# An Interdisciplinary Approach for a Holistic and Embodied Emotion Model in Humanoid Agents

Samer Schaat<sup>1</sup>, Matthias Huber<sup>2</sup>, Klaus Doblhammer<sup>1</sup>, Dietmar Dietrich<sup>1</sup>

<sup>1</sup>Institute for Computer Technology, Vienna University of Technology, A - 1040 Vienna {schaat, doblhammer, dietrich}@ict.tuwien.ac.at <sup>2</sup>Department of Education, University of Vienna, A - 1010 Vienna matthias.huber@univie.ac.at

Abstract. Computational models of the human decision making process have already enabled several insights and applications. Nevertheless, such models have only recently begun to consider that simply modelling rational decision making was insufficient to represent human behavior. Recent efforts have started to consider this factor by utilizing computational models of emotions. One of the most significant challenges to be faced here is the interdisciplinary cooperation required in order to develop a holistic and integrated model, which reflects aspects of embodiment and is integrated in a holistic cognitive architecture. In this paper we will analyse how models from relevant disciplines can support us in outlining an overview model which considers the mentioned criteria.

**Keywords:** Emotions · Humanoid Agents · Evaluation Models · Decision Making · Simulation · Artificial General Intelligence · Neuroscience · Neuropsychoanalysis.

#### 1 Introduction

If one considers humans as being the archetypes of intelligent information processing, the fundamental mechanisms of human decision making must be considered. This is especially relevant for a holistic model of decision making. Over the last years the focus in this area was on rational decision making, with little consideration for the fact that this is actually based upon *unconscious processes*. In particular the consideration of affective processes is needed in a computational representation of human capabilities that are used to cope with every-day problems, such as "intuition", "gut feeling" etc. A further tendency in humanoid agents is the reduction of decision making are generally valid, the definite rules followed by humans cannot be generalized, and therefore must be generated according to a *subjective approach*, based on memory and also in an agent specific manner. Besides, conventional decision making models are mostly abstracted from the body. Following an *embodied approach*, human cognition and particularly decision making is however based upon the interaction of the brain with the rest of the body.

These topics can only be tackled with an interdisciplinary approach, where computer science is not used as a tool, but its methodologies are applied on the content of other disciplines. Such approach not only enables to develop computer models for technical applications, but also serves precision, concretization and sharpening of concepts from different disciplines. In order to develop a deterministic model of an agent decision making process according to human-based principles, we will leverage insights from different disciplines, considering embodiment, subjective approaches and unconscious processes. In particular we consider neuroscience, psychoanalysis, neuropsychoanalysis, and current computation models of emotions. Although the unconscious is a key concept in psychoanalysis, as far as we know, it is not analyzed yet for computational models of affective processes, and neuroscientific theories are not reflected in a holistic and embodied computational model of human decision making yet. In order to further construct a model transformation upon this and to integrate neuroscientific and (neuro)psychoanalytic models into a technical model, it is necessary that they are described *consistently and holistically*. Even though this paper shall focus on a model of the basic mechanisms of decision making, the holistic perspective may not be neglected. This represents the integration of cognition and emotion, which are thereby not modelled separately. Upon highlighting how each discipline (by an expert of neuroscientific emotion theories, a psychoanalyst and a computer scientist) perceives the fundamental mechanisms of subjective decision making in terms of the criteria described, we shall sketch an evaluation model for subjective decision making in humanoid agents which integrates the insights of the various disciplines.

#### 2 Emotions in Neuroscience

Currently several epistemologically divergent neuroscientific research movements are endeavoring to describe human emotionality in a holistic context. In particular the work of Jaak Panksepp is of note, specifically his theory of mutually linked (neuronal interacting) basic emotion systems (seeking, fear, panic, rage) [1]. Also of note is the work of Gerald M. Edelman and his theory of Neuronal Darwinism in which emotions, in the context of values and value systems, are introduced in a holistic theory of consciousness [2], a precursor to a modern integrative comprehension of emotions. The research of Joseph LeDoux concerning the "Amygdala Fear System" [3] and also the work of James McGough on the significance of emotions in learning and memory [4] are also worthy of mention. The Portuguese Antonio Rosa Damasio is recognized as one of the most significant and influential contemporary neuroscientists. His "Theory of Emotion" is highly regarded far beyond neurobiology circles, as his comprehension of emotions and feelings possesses great explanatory potential, confirmed through diverse studies [5], [6], and a plethora of neurophysiological data [7], [3], [8]: On the one hand Damasio is able to justify the holistic claim as his comprehensive body of work includes, along with the Theory of Emotion, a "Theory of Self", a "Theory of Consciousness", a "Theory of Mind" along with a theory governing the automatic regulation of life. On the other hand, Damasio also took on the challenge of creating a model for the dualistic juxtaposition of body and mind, based on the inevitable interdependence of reason and feelings, thereby transporting an anthropological comprehension which satisfies both the psychoanalytical developmental theories and phenomenological traditions.

