Towards Formal Modeling of Affective Agents in a BDI Architecture

Bexy Alfonso, Emilio Vivancos, and Vicente Botti

Universitat Politècnica de València,
Departamento de Sistemas Informáticos y Computación, Spain
{balfonso,vivancos,vbotti}@dsic.upv.es

Abstract. When designing agents to simulate human behavior, the incorporation of personality, emotions and mood into the agent reasoning process provides the agent with a closer to human behavior. We have designed an Open Affective Agent Architecture (O3A) based on widely accepted psychological theories. O3A offers a flexible way of integrating the affective characteristics of agents into their logic. We extend the operational semantic of the BDI agent language AgentSpeak modifying the traditional BDI reasoning cycle to incorporate affective components. This informal semantic description allows an agent to have a representation of not only the current state of the environment, but also of the agent affective state.

Keywords: Agents, Emotions, Semantic, Formalization

1 Introduction

Researches on multi-agent systems have traditionally focused on the search of rational solutions that maximize the quality or utility of the result. However, when an agent needs to simulate humans behavior, this kind of approaches is not the most appropriate. Human decisions are influenced, at greater or lesser extent, by affective characteristics such as the personality, the emotions or the current mood of the individual. In environments where agents must act like humans, the incorporation of emotions into the agent-based reasoning process provides the agent with a closer to human behavior. Many representations and formalizations of affective agents are based on the cognitive perspective of emotions. These formalizations model the appraisal process, the emotions dynamics, or the influence of affective traits on the cognitive processes. However a global formalization of the interrelation between the cognitive and the affective reasoning process is still required.

We have designed O3A, an Open Affective Agent Architecture, which is based on widely accepted psychological and neurological theories. An informal presentation of the main components of O3A can be found in [2]. O3A is built over a traditional BDI architecture and offers components to represent affective traits like personality, emotions and mood. The interaction of the architecture components with the cognitive processes of the agent, produces behavior biased by the agent mood. Our final aim is to to offer a feasible and comprehensive way of building affective agents using the O3A features. We also extend the reasoning cycle of AgentSpeak [4, 24], with new steps to facilitate emotional-based reasoning. The new components that we are proposing are flexible enough to be adapted to any particular requirement of the agent emotional processing.

2 Background

Several authors have proposed mechanisms to incorporate affective components into intelligent agents [3, 8]. For example in [8] Dias *et al.* propose FAtiMA, a BDI architecture that uses emotions and personality to influence agents' behavior. In contrast to O3A, FAtiMA does not have an explicit model to address personality. The agent's personality is implicitly represented in different agent processes and internal structures. In O3A the setting and fitting of the agent personality can be performed in a centralized way according to the widely accepted model of personality FFM (Five Factor Model) [11]. A more detailed comparisons of O3A with other similar approaches can be found in [2].

On the other hand, some works have proposed to incorporate affective traits into agent-based systems in a formal way. Some leading works [16, 19] are considered the base for further approaches that have modeled and formalized the emotion logics. Oatley & Jenkins' model of emotions [16] has inspired works like [22], and [23]. These works extend the KARO framework [14, 15] in order to use this formalism to give a logical account of emotional agents. In [14] Meyer models the dynamics of emotions and the influence of emotions on how an agent deals with its goals and plans. This paper distinguishes four emotions: happiness, sadness, anger and fear. These four emotions are also analyzed in [22], where the authors formally specify the emotion influence on behavior following the OCC model [17]. In [21] Steunebrink et al. make a qualitative formalization of the OCC model offering a method for calculating emotions intensities. Other KARO based model is the one proposed in [23]. Steunebrink *et al.* formalize the triggers conditions for some of the OCC emotions in BDI agents with achievement goals, specifically the appraisal part of OCC.

Rao and Georgeff's BDI_{CTL} logic [19] has been the starting point for some works like [18]. Pereira *et al.* present an improved version of their $EBDI$ logic to model the role of fear, anxiety, and self-confidence in a emotional BDI agent.

Other authors have focused on modeling the eliciting conditions for a subset of emotions, or on the emotions influence on the agent cognitive process. Our aim is to reach a formalization of affective agents at a higher level, offering a flexible approach where the processing of the emotions can be easily adapted to the specific emotional characteristics of the problem to be solved. Therefore, the formalization of our architecture uses general components in order to integrate the affective components with the BDI agent cognitive process. Mood represents the agent emotional state at every moment, and this representation is based on a dimensional theory. The agent has also a personality and emotions that are automatically generated as a result of the agent internal reasoning process and the agent interaction with the environment. In this article we present an ongoing work to extend the operational semantics used in Jason [4, 24], a well know agent-oriented programming language grounded in a logical computable language (AgentSpeak)[4].

