A Psychoanalytically-Inspired Motivational and Emotional System for Autonomous Agents

Samer Schaat, Klaus Doblhammer, Alexander Wendt, Friedrich Gelbard, Lukas Herret, and Dietmar Bruckner
Institute of Computer Technology
Vienna University of Technology
Vienna, Austria
{schaat, doblhammer, wendt, gelbard, herret, bruckner}@ict.tuwien.ac.at

Abstract—When developing autonomous agents, the problem of motivation poses one of the key questions. Following a bionic approach inspired by psychoanalysis, a multi-level motivational system to enable autonomous agents to pursue their own agenda and cope with their internal and external environment is presented. Considering embodiment, both homeostatic (extrinsic) and non-homeostatic (intrinsic) internal motivations are modeled. In particular, drives, which are represented on multiple levels, are modeled as the motivational element for agents’ actions. Representing an additional level of motivations, emotions are integrated as a co-determinant of decision making.

Keywords—Cognitive Automation; Artificial General Intelligence; Artificial Recognition System; Autonomous Agents; Motivations; Emotions; Cognitive Architecture; Artificial Intelligence

I. INTRODUCTION

An autonomous agent is a system situated within an environment, on which it acts in pursuing its own agenda [1]. There, central questions arise, such as: How and why does such an agent perform any action at all? Which mechanism is the source for such actions? How are decisions made and which external or internal actions are to be executed? How does the agent regulate its internal state? The problem of motivation is omitted in non-autonomous systems, which simply fulfill a task as a tool. In such case the motivation is situated outside of the system.

Hence, when developing autonomous systems the question of motivation must be solved first. Following a bionic approach, in the ARS (Artificial Recognition System) project [2] the human motivational system is used to develop a motivational system for an artificial agent. As an appropriate model for motivational processes the ARS project uses parts of the psychoanalytic model of human personality, which describes the functionality of motivations of a human being, presented in the psychoanalytic drive theory [3] [4]. Such a model is particularly appropriate for a functional approach to autonomous agents, due to its focus on internal motivations, as opposed to behavioristic approaches that focus on external stimuli as source for an agent’s behavior. Therefore the corresponding psychoanalytic theory is briefly introduced next.

The psychoanalytic drive theory describes how a bodily state of tension is represented in the human psyche and how it operates therein, with the result of a concrete operation and an emotional state. The psyche represents a drive, i.e. a somatic organ tension, in a drive representation with a drive source (the organic point of origin), a drive aim (an activity that decreases the organic tension) and a drive object (with which the drive can achieve its aim) [5]. In addition to this, the drive’s original organ tension, particularly its quantity, is represented in the psyche as a quota of affect, which is split in aggressive and libidinous components (see Section IV). In further processes in the psychic apparatus this quota of affect can be used in its components for cathexes1. Furthermore, the quota of affect causes emotional states and finally brings pleasure when drive tension is discharged, i.e. when drives are satisfied, using the drive object and drive aim. This also leads to consciously perceived feelings. Discharge of quota of affects leads to pleasure; accumulation of quota of affects leads to unpleasure. In combination with the pleasure principle these are basic psychoanalytic rules. The pleasure principle states that all psychic activity has primarily the aim to gain pleasure and to avoid unpleasure. The generation of pleasure can be regarded as a reward and feedback mechanisms, with the amount of pleasure indicating the goal achievement, i.e. reduction of the tension at the drive source.

Hence, psychoanalytic drive theory establishes a connection between bodily states, the psychic processing of them, and conscious actions, which result thereof. In addition, this theory serves as a basic concept for the generation of emotions and feelings and explains their impact on psychic processing, action planning and action execution. Emotions are an additional adjustment factor for decision making and execution. According to psychoanalytic theory, finally, an action in the external world is selected that leads to a maximal pleasure-gain – considering temporary accepted unpleasure due to necessary scheduling – over a certain period of time. This is called the reality principle, which extends and modifies the pleasure principle.

By searching and finding an object for drive satisfaction – a process that a drive necessarily implies – a connection between an agent that wants to satisfy its drives and the environment is established. Objects of the external world get the state of

1 Cathexis describes the attribution of quota of affect to psychic content. As a result this content is valued and activated for processes in the psychic apparatus and gets a specific relevance for inner psychic processes.
satisfactory objects. This principle also enables simple social interaction between agents, since other agents can also be regarded as drive objects. Hence, psychoanalytical drive theory is used to cope with the external world in satisfying an agent’s needs and as the basis of any object and social interaction.