The starting point of Antonio Damasio's deliberations is the organism, as a holistic, open system, that progressively interacts with the environment and is organized according to the operating principle of homeostasis (automatic life regulation), whereby emotions form the ultimate level of these regulative (permanently active) mechanisms. This also clarifies why emotion and feeling take on a biological (and also evolutionary) function as necessary survival regulators for the protection of an organism. Damasio differentiates in this context between three processing stages: (1) an emotional status, which can be unconsciously triggered and executed, (2) a feeling state, which can be unconsciously represented along with (3) a consciously generated feeling state, in which the organism knows that it possesses emotions but also feelings [6, p. 51]. Furthermore three types of emotions are differentiated: (a) the primary emotions (fear, anger, sorrow, happiness, disgust, surprise) which are congenital (preorganized in terms of classical "Jamesist" feelings) and can be thought of as being genetically determined, universal and culturally dependent basic emotions. (b) The secondary or social emotions which in contrast develop over the course of ontogeny and emerge as soon as systematic connections of objects and/or situations with primary emotions are formed. Social or secondary emotions (e.g. compassion, embarrassment, shame, guilt, pride, jealously, awe, envy etc.) are thereby acquired and respectively triggered through mental registration with respect to the cognitive processing of situations and/or objects. (c) Background emotions are seen by Damasio as being the consequence of certain combinations of simple regulative (homeostatic) reactions (e.g. desire) [9, p. 56]. They are ever-present but are seldom consciously perceived and may be interpreted as being an expression of well-being or discomfort. Emotions thereby fulfill a double biological function: as already mentioned they must continuously regulate the inner status of the organism. Above and beyond that they must also trigger a specific reaction to a particular stimulus or situation. Two mechanisms are available for this: emotions are formed either when our sensory organs process certain objects and/or situations (the body-loop) or when the organism retrieves certain objects and/or situations from memory and represents them as imaginary images in the thought process (the as-if-loop). It has already been stated that emotions and feelings are temporally and structurally predetermined: the key content of a feeling is the illustration of a particular bodily state. A feeling is thereby a projection of the body under certain conditions and the feeling of an emotion is the projection of the body while under the influence of a particular emotional process. Additionally, along with the bodily-related projections, in certain situations specific projections of the thought process are also relevant. Therefore a (conscious) feeling consists of the perception of a certain bodily state along with the perception of a certain associated mental state (the cognition of the emotion). As before, Damasio also differentiates here between three types of feelings: (a) feelings of basic universal emotions (in terms of the primary emotions), (b) feelings of differentiated universal emotions (as a connection of cognitive states with emotional bodily states; depending on experience, in terms of the

social emotions), along with (c) background sensations (in terms of background emotions, although not formed by emotions in the strictest sense) [5, p. 208].

In the context of this "Theory of Emotion", an elementary principle is clearly evident in Damasio's work: emotion, feelings and consciousness are continually dependent on the representations of the organism and their common entity is and remains the body. Human thoughts and actions are therefore dependent on the emotional constitution and respectively to certain changes in bodily state. The purpose of thought and the prerequisite for action is however always a decision, whereby the essence of a decision lies in choosing a certain response (e.g. a course of action) [5, p. 227].

