A Multi-level Model of Motivations and Valuations for Cognitive Agents

Samer Schaat, Klaus Doblhammer and Dietmar Dietrich
Institute of Computer Technology, Vienna University of Technology, Gusshausstrasse 27-29, Vienna, Austria

Keywords: Cognitive Architectures, Artificial General Intelligence, Cognitive Agents, Motivations, Valuations, Emotions.

Abstract: In developing cognitive agents using a functional model of the human mind as their decision unit, a model of motivations and valuations is needed as the basis for the agents' decision making. This enables agents to cope with their internal and external world while pursuing their own agenda. We show that a technical model based on the psychoanalytical drive concept and Damasio's neuro-biological findings is appropriate for human-inspired cognitive agents. In particular, after overcoming the hurdles of interdisciplinary work between hermeneutic and axiomatic approaches, a transformation of psychoanalytical and neuro-biological concepts into an consistent and deterministic model solves the problem of motivations and valuations in artificial cognitive agents. This multi-level model is presented, in which multiple principles and influences of valuation are used to incrementally generate and decide an agenda for the agent's behavior.

1 INTRODUCTION

When following a holistic and integrated approach to develop a functional model of the human mind as a decision unit for cognitive agents, basic questions are *(1) what is the source for the agent's agenda and (2) how can the agent cope with the external world while pursuing this agenda?* These are key questions since a cognitive agent is also an autonomous agent, i.e. a system situated within an environment which it acts on in pursuit of its own agenda (Franklin, 1997). Additionally, when using the human mind as an archetype, cognitive processes are dependent on motivations and valuations, e.g. decision making is based on emotions (Damasio, 2003). The first question above can thus be phrased as the problem of motivation, i.e. what should agents do and why should they do anything at all? The second question can be formulated as the problem of valuation in cognitive agents, i.e. the *determination of relevance of processed data* (e.g. objects, plans, actions) for the agent's motivations. Hence, these two questions can be regarded as interdependent, since valuation is a means of fulfilling motivations.

When considering a holistic model of the human mind as a decision unit in cognitive agents, embodiment must be considered when tackling the problem of motivation and valuation, in particular the interaction between the psyche and body. In this sense, autonomy and embodiment are foundations of higher cognitive functions such as planning, and motivations are based on the agent's bodily state on the one hand while serving as the basis for planning on the other hand. Hence, the model presented follows a *generative approach*, i.e. motivations are generated based on the agent's bodily needs and in turn are the source for generating goals for planning.

As a cognitive theory the ARS project has chosen the second topographical model of Sigmund Freud (Freud, 1915), since it is the only appropriate holistic and functional model of the human mind (Dietrich et al., 2009) (see 2) and is also backed up by neuroscientific data (Solms and Turnbull, 2002). However, it can only be used as a framework and general basis and is not sufficient to develop a human-inspired cognitive architecture. For instance, Damasio's concept (Damasio, 2003) is used to support the concretization of motivations and valuations in the ARS model.

2 RELATED WORK

Motivations and valuations are key methods used by humans to cope with and adapt to the world. Some robot architectures, such as (Parisi and Petrosino, 2010)(Konidaris and Barto, 2006), consider drives for motivational purposes. Nevertheless, a bionic and functional approach to motivations and valuations are often neglected in current cognitive architectures. Established cognitive architectures, such as ACT-R (An-

Schaat S., Doblhammer K. and Dietrich D..
A Multi-level Model of Motivations and Valuations for Cognitive Agents.
DOI: 10.5220/0004916502550260
In *Proceedings of the 6th International Conference on Agents and Artificial Intelligence* (ICAART-2014), pages 255-260
ISBN: 978-989-758-016-1