II. REVIEW OF THE STATE OF THE ART

In general, motivations can be distinguished in internal, external, intrinsic and extrinsic [6]. Internal and external motivations differ in the source of reward, which is produced within the organism in the former case, and outside the organism in the latter case. As opposite to extrinsic motivations, which head for rewards, intrinsic motivations, which drive exploration and curiosity, lead to activity for its inherent satisfaction, in particular they aren’t homeostatic.

Although a motivational and emotional model has to be a central and integrated part of a process model of the human mind, most cognitive architectures (e.g. ACT-R [7]) lack this aspect. Some cognitive architectures, such as MicroPsi [9] and CLARION [8], consider motivational aspects. Nevertheless, the former does not consider a holistic approach and the latter lacks integration of an emotional system with its motivational model. Both mentioned models use the concept of drives, but as elaborated below, not in terms of the Freudian model. MicroPsi uses innate drives for intactness, energy, affiliation, competence and reduction of uncertainty. The latter two are considered cognitive urges that lead the agent into exploration. Events that raise the agent’s drives lead to negative reinforcement signals, whereas the satisfaction of a drive creates a positive signal.

CLARION is an appropriate example to elaborate the difference of current models to the motivational system within the ARS agent. It uses Hull’s concept of drives [10] to represent motivations. Nevertheless, a generalized notion of drive is used; drives represent “…internally felt needs of all kinds that likely may lead to corresponding behaviors…” [10]. Primary and secondary drives are distinguished. The former are separated in low-level primary drives, mostly concerned with innate physiological needs (such as hunger), and high-level primary drives (such as affiliation and belongingness), mostly concerned with social needs. Secondary drives are derived from primary drives and are learned by conditioning (e.g. the pursuit of money). Regarding representation CLARION distinguishes implicit drive representation and explicit goal representations, which are based on drives and are generated on the fly during an agent’s interaction with various situations. Generally, it seems that the distinction between implicit and explicit representations in CLARION complies with the distinction between unconscious and conscious processing in psychoanalysis. A clear distinction to the concept of drives in ARS is the determination of the drives’ strength, which in CLARION is determined by an internal deficit and an external stimulus (e.g. food in case of hunger). The latter determinant emphasizes Hull’s behaviorist approach and shows a central difference to a psychoanalytical approach to drives, where the inner world (e.g. the bodily deficit) is the point of departure and sole basis for determining a drive’s strength. The multi-level approach followed in the ARS model, especially the distinction in a primary and secondary process with distinct principles, enables a more flexible motivational system. The first level emphasizes a subjective approach to motivations, without the considerations of the external world. The agent focuses on its experience and internal state in the pursuit of maximal satisfaction. After the consideration of further levels, e.g. the defense mechanisms (see Section IV), only in the secondary process are external world and reality principle considered. This enables dynamic and flexible behavior. Another major difference to the ARS approach is the consideration of sexual drives only for the sake of reproduction. Following a psychoanalytical approach, where the genital factor is only one amongst many sexual modalities, sexual drives are the central source for intrinsic motivations. That is, the pleasure gained from the satisfaction of sexual drives is not coupled to the satisfaction of a bodily need. Moreover, the reduction of sexual drives is an end in itself and only aims for pleasure-gain without consideration of homeostatic aspects. Finally, unlike the usage of high-level motivations for primary drives, the ARS model is more consistent to the principle of embodiment, since all drives are based on bodily sources, but form the basis for higher motivations and goals.

Emotions are a significant part of a holistic motivational system. Most generally they can be regarded as an information framework (emotions as embodied information about value and importance [11] [12]) and as a representation of an agent’s internal state [12]. Computational models of emotion differ in various dimensions. One distinction is given by the design decision if emotions in an agent should be engineered or dynamically evolved. Another distinction of emotional models in agents is based on the theory they are based on. Emotional theories differ in which components are considered as intrinsic to an emotion (e.g. cognitions, somatic processes, behavioral tendencies and responses), their relationships and representation [12]. The currently predominant theory for computer models of emotions is cognitive appraisal theory, due to its focus on the connection between emotion and cognition [12]. In appraisal theory, emotions are argued to arise from the appraisal of external events and situations in relationship to an individual’s beliefs, desires and intentions. As a result, appraisals trigger cognitive responses, typically referred to as coping strategies (e.g., planning, procrastination) [12]. The focus in appraisal theories is to find a minimal set of appraisal criteria that explain the elicitation and differentiation of emotions. A popular model used in AI (e.g. EMA [13]) is the OCC model [14], a typical appraisal model, with appraisal criteria such as unexpectedness, appealingness and desirability. In appraisal theory emotions are only elicited by evaluations of external events and hence only considered for interaction purposes. This is not the case in the ARS model, where the generation of emotions is indeed influenced by external events, but is based on the agent’s drive state. In particular, the agent may be in a complex emotional state without considering the external world.