If one considers the decision making process on the basis of an undesirable development, one thereby creates an undesirable/negative outcome, that is connected to the associated response and is consciously perceived, even when short-lived, as an uncomfortable/negative feeling. As a feeling (from emotion) affects the body, Damasio choose the term somatic (soma = body), and as the feeling identifies, or marks a projection, he also chose the term marker. A somatic marker is thereby understood as being the perception of an automatic reaction of the body to a certain projected image (a situation or an event) respectively, as a bodily signal marking a particular scenario as being either good or bad. [5, p. 238]. Accordingly a positive somatic marker functions as a start signal and a negative somatic marker functions as an inhibitor. Somatic markers are formed throughout the course of upbringing and socialization through connecting certain classes of stimuli with certain classes of somatic statuses. Therefore they touch on the process of the secondary emotions. The adaptive function of the somatic marker (as an assistant with anticipatory skills) orientates itself towards congenital, regulatory dispositions (internal preference system) which ensure the survival of the organism and take care of avoiding pain and seeking or increasing desire.

Looking back at the functional mechanisms of emotion, one can differentiate four forms of the decision making process: (A) in the context of the body-loop, the body is actually (from the prefrontal cortex and the amygdala) prompted to take on a certain state profile, the result of which (via the somatosensory cortex) is considered with attention and perceived. (B) In the context of the as-if-loop the somatosensory cortex functions (as instructed by the prefrontal cortex and the amygdala) as if the signals were received from the body. Therefore the body is taken out of the loop, nevertheless the as-if activity patterns influence the decision making process, as it suggests that real bodily states are symbolically processed. (C) Additionally somatic markers (e.g. feelings) can represent very concrete components or triggers for decisions, regardless of whether they follow a real or representative route. (D) Very often decisions are made, where it appears that no feeling at all was involved. Therefore it is not - and that is key here - that it doesn't come to an evocation of a bodily state or that of a representative surrogate, but rather just that of the bodily state with which the signal function is activated, it is simply just not considered and therefore not consciously perceived [10, p. 84]. By this means somatic markers operate permanently outside of consciousness and persistently influence conscious thought and decisions. Therefore one differentiates between somatic markers with respect to the influence of emotion and feeling on the decision making processes based on their neural route, (A) the real

body-loop versus (B) the representative as-if-loop and on the basis of their influence, (C) manifest versus (D) covert.

### **3** Evaluation Models in Psychoanalysis

There are multiple aspects to the description of decision making in the "psychic apparatus" in classical psychoanalytic theory [11, 14]: On the one hand the body delivers via homeostatic differences drive tensions, so-called "quota of affects" represented in the psyche, which can consequently cathect<sup>1</sup> psychic contents. This allocation of the "quota of affects" to psychic contents already activates content in the unconscious whereby these become accessible for mental processing. The level of cathexis is a measure of the grade of the activation, representing an evaluation, which ultimately is a key factor in determining, if content shall be processed and ultimately become consciously perceivable and actionable. The psychic contents on the other hand, come from *memory traces* which are associated via perceptional data and drive representatives and by this means serve as a basis for cathexis. A cathected association complex is called thing presentation. This is, during the transition from (unconscious) primary process to secondary process - upon going through a conflict regulating defense linked with a so-called word presentation, which means, that from now on a psychic content can be treated within general (formal, verbal) logic. This was not yet possible in the primary process, as this is governed by a pre-logical order of associations. Cathexes, which in the course of the mental processing of the primary process have been displaced many times, remain intact throughout these transitions. In the secondary process, the topical description of which encompasses the psychoanalytical preconscious and conscious areas, are now fixed to the "quota of affects" of certain contents and contribute, along with the logical links via word presentations to the evaluation of the association complexes.

The overall evaluation of action chains generated in this manner is regulated on the one hand – in the primary process – by the so called "*pleasure principle*" and in the secondary process by the "*principle of reality*". The "pleasure principle" states that the overall goal of all mental activity is to avoid unpleasure and to aim for pleasure (as in [12, p.321]), the "principle of reality" is a variation upon this, namely the moments in which the outside world becomes included into these activity designs (as in: [12, p.378]). Pleasure is created if psychic energy is discharged, unpleasure is equivalent to the "quotas of affect" present within the apparatus. Both of these principles are ultimately relevant for decision making and choosing a course of action, so much so, that in total a maximum of expected pleasure less the necessary unpleasure in order to achieve it is always sought. Primary and secondary process oriented thinking and evaluation mechanisms complement each other towards taking action. Thinking is essentially an experimental kind of acting. [13, p.220]. If action is undertaken, the "quota of affects" is discharged and alongside the physical impulses, consciously perceptible affects and impulses of feelings in particular shades are formed. Uncon-

<sup>&</sup>lt;sup>1</sup>In psychoanalysis this also known as the economic aspect of psychological operation. Cathexis (a psychoanalytic term) is the allocation of quota of affects to psychic contents.

scious affects, feelings and sensations are of no relevance in psychoanalytic theory, they are, - in contrast to psychic contents –virtual qualities, which with respect to the occupation conditions of the "quota of affects" can be constructed retrospectively [14, p.176].