derson et al., 1997), focus on 'higher' cognitive functions without considering the basis for these higher functions, i.e. generation and valuation of goals. Only a few cognitive architectures, such as MicroPsi (Bach, 2011) and CLARION (Sun, 2007), also focus on motivational aspects. However, these approaches do not consider a generative and embodied approach. For instance, MicroPsi distinguishes physiological, cognitive and social needs, which are all hard-wired. In the embodied approach of the ARS model all needs are grounded in physiological needs. Based on social rules and the agent's memories of how to satisfy physiological needs in the short and long term, they are transformed into complex motivations and goals. CLARION in fact considers the derivation of 'secondary' drives from 'primary' drives, but uses the drive-concept more in terms of a behavioristic approach, since it is based on Hull's concept of drives (Hull, 1951). In contrast to the concept of drives in ARS, the drives' strength in CLARION is determined by an internal deficit and an external stimulus (e.g. food in case of hunger). In this regard, the ARS multilevel approach enables a more flexible motivational system: generally speaking, a first level of motivation and valuation considers only how to satisfy its motivation best according to the agent's memories and without consideration of the external world. After further levels of motivation and valuation, it is only in the ARS secondary process that the reality imposed by the external world is considered.

Emotions are a central aspect of motivations and valuation; they can be seen in general terms as a representation of an agent's internal state (emotions as embodied information of valuation and importance) (S. C. Marsella and Petta, 2010). Recently, various stand-alone computational models of emotion have been developed, i.e. they are usually not integrated into a full-fledged cognitive architecture. Due to its focus on the connection between emotion and cognition (S. C. Marsella and Petta, 2010), cognitive appraisal theory is currently the dominant theory for computer models of emotions. In this theory, emotions emerge from the appraisal of external events and situations under the consideration of the agent's beliefs, desires and intentions. The result of such appraisal is the triggering of cognitive responses, in particular coping strategies (e.g. planning, procrastination) (S. C. Marsella and Petta, 2010). Appraisal theories focus on determining a sufficient set of appraisal criteria to explain the elicitation and differentiation of emotions. A widely used model of cognitive appraisal theory is the OCC model (Ortony et al., 1990) (e.g. used in EMA (Marsella and Gratch, 2009)), with appraisal criteria such as unexpectedness, level of appeal and desirability. In appraisal theory, emotions are only elicited by evaluations of external events and hence only considered for interaction purposes. This is a major difference to the ARS model, where the generation of emotions is influenced by external events on the one hand and based on the agent's drives on the other hand. In particular, this means the agent may be in an emotional state even without consideration of the external world. Another category of emotion theories follows a dimensional approach and models emotions not as discrete entities but as points in a continuous dimensional space (S. C. Marsella and Petta, 2010). A typical example for this is the three-dimensional PAD model (Mehrabian and Russell, 1974), with pleasure (a measure of valence), arousal (indicating the level of affective activation) and dominance (a measure of power or control) as dimensions.

3 ARS APPROACH

The leitmotif of the ARS approach (Dietrich et al., 2009) is to model the functions behind the desired capabilities that generate behavior instead of simply describing behavior. This complies with the generative and broad approach of Artificial General Intelligence (AGI).

In the ARS project, the human cognitive architecture (i.e. the brain) is considered as an information processing system that stores, manipulates and transfers information. Following the standard approach in computer technology, it is described in a top-down design process using a layered model, starting with three functional layers. The first layer, the neurons, can be described as hardware under consideration of the laws of physics. The next layer is called the neurosymbolic layer, which handles the symbolization of the neural layer. The third layer represents the psychic layer, which is described in functional terms on an algorithmic level. Following a monistic view, psyche and brain are of course the same, with only different models being used in their respective descriptions. Since only a functional description is relevant for artificial systems (but not how these functions are implemented), the ARS project focuses on the description of the psychic apparatus. As mentioned in the introduction, the second topographical model of Freud (Freud, 1915) was chosen as a general framework, which uses the abstract functions Id, Ego and Super-Ego to describe the human psychic apparatus. The Id represents drives, which are in effect bodily demands coming from internal sensors, the Super-ego represents internalized moral demands and the Ego me-

diates between the Id and the Super-Ego under consideration of the external environment. These three abstract functions are concretized incrementally until their description can be used for a functional model that can be implemented in a computer simulation. The most concrete level of the ARS functional model is shown in Fig. 1 using a track view in which function modules are summarized into tracks.