3 Extension of the Jason operational semantics

3.1 Considerations to formalize the O3A architecture

Psychological and neurological theories try to explain the influence of emotions in human decisions. Our proposal considers two kind of emotions: primary and secondary. Primary emotions are "infant like" fast reactions easily deductible, while secondary emotions are the result of a more complex reasoning based on expectations or previous experiences [6, 17]. The O3A architecture proposes a set of new components to be included into a traditional BDI architecture with the ultimate goal of offering a computational model of these psychological concepts.

O3A is based on some of the most relevant theories of emotions and personality [5, 6, 17, 20]. These theories helped to build a formal specification of O3A from a cognitive perspective based on the appraisal theory, where the emotional state of an agent responds to a dimensional approach of the mood.

O3A uses *Primary and Secondary Emotions*. Lets start defining how these primary and secondary emotions are derived. We assume that percepts from the environment are labeled with the most common reactive emotions that individuals can experience. This assumption is based on the idea that an event often cause similar emotions in different individuals. For example, when facing a hurricane, people generally will feel fear. The *emotion reactive component* of O3A is responsible of deriving these primary emotions from percepts. On the other hand, secondary emotions are the result of a more complex and internal reasoning. In our approach, secondary emotions can emerge when events of any nature (internal or external) appear. The *emotion deliberative component* of O3A considers some variables: desirability, likelihood, expectedness, causal attribution, and controllability[1][10]. These variables produce an appraisal pattern that is used to derive secondary emotions. While primary emotions are reactive and fast responses, secondary emotions are the result of a "though process" and a more complex reasoning process that considers expectation and experience [6]. We are currently considering that both primary and secondary emotions belong to the same set of possible emotions: the OCC model, but a more detailed study is required to select the set of primary emotions and the set of secondary emotions to be considered for any specific domain of application.

In our approach the mood is represented in a three-dimensional space where three values describe the agent mood: Pleasure, Arousal and Dominance (PAD) [13]. We map each emotion of the OCC model to three values representing a point in the PAD space according to [9]. These three-dimensional points can be synthesized in a tuple which represents the mood. In our approach this 'synthesis'

[1] A more detailed explanation of their meaning can be found in [1]

is made by averaging the values of all the points corresponding to the appraised emotions. Then the calculated mood and the previous mood are merged following the proposal of [9]. When the calculated mood is similar to the previous one, the intensity of the new mood will be increased. Although this change doesn't depend on other processes, we assume that the mood update will occur in each reasoning cycle of the agent. We also propose a function that modifies the value of the mood according to the current mood and the agent personality. The personality is represented in O3A using the Five Factor Model [11] which describes quite accurately individual traits through five dimensions: openness, conscientiousness, extraversion, agreeableness, and neuroticism. The initial mood of an agent is calculated using the personality. The mapping from the agent five dimensions of personality to the three dimensions of the PAD space is done according to Mehrabian's work [12].

In O3A the calculated mood influences the agent behavior. Mood helps to prioritize the agent intentions to affront a given internal or external change. According to [7] the trait Dominance of the PAD representation is a good indicator of how much risk can be taken in order to achieve a goal. If an O3A agent has more than one option to respond to an event, the option selected will be that whose value of risk[2] is closer to the value of the agent dominance.

O3A has a function to determine *when and how beliefs are affected by the agent current mood*. As a result of that function, a belief will be evaluated depending on the current agent mood and knowledge. From this evaluation the belief can be considered for example positive or negative. This evaluation is implicit when percepts are labeled with the most common emotions categories experienced after this percept. For instance, the emotion "joy" could indicate a "positive" percept according to agent interests. Nevertheless, if the percept has not emotional category labels, the current state of the agent and the mood will be used to assign primary emotions to this percept. For example the percept *human_shadow* normally wouldn't produce any specific reaction, nevertheless, if the agent is very scared, this percept may produce the primary emotion "fear", and consequently a reaction to this fear. In [1] we show examples of O3A agents where emotions are used to improve the solution to some classical problems in behavioral economics.

3.2 Extension of the BDI reasoning cycle

In order to offer an integral description of a emotional BDI agent, we have extended the operational semantic of Jason [4, 24] with affective traits. The agent configuration is defined by a tuple $\langle ag, C, M, T, Mo, P, s\rangle$ where ag is the agent program, which contains a set of beliefs (bs) and a set of plans (ps). C, M, T, and s represent respectively the agent circumstance, communication parameters, temporary information for a reasoning cycle, and the label of the current step in the reasoning cycle . In order to include the affective state in the agent configuration, two elements have been added to the agent configuration tuple:

[2] The risk value is a property of plans and it is set by the programmer.

the agent current mood (Mo) and the agent personality