#### 4 Emotions in Neuropsychoanalysis

Neuropsychoanalysis seeks to forge a connection between psychoanalytical models and related neuroscientific findings [7, 15]. It seeks to assume a neuroscientific perspective of every mental function and thereby wishes to reassess Freud's description of the functional organization [15, p. 830]. The results of these comparisons may remain patchy in accordance with the method applied and usually only the most significant theses and statements of these disciplines are studied.

Considering evaluation models for decision making, neuropsychoanalysis holds that the conscious decisions for actions, in reality follow their unconscious initiation [cf. 15, p. 384]. Feelings of pleasure and unpleasure with respect to an object or a situation represent the most elementary evaluation of a consciousness, according to Solms, Panksepp and Damasio [15, p. 836]. Likewise the basic units of these evaluations, of the driving forces, can be illustrated neurologically and yield that: An amazingly large consensus emerges between Panksepp's SEEKING-System and Freud's Libido-System and the highest priority of a regulating function of consciousness is to generate feelings of pleasure and unpleasure, which in turn are then associated with the objects which are best suited to their generation [cf. 15, p. 848].

#### 5 Emotions in Agent-based Systems

In recent years several computational models of emotions have been developed and integrated in the decision making process of artificial agents. These models differ in various aspects. Generally they differ in the components which are considered as being intrinsic to emotions (e.g. bodily processes, behavioral tendencies), in their relationship to cognitive processes and in their representation [16]. The most important difference lies in the supporting theory, which in most cases originates from psychology. This in turn influences the distinction if emotions are generated dynamically (emergent emotions) or are designed explicitly (discrete emotions). These aspects of distinction are mirrored in the division of "appraisal, dimensional and anatomical" computational models of emotion [16], whereby the former is the most widespread due to its aptness in linking emotion with cognitive processes.

In appraisal models (e.g. OCC Modell [17]), emotions are formed through the evaluation of external events regarding the agent's beliefs, desires and intentions, whereby coping strategies (e.g. planning, delaying) are triggered. A computational model of this, which offers the chance to adjust the appraisal process, is FAtiMA Modular [18]. The agent architecture consists of an extensible core architecture which offers the framework for various implementations of appraisal models, enabling easier comparison. The appraisal process is split into the appraisal derivation and the affect

derivation. The former evaluates the relevance of an event and creates appraisal variables (e.g. unexpectedness, appealingness and desirability). The latter builds upon these variables by creating the associated emotions, determined by a specific appraisal theory. According to a set of rules, emotions then influence the choice of action in either a reactive or deliberate form.

In dimensional models, emotions are located in a dimensional space instead of being formulated as discrete entities. A typical model is the PAD model [19], with the dimensions pleasure, arousal and dominance. Some computational models such as ALMA [20] and WASABI [21] utilize both, appraisal models to model appraisal processes, and dimensional models to model mood processes. Other models such as MicroPsi [22, p.143-155] describe emotions implicitly as regions of a multi-dimensional space, defined by the parameters which determine the behavior of the agent. These are: arousal, resolution level, dominance of the leading motive, the level of background checks (the rate of the securing behavior), the level of goal-directed behavior, and valence. Thus, explicit emotions do not exist for agents but rather emotions are first attributed to an agent upon (self) perception.

In conclusion, it can be stated that none of the models mentioned offers an embodied model that holistically considers the various aspects of emotion, or is integrated in a holistic cognitive architecture.

## 6 A Holistic and Embodied Emotion Model for Evaluation in Decision Making

Building on the findings of the various disciplines, we sketch a holistic and embodied model. As shown above, the models of Damasio and psychoanalysis fulfill the initially mentioned criteria especially well. Insights from both theories confirm and complement each other. For a technical model of the basic mechanisms of decision making, the psychoanalytic findings mentioned above are particularly well suited as an abstract framework (which is required for a holistic and coherent model), and Damasio's model is especially well suited for its concretization due to its consistent and holistic character in considering the interaction between body and mind.