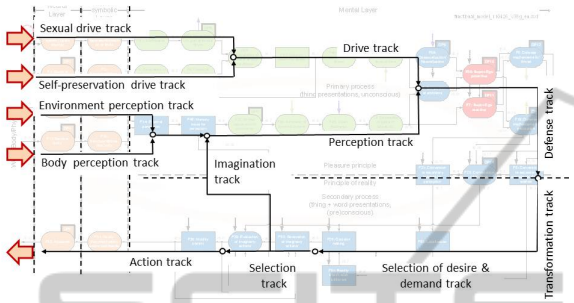

Figure 1: ARS functional model - track view.

4 PSYCHIC INTENSITY

Following an interdisciplinary approach, the general psychoanalytical framework is described based on terminology used in the ARS project, and then concretized by a technician in the subsequent sections to a coherent and deterministic *multi-level model of motivations and valuations*.

4.1 Economic Point of View

The economic point of view describes functionalities that focus on circulation, distribution and transformation of quantities of excitation. Freud called these quantities 'psychic energy' , which became a fundamental term in psychoanalysis (Freud, 1940, p. 158) to designate a representation of a bodily drive-tension, which is created by differences in homeostatic levels. The quantification of 'psychic energy' in the psychic apparatus is called quota of affect. With this value, psychic content can be used for cathexis, i.e. 'psychic energy' is assigned to certain psychic content via its quota of affect. This represents the valuation of psychic content. Following the so-called pleasure principle and the principle of reality, cathexes are essential for a psychic economy since they provide the direction for decisions, actions and plans that an agent executes or wants to execute. The pleasure principle says that the entirety of psychic activity is aimed at avoiding unpleasure and generating pleasure, while the principle of reality is a modification of it as aspects of reality are involved in this design of activity. Pleasure arises from the discharge of 'psychic energy'; unpleasure is equivalent to the sum of all quotas of affect within the psychic apparatus. Since emotions are experienced as an aspect of pleasure and unpleasure, the term 'psychic energy' is also a basic term for the generation of emotion and feeling, which additionally influence decision making and bodily states.

An adult psyche does not assign all 'psychic energy' to psychic content. Instead, a certain amount of it is available for particular psychic functions, the so-called Ego-functions. This 'neutralized drive energy' can be regarded as an engine for these functions, which deal with thinking, focus of attention, planning, adaption to reality and other concepts. This implies discharge and produces pleasure as well.

4.2 Psychic Intensity in the ARS Model

For an embodied approach, the connection between body and psyche, i.e. between the different description languages of physics and information theory (see 2), must therefore be considered. This is done by use of the neurosymbolic layer. For the questions of motivation and valuation, this means that we must consider the bodily source of motivation and valuation (i.e. the connection between body and psyche), and in particular the symbolization of the bodily source and its psychic representation. The use of the term 'psychic energy' shows that these aspects are not considered in classical psychoanalysis. Nevertheless, in a concrete technical model they must be concretized and structured using the appropriate level and language of description.

But the term 'psychic energy' is misleading and not compliant with the ARS approach, since it uses the description language of the physical world which is only appropriate for the first layer (see above). But it is misleading even as a metaphor for describing the concept's functionality, e.g. the abstract and functional concept of money is better suited as a metaphor. Since the essence of the concept is valuation, i.e. the quantification of the importance of processed content, the term *psychic intensity* was chosen as an appropriate representation of the underlying functionality. As with money, the physics of psychic intensity is irrelevant for its description; only a functional description can cover the essence of the concept. In particular, psychic intensity serves as an *information value*. In the algorithmic level of the ARS model, i.e. the psychic layer, the concept of psychic intensity is used as the *link between bodily needs (i.e. homeostatic needs) and motivations*, in particular as a *valuation system* for determining the relevance of memories and perception and as a foundation for decision making.

Next, beginning with the concept of psychic intensity we describe the coherent multi-level model of motivations and valuations.

