Cambridge (1995)
66. Rosenblatt, F.: The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. Psychological Review 65, 386–408 (1958)
67. Rumelhart, D.: McClelland and the PDP Research Group. In: Parallel Distributed Processing. MIT Press, Cambridge (1986)
68. Russel, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice-Hall, Englewood Cliffs (2002)
69. Samuel, A.: Some Studies of Machine Learning Using the Game of Checkers. IBM Journal of Research and Development 3, 211–229 (1959)
70. Selfridge, O.: Pandemonium: A Paradigm of Learning. In: Symposium of the Mechanization of Thought Process, pp. 513–526. HMSO (1959)
71. Shannon, C.: Programming a Computer for Playing Chess. Philosophical Magazine 41 (1950)
72. Simon, H.: Motivational and Emotional Controls of Cognition. Psychological Review 74, 29–39 (1967)
73. Staugaard, A.: Robotics and AI: An Introduction to Applied machine Intelligence. Prentice-Hall, Englewood Cliffs (1987)
74. Sutton, R.: Learning to Predict by the Method of Temporal Difference. Machine Learning 3, 9–44 (1988)
75. Turing, A.: Computer Machinery and Intelligence. Mind 59, 433–460 (1950)
76. Warren, D., Pereira, L., Pereira, F.: PROLOG – the Language and its Implementation Compared with LISP. Proc. Symp. on AI and Programming Languages. SIGPLAN Notices 12(8) (1977)
77. Watkins, C.: Learning from Delayed Rewards. Ph.D. Thesis. King's College, Cambridge (1989)
78. Widrow, B., Hoff, M.: Adaptive Switching Circuits. IRE WESTCON Convention Record, pp. 96–104 (1960)
79. Widrow, B., Gupta, N., Maitra, S.: Punish/Reward: Learning with a Critic in Adaptive Threshold Systems. IEEE Trans. Systems, Man, and Cybernetics 3(5), 455–465 (1973)
80. Winograd, T.: Understanding Natural Language. Academic Press, London (1972)
81. Wirth, N.: Algorithms + Data Structures = Programs. Prentice Hall, Englewood Cliffs (1976)
82. Siegelmann, H.: Computational Models of Emotion. CS691I Seminar, University of Massachusetts at Amherst (2007), http://www.cs.umass.edu/~molsen/emotion