Another class of emotional theories follows a dimensional approach and conceptualizes emotions as points in a continuous dimensional space instead of as discrete entities [12]. A typical dimensional model is the three-dimensional PAD model [15], with the dimensions pleasure (a measure of
valence), arousal (indicating the level of affective activation) and dominance (a measure of power or control). A recent example of a computational model is WASABI [16], where the PAD model is used to model primary and secondary emotions.

III. ARS APPROACH

Using a top-down design approach, a process model of the human cognitive architecture is developed in the ARS project [2]. Since it considers holistic and functional aspects, the second topographical model of Freud [17] is chosen as a framework for developing a functional model of the human mind, since it is recognized as the only appropriate unitary cognitive theory of the human mind that considers top-down and functional aspects [18]. The leitmotif of the ARS model is to describe the functions that generate behavior instead of developing a narrow behavior model. This complies with the generative and broad approach of Artificial General Intelligence (AGI).

Starting from the topmost level in the top-down design process (i.e. the second topographical model’s three abstract functional units Id, Ego and Super-Ego), together with psychoanalytical consultants a finer-grained, more detailed description of the functions is generated with each new level of the model’s description. The third level of the ARS functional model is shown in Fig. 1 [19].

The functional model distinguishes models of the three functional units of the second topographical model (see Fig. 1 - green: Id, red: Super-Ego, blue: Ego). Additionally it considers the separation of the primary (see Fig. 1 upper part) and secondary process (lower part), which refer to unconscious and conscious processes, respectively. The two processes operate under different principles. The primary process does not consider logical principles or order, e.g. temporal and spatial relations. The central principle of the primary process is the pleasure principle, which represents the dynamics of drive wishes, i.e. the aim for maximal satisfaction of a drive, which is the representation of an agent’s bodily need. Hence, the motivational system, represented by the drive system, is the basis of the ARS functional model. This is not only the case for the primary process, the most flexible domain for drives, but also for the secondary process, which has to cope with the demands that are generated by drives in the primary process. Moreover, drives form the basis for further representations of motivations and appraisals, namely emotions and feelings. In this regard the motivational system in the ARS agent is a multi-level model, which breaks the complexity of the motivation system in multiple levels that process under different principles and influences.

IV. GENERATION OF DRIVES

As stated in Section I, drives are a pivotal component of the ARS cognitive architecture. This is in accordance with psychoanalytic theory where the satisfaction of drives or altered drives is the motivational element for human actions. Thus, in the following, the creation and processing of drives in the ARS system is described and the data structures are elucidated.

As shown in Fig. 2 below, drives are initiated in the ARS agent by certain drive sources. A drive source can be a body organ or an orifice of the body. These bodily values are simulated in the ARS model and are converted after the initialization phase into ARS data structures (see Fig. 3).

In the ARS cognitive architecture we distinguish between self-preservation drives und sexual drives. Self-preservation drives are, for example, the desire to eat or to drink. Self-preservation drives in ARS are always initiated pair-wise. One libidinous drive is accompanied by an aggressive opposite drive. For example, the libidinous drive to eat is accompanied by the aggressive drive to bite. On the other hand, sexual drives in ARS represent intrinsic motivations (see Section II) and consist of four components: anal, oral, phallic, and genital, which are also split in libidinous and aggressive components, the so-called drive components. The drive source of sexual drives is the libido stasis, a virtual value representing the bodily source of sexual drives.

Additionally, during the initialization phase of drives in ARS, using the agent’s experience with similar situations, the drive aim is added for each drive. E.g., if the agent is hungry, its drive aim is to eat, if the agent is thirsty, its drive aim is to drink. One can easily spot that the drive aims are later on converted into goals of the agent. Furthermore, as shown in
Fig. 2, using the agent’s experience, a drive object that can satisfy the drive is memorized for each drive. For example, according to the agent’s memory the drive object for the drive to eat can be a Wiener Schnitzel. Finally, each drive has to pass the defense mechanisms, which control the compliance with internalized rules (e.g., social rules). Drives can pass defense mechanisms, can be altered, or can be repressed by them. Particularly, primitive drives like the desire to eat or to drink can be sublimated into socially higher drives like support others with food, or, for example, collect items and store them for future use.