5 A MULTI-LEVEL MODEL

The foundation for motivations and valuations in the ARS model is provided by the concept of *drives*. This is true not only for the ARS primary process, their principal domain, but also for the secondary process, which has to cope with the demands generated by the drives under consideration of social rules, the agent's current plans and the external world. Based on the drives, an extended representation of motivations and valuations is given by *emotions*, which additionally consider the agent's perception. Hence motivations and valuations are represented in the ARS agent using a *multilevel model*, which breaks the complexity of the motivation system up into several levels. This allows an *incremental generation* of motivations and valuations, where motivations and valuations are extended and/or adapted using different principles (rules of the ARS primary and secondary process) and influences (memories, perception, plans). Psychic intensity is thus extended to - and concretized on - multiple levels, which nonetheless all serve the same purpose: motivation and valuation.

Such a multi-level model of motivations and valuations based on the agent's body and memories complies with Damasio's model (Damasio, 2003), where the homeostatic mechanisms for the regulation and control of the organism's processes are considered to be multiple but coherent levels which (in terms of evolution theory) evolved as adaptations for coping with the world.

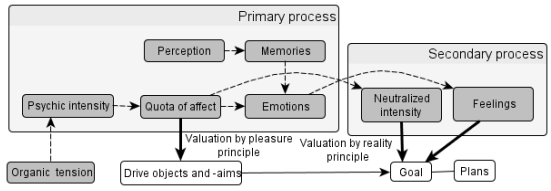

Figure 2: Different valuation-principles and influences (memories and perception) are considered.

Each *homeostatic need* is represented by an *organic tension value* (e.g. the volume of stomach content). As described above, this value is represented as *psychic intensity* in the psychic layer. A range of 0-1 is used for the quantification of psychic intensity. A personality parameter determines the mapping from drive tension to psychic intensity, i.e. in which intensity the agent's psyche represents the bodily need. Next, in the process of generating a drive representation (see Ch. 5), psychic intensity is represented as a *quota of affect* which is used to valuate memories, e.g. objects and actions that have previously satisfied the drive, and hence represents the first level of valuation. *This first level of valuation lead to the first level of motivations, drive representations*. This valuation is used under the consideration of the *pleasure principle*, in particular to determine objects that have provided the best satisfaction of the drive (i.e. the bodily need) according to the agent's memories without the consideration of reachability. The valuation of content may be changed by *defense mechanisms*, which take into account internalized rules such as social and moral rules (e.g. a rule that forbids eating in a specific situation may result in changing the valuation of food objects). Next, through the function of 'neutralization' , an amount of the drive's quota of affect, determined by a personality parameter, is reduced and used as *neutralized intensity* for a possible extension of already valuated content, using the *reality principle* in the ARS secondary process as a valuation criterion. This is primarily the consideration of the compatibility of the valuated objects and actions with the current reality provided by the external world and with the agent's current plans. In general, logic and time are considered in this valuation process. This additional valuation transforms a drive representation to a valuated *goal*, where the goal's object, action and plan are valuated. The valuation of goals can be extended by *emotions*, which are another level of valuation and are generated based on the remaining quota of affect (after neutralization) and memorized emotions that are activated by perception. An overview of this incremental multi-level valuation is provided in Fig. 2.

In the next section, the key elements of motivations and valuations, drives and emotions will be outlined.

5.1 Drive Representations: Motivations in the ARS Model

The representation of a drive, i.e. a somatic organ tension, in the psyche is provided by a *drive representation*, which consists of a *drive source* (the organic origin), a *drive aim* (an activity that decreases the organic tension) and a *drive object* (with which the drive can achieve its aim). The action given by the drive aim is always a strategic action, which is implemented as a tactical plan. In particular, how this action is fulfilled (the tactic) is the domain of reasoning in the ARS secondary process. *Self-preservation* and *sexual drives* are distinguished. The drive source for the former is the respective organ (e.g. stom-