The role of computer science here is to integrate the various models from other disciplines in a consistent and coherent model of decision making, which is deterministic and can therefore be validated by means of simulation. Thus, computer science enables a model building methodology and evaluation tool, by the means of agent based simulation. The basic principle of this approach has been illustrated in a previous article [23], where a functional view of emotions in the decision making unit of an agent was integrated. However, a holistic view considering the theories of Damasios was neglected. Nevertheless, the fundamental principle remains intact, i.e. affective evaluation processes are the foundation of evaluating data (psychic contents, thereby also actions and plans). These are those processes which use "quotas of affects" or derived evaluation variables to determine the relevance of data, based on memories, for decision making in a given situation (see Fig. 1). Thus, the evaluation of data is an incremental process on multiple levels – considering various evaluation

principles (pleasure and reality principles) and evaluation influences (bodily influences and through perception activated memories and fantasies).

Emotions are an additional level of this incremental and hierarchical multi-level evaluation model. They represent (1) "quotas of affects" from the drives, (2) emotions activated through perception and fantasy (memories associated with emotions), (3) the current pleasure. Hence emotions form a holistic representation of the *psychobiological* status of the agent (having information concerning the body and mental status) and can therefore consider the overall status of the agent in the evaluation. The final evaluation step is carried out by feelings, whereby depending on the intensity of the emotion, it is transformed into a preconscious feeling and subsequently a consciously "felt feeling" (in the sense of Damasios). The latter can be described as an inner perception, upon which the agent can reflect. As with the other valuation variables, feelings evaluate goals and plans by activating memories.

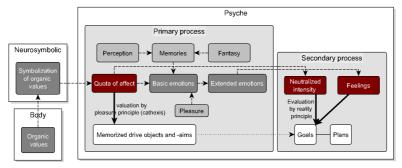


Fig. 1. Evaluation is an incremental process that considers multiple influences and principles.

By considering perception and fantasy, evaluation through feelings not only occurs in the terms of gaining pleasure, but also in terms of avoiding unpleasure, that is to say not just to support the fulfilment of drives but also to evaluate external events in terms of their potential to increase unpleasure. Evaluation generally serves to prioritize and select actions, mediating between the environment and the internal state (e.g. to fulfil desires in the environment and to adapt desires to external conditions).

The representation of the biological aspect of the psychobiological status is achieved through drives and body perception (proprioceptive and external perception). Whereby it must be emphasized that memory-based psychic representation (representation of drives and body representation) is used for emotion generation (and not the body signals as such). The psychic aspect of the psychobiological status is represented by the memories activated by environment perception and fantasies. In the sense of Damasio, one can conceptually speak of background emotions (red influences in Fig. 2), which can be considered as moods, and emotions triggered from the outside world.

The key concretization, when integrating Damasio's model, concerns the consideration of the embodiment by means of a mental representation of it. In this regard we follow the approach of considering the psyche as an information theoretical level of the physical world. This is reflected in the differentiation between the neural, neurosymbolic and physical levels (see Fig. 2), and also in the application of a memory based physical representation (see Fig. 2).

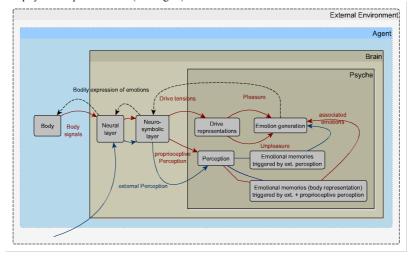


Fig. 2. A holistic and embodied emotion model.

#### 7 Conclusion

Interdisciplinary cooperation enabled us to outline a holistic model of emotion for humanoid agents. Due to the consideration of bodily influences and the evaluation of perceived events, the model outlined can be considered as being a combination of both appraisal and dimensional models, whereby neuroscientific inputs and embodiment are considered in developing a holistic model. Psychoanalysis and Damasio's neuroscientific model fulfill the initially mentioned criteria particularly well. Whereby psychoanalysis offers us an abstract holistic framework which can be concretized by means of Damasio's model. Both models also complement each other, as Damasio considers the interdependence of body and mind more concretely and is more consistent. Neuropsychoanalysis supports our approach of combining psychoanalysis with neuroscientific models, by revealing supporting evidence. Computer science enables the combination of the various models in one consistent and holistic model and offers an evaluation tool by means of simulation. Having outlined such a simulation model, the next steps lie in extending an existing implementation of a holistic functional model of the human mind [23], to integrate these new findings in a holistic model of human information processing. We expect that the integration of the various models in an overall evaluation model will yield new discoveries and opportunities in simulations.