Hence, the basic components that are later needed in the ARS action planning module are stored within a drive’s data structure. Fig. 3 shows the data structure of drives in detail.

V. GENERATION OF EMOTIONS

Following a functional approach, emotions can be regarded as an adaptation to deal with internal (bodily needs) and external (perception) stimuli and events. Based on the agent’s drives and influenced by perception and activated memories, emotions are a further representation of the agent’s internal state. Since its consistent multi-level generation from bodily needs, drives and emotions around the agent’s autonomy, adaptation and social interaction on the body.

Emotion generation in the ARS model is a two-step process. In a first step basic emotions (anger, mourning, anxiety, joy, saturation, and elation) are generated based on the agent’s actual drives and perception. Afterwards, these emotions can be transformed into complex emotions (guilt, melancholia, shame, pity, disgust, hate, love, envy) in the course of defense mechanisms.

As shown in Fig. 4, besides the drives’ quota of affect the agent’s memories, which are activated by perception, are the second source for the emotion factors. They are activated by perceived situations. Using spreading of activation [20] in the agent’s associated memories, memorized situations that are activated by the agent’s perception subsequently activate memorized emotions, which influence the determination of the agent’s actual emotions via their four factors. In the course of spreading activation the association strength of memories with emotions co-determines their activation. Hence, the stronger a memory is associated with an emotion, the more likely it will be activated and influence the agent’s actual emotional state. This boost emphasizes the impact of emotions in a memory-based agent.

Following a memory-based approach, the activation of memories is the basis for the appraisal of objects and situations. In the course of the search after appropriate drive objects, perceived objects are evaluated according to their potential to satisfy actual drives. The impact of perception on the generation of emotions results in an additional evaluation of the external world. Hence, the ARS model considers a multi-level appraisal of the external world, a multi-level representation of the agent’s internal state and a multi-level adaptation to the internal and external world.

VI. DRIVES AND EMOTIONS IN DECISION MAKING

As a first step of conscious decision making, drives, emotions and images are extended with data structures for their use in the secondary process. The secondary process data structures consider logical associations, with assigned predicates, allowing the usage of reasoning. After their adaptation to secondary process data structures emotions are called feelings in the secondary process. Supported by feelings, drives define goals, which the agents will try to reach. They are the motor of decision making. A goal consists of the goal content (the goal type and goal source), a goal object, a preferred action to satisfy the goal, supportive data structures and the subjective importance of the goal. An example for a goal object is a food source, the action to reach the goal is to eat, supportive data structures are information either from perception or an act (i.e., a memorized sequence of images, i.e., situations) and the importance is represented by the quota of affect. As feelings are influenced by perceived situations or memorized contexts, they either reinforce or reduce the importance of a certain drive goal in a certain context or they evoke a reaction on something in the context or they evoke a reaction on something in the context or they evoke a reaction on something in the context or they evoke a reaction on something in the context or they evoke a reaction on something in the context.
experienced acts. It is save in the agent’s memory in the form of
that way, performed actions are evaluated. The results of the
successful in reducing bodily tension, it generates pleasure. In
if the action was
action plan is executed. The perceived feedback of the action
it. After choosing a goal, the associated
dependent on the current
drives, the
current feelings and the effort of reaching it. After choosing a goal, the associated
action plan is executed. The perceived feedback of the action
then potentially generates pleasure. If the action was
successful in reducing bodily tension, it generates pleasure. In
that way, performed actions are evaluated. The results of the
actions are saved in the agent’s memory in the form of
experienced acts.

Next, all possible, reachable goals from perception and
respective acts are extracted. The overall purpose of the
secondary process is to try to fulfill the agent’s drives or to
react on situations with matching feelings. The first stage of
decision making is to decide which drive shall be fulfilled or
which reaction is needed (signaled by a feeling). This is done
by evaluating the memorized goals regarding their ability to
fulfill the most important experienced goals (drive or feeling).
Then, the matched memorized goal’s associated feelings are
compared to the experienced feelings. The memorized feelings
can either increase or decrease the importance of a memorized
drive goal. The stronger the experienced feeling is, the more it
influences the memorized goals. For instance, if the agent
suffers big appetite (drive) and sees a food source together
with an enemy agent (perception), it recalls a similar act,
which reminds the agent that he is afraid (feeling) of the other
agent, which is causing the agent to feel anxiety. Although the
food source would satisfy the drive demand perfectly, the
memorized feeling (anxiety) of the situation reduces the
importance of that goal and the agent therefore prefers another
goal. The equivalent case happens if a food source is far away,
but the perceived situation activates a positive memorized
feeling like joy. Then, the agent may follow that goal although
it may be more difficult to reach.

In the second stage of decision making, potential action plans are added
to the reduced list of possible reachable and evaluated goals. In the case of goals
that origin from acts, action plans are considered, which were executed as the
act had been experienced. In that case, the agent has the possibility either to take
the same action as last time or to try something new. Each goal is then
evaluated regarding the effort needed to reach the goal or to react to a certain
situation. Effort evaluation influences the importance of reachable goals. Finally,
the goal with the highest importance is chosen. Hence, the goal is chosen
dependent on the current drives, the current feelings and the effort of reaching it.

To validate the functionality of the ARS model, use cases are defined [19]. A use case defines initial situations with possible outcomes. The initial situation is defined by four
determinants, comprising the agent’s memories, personality parameter, external world and the agent’s initial bodily needs. These factors determine the agent’s behavior and actions. By
varying the determinants and observing the agent’s behavior, we evaluate the functional model, i.e. we show that it is not a behavioral model and that we consider aspects of AGI. These
use cases are executed within the multi-agent simulation framework MASON [21] to validate the ARS model. By observing the agent’s behavior and using visualization tools, the result of the simulation is compared with the defined outcomes of the defined use case 1 (see below) to verify the
functionality. To display the internal states of the agents, so-called inspectors are implemented. These inspectors enable the
observation of the results of each function module in the
agent’s decision unit. Several inspector types to display data in
various ways are developed. For example, the line chart
inspector is shown in Fig. 5 and Fig. 6.

The initial situation of use case 1 is a world containing two
agents and a food source (Wiener schnitzel). One agent is
called Adam and contains the ARS decision unit. The other
one is a passive agent named Bodo. The goals to be
investigated with this use case are:
- Usage of drives as a motivation for behavior
- Activation and association of memory
- Using defense mechanisms to support Super-Ego rules

![Fig. 5. Adams drives during use case 1](image)

At first, Adam is initialized with medium strength of
hunger and has nothing in the range of vision. Fig. 5 shows
Adam’s drives during the use case. Due to his personality, the
aggressive part of the drive that represents hunger (green
curve) is higher than the libidinous part (orange curve). In the
beginning of the simulation, motivated by his drives, the agent
starts to search for a food source, i.e. drive objects that are memorized to satisfy his drives. After eating the food, the stomach tension, represented by the stomach drive, decreases. The reduction of the stomach drive generates pleasure (black curve). After a delayed time due to digestion, the rectum tension, represented by the rectum drive (rose and brown curve), rises and indicates the need to empty the rectum. From these drives and values from perception emotions are generated (see Section V). Fig. 6 shows Adams emotions during the simulation of use case 1. The emotions are generated depending on Adams drives and his perception (see Fig. 6). First, anger and anxiety grow because of displeasure and the sum of all aggressive drive components. After eating, displeasure and aggressive drive components decrease and pleasure is generated. This results in the generation of joy and reduction of the intensity of anger and anxiety.

In the secondary process reachable goals are extracted. In decision making, the strongest drive is selected to be fulfilled and only those reachable goals, which match the drive object of that drive, are further processed. Out of the different possibilities to fulfill the drive for hunger, the one is selected, which provides the highest pleasure, indicated by joy.

VIII. CONCLUSION

This paper presents a holistic and integrated model of a motivational and emotional system as a central part of a cognitive architecture and shows its validation through the simulation of a use case. We show how the psychoanalytic concept of drives and emotions can be adapted for a model that enables an artificial agent to pursue its own agenda. Particularly, we show how drives and emotions are used as a representation of the agent’s internal state for the adaptation to its internal and external world, especially for appraisal and action control. Due to its focus on internal motivations and its holistic approach, a psychoanalytic approach is more appropriate than a behavioralistic approach, which is used by other cognitive architectures for modeling motivations. In this way, appraisal and action control are not focused on external stimuli, as in other models, but based on the agent’s internal state.

REFERENCES