ach), the drive source for the latter is called *libido*, a value that represents hormonal sources. Depending on erogenous zones on four different body parts, four so-called partial sexual drives are distinguished: anal, phallic, oral and genital. Only the last of these drives reproduction, while the others may be regarded as *intrinsic motivations* , i.e. pleasure gain from satisfying sexual drives is independent from the satisfaction of bodily needs. To complete the representation of a drive, the agent has to choose an object and action with which the bodily need, represented by the drive, can be satisfied. Psychoanalytic theory describes this process as cathexis, which corresponds to the valuation of objects and actions. Drive objects and aims that were best valuated in the past (i.e. brought the amounts of highest satisfaction) are valuated best. In the ARS agent, only valuated data can be considered in the function modules. The valuation done by cathexis is based on (1) the memorized valuation and (2) the quota of affect available. Memorized valuation is represented by an association between a memorized drive representation and an object. The valuation process consists of the following steps: (1) Activation of memorized drive representations with the search criteria of drive source and drive component using an associative search algorithm, (2) Assignment of the drive representation's provided quota of affect to the data associated to the memorized drive representations, (3) Consideration of accumulation by multiple valuations.

5.2 Emotions and Feelings

Based on the agent's drives and influenced by memories activated by perception, emotions are an additional representation of the agent's internal state under consideration of the external world - in particular, the agent's experience with similar situations. The transformation of emotions to feelings, based on conscious perception thereof (Damasio, 2003), forms a new valuation-level, which may change the valuation of goals.

Currently basic emotions (anger, mourning, anxiety, joy, saturation, and elation) are considered in the ARS agent, which are generated based on the agent's actual drives and perception, in particular the memorized emotions associated with perception. They are determined by four factors, namely *unpleasure* (the sum of the quotas of affect), *pleasure* (the value generated after reducing psychic intensity), the sum of all *aggressive quotas of affect* and the sum of all *libidinous quotas of affect*. Two sources for these factors are considered: (1) drive representations and (2) the factors of memorized emotions that are activated by perception. Dependent on the dominance of a subset of these four factors, which also determine the intensity of the emotions, different emotions are generated.

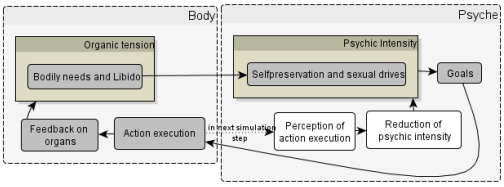

Figure 3: Feedback on body or psyche.

5.3 Motivation Control Loop

In terms of a control loop, the *feedback* on the source of motivations must be considered. In the embodied approach of the ARS model, this can occur in two ways: the first possibility is the implicit reduction of psychic intensity (the source of motivations) through reduction of the respective organ tension via a specific action. For instance, eating influences the state of the stomach, which in turn changes drive tension and the psychic representation thereof, psychic intensity. In case of actions that do not produce direct feedback on the bodily source (e.g. beating), a second possibility of feedback is provided by the perception of the currently executed action. Since this feedback occurs on the psychic layer, it affects psychic intensity directly.

6 SIMULATION

To find the functions behind any given behavior, a *use-case driven approach* is employed: the behavior to be generated by the functions of the ARS model and their corresponding inner processes are described together with psychoanalysts and structured as use-cases. Thus requirements are specified that are then fulfilled by functions and data which generate the described behavior in an agent-based simulation. For simulation the multi-agent simulation framework MASON is used. After presenting a simulation of an agent-object interaction in, here an overview of preliminary results of simulating a social interaction is given next. The use-case defines a situation with two agents (a green one and a red one) and a Wiener Schnitzel as food source.

The green agent's drive state, especially the domination of aggressive drives (see Fig. 4), the memorized valuation of the Schnitzel and the red agent, and the green agent's personality parameters result in the green agent beating the red agent. In particular, due to the green agent's personality parameters, the quota of affect of aggressive drive representations is

higher than that of libidinous drive representations, and the quota of affect is only reduced minimally by increasing neutralized intensity. Thus valuation by the green agent occurs primarily via cathexis following the pleasure principle, and due to the green agent's memories, the drive object 'red agent' and -aim 'beat' (which are transformed into the respective goal), are valuated the highest.