#### References

- Panksepp, J.: Affective Neuroscience. The Foundations of Human and Animal Emotions. Oxford University Press, New York (1998)
- Edelman, G. M., Tononi, G.: Neuronaler Darwinismus. Eine selektionistische Betrachtungsweise des Gehirns. In: Meier, H., Ploog, D. (eds.): Der Mensch und sein Gehirn. Die Folgen der Evolution, 187-233, Piper Verlag, München (1997)
- Ledoux, J.: The Emotional Brain. The Mysterious Underpinnings of Emotional Life. Simon and Schuster, New York (1996)
- Cahill, L., Mcgough, J. L.: Mechanisms of emotional arousal and lasting declarative memory. In: Trends in Neurosciences 21, 284-299 (1998)
- Damasio, A. R.: Descartes' Irrtum. Fühlen, Denken und das menschliche Gehirn. [Descartes' Error. Emotion, Reason, and the Human Brain Putnam. Penguin edition, London (2005)] Deutscher Taschenbuch Verlag, München (2001)
- Damasio, A. R.: Ich f
  ühle, also bin ich. Die Entschl
  üsselung des Bewusstseins. [The Feeling of What Happens. Body and Emotion in the Making of Consciousness. Harcourt, New York (1999)] List Verlag, Berlin (2007)
- 7. Solms, M., Turnbull, O.: The Brain and the Inner World. An Introduction to the Neuroscience of the Subjective Experience. Other Press, New York (2002)
- Kandel, E. R.: In Search of Memory. The Emergence of a New Science of Mind. Norton, New York (2006)
- Damasio, A. R.: Der Spinoza-Effekt. Wie Gefühle unser Leben bestimmen. [Looking for Spinoza. Joy, Sorrow, and the Feeling Brain. Harcourt, New York (2003)] List Verlag, Berlin (2007)
- Freud, S.: The Interpretation of Dreams. The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume IV. Hogarth Press, London (1900)
- 12. Laplanche, J., Pontalis, J.B.: The Language of Psycho-Analysis.London: The Hogarth Press and the Institute of Psycho-Analysis, London (1973)
- Freud, S.: Formulations on the Two Principles of Mental Functioning. The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XII, Hogarth Press, London, 213-226 (1911)
- Freud, S.: The Unconscious. The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XIV. Hogarth Press, London, 159-215 (1915)
- Solms, M.: Sigmund Freud heute. Eine neurowissenschaftliche Perspektive auf die Psychoanalyse. Psyche. 60, pp. 829-859 (2006)
- Marsella, S. C., Gratch, J., Petta, P.: Computational models of emotion, In: K.R. Scherer, T. Bänziger, and E. Roesch, (eds.), Blueprint for Affective Computing. Oxford University Press, New York (2010)
- Ortony, A., Clore, G. L., Collins, A.: The Cognitive Structure of Emotions. Cambridge University Press, New York (1990)
- Dias, J., Mascarenhas, S., Paiva, A.: FAtiMA Modular: Towards an Agent Architecture with a Generic Appraisal Framework. In: Proceedings of the International Workshop on Standards for Emotion Modeling (2011)
- Mehrabian, A., Russell, J.A.: An approach to environmental psychology. MIT Press, MA (1974)

- Gebhard, P.: ALMA: A layered model of affect. In: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems. pp. 29-36 (2005)
- Becker-Asano, C., Wachsmuth, I.: Affective computing with primary and secondary emotions in a virtual human. Auton Agent Multi-Agent Syst. 20, pp. 32-40 (2010)
- Bach, J.: Principles of Synthetic Intelligence, Psi: An Architecture of Motivated Cognition. Oxford University Press, New York (2009)
- Schaat, S., Doblhammer, K., Wendt, A., Gelbard, F., Herret, L., Bruckner, D.: A Psychoanalytically-Inspired Motivational and Emotional System for Autonomous Agents. In: Proceedings of the 39th Annual Conference of the IEEE Industrial Electronics Society, pp. 6648 - 6653 (2013)