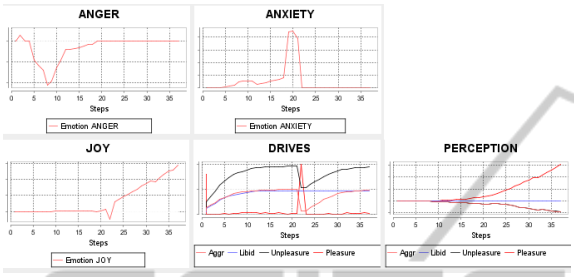

Figure 4: Beating in simulation step 21 reduces aggressive drives, which increases pleasure. Together with memorized valuations activated by perception this leads to the generation of the shown emotional state.

The emotions shown in Fig. 4 represent the drive representations and memories activated within the green agent by perceiving the Schnitzel and the red agent. Since the red agent looks similar to the red agent's memorized brother - who, in combination with a food source, is associated with anger - and aggressive drive representations dominate, anger also dominates; the low libidinous drive representations, together with the memorized joy activated by perceiving the Schnitzel, lead to a low generation of joy compared with beating the red agent. Beating the red agent leads to reduction of the corresponding psychic intensity and hence to generation of pleasure within the green agent, which in turn increases joy and reduces anxiety.

7 CONCLUSIONS

Valuation of objects and actions using the concept of psychic intensity lead to the generation of drive representations from bodily needs, which together with emotions represent the agent's motivations and ground its autonomy in its body. This shows that the development of a consistent and deterministic model of psychoanalytic and neuro-biological concepts for the simulation in agent-based systems is possible. In particular, the connections between body, drives, emotions and goals and the feedback on the body are considered in a deterministic and cyclic fashion. The presented multi-level model, in which different valuation criteria are considered, shows how human-inspired complex valuation enables an incremental determination of the relevance of objects, actions, goals, and plans to reach them.

REFERENCES

Anderson, J. R., Matessa, M., and Lebiere, C. (1997). Act-r: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4):439–462.

Bach, J. (2011). A motivational system for cognitive ai. In J. Schmidhuber, K. R. T. and Looks, M., editors, *Proceedings of the 4th international conference on Artificial general intelligence*. Springer, Berlin.

Damasio, A. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Harvest Books.

Dietrich, D., Fodor, G., Zucker, G., and Bruckner, D. (2009). *Simulating the Mind - A Technical Neuropsychoanalytical Approach*. Springer, Wien.

Franklin, S. (1997). Autonomous agents as embodied ai. *Cybernetics and Systems*, 28(6).

Freud, S. (1915). *The Unconscious*, volume XIV (1914-1916) of *On the History of the Psycho-Analytic Movement, Papers on Metapsychology and Other Works*. Vintage.

Freud, S. (1940). An outline of psycho-analysis. *International Journal of Psycho-Analysis*, 21:27–84.

Hull, C. (1951). *Essentials of Behavior*. Yale University Press, New Haven.

Konidaris, G. and Barto, A. (2006). An adaptive robot motivational system. In *From Animals to Animats 9*, pages pp 346–356. Springer, Berlin.

Marsella, S. C. and Gratch, J. (2009). Ema: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90.

Mehrabian, A. and Russell, J. (1974). *An approach to environmental psychology*. MIT Press, Cambridge.

Ortony, A., Clore, G. L., and Collins, A. (1990). *The Cognitive Structure of Emotions*. Cambridge University Press, New York.

Parisi, D. and Petrosino, G. (2010). Robots that have emotions. *Adaptive Behavior*, 18(6):453–469.

S. C. Marsella, J. G. and Petta, P. (2010). Computational models of emotion. In K.R. Scherer, T. B. and Roesch, E., editors, *Blueprint for Affective Computing*. Oxford University Press, New York.

Solms, M. and Turnbull, O. (2002). *The Brain and the Inner World: An Introduction to the Neuroscience of Subjective Experience*. Karnac/Other Press, Cathy Miller Foreign Rights Agency, London, England.

Sun, R. (2007). The motivational and metacognitive control in clarion. In Gray, W. D., editor, *Integrated Models of Cognitive Systems*, pages 63–75. Oxford University Press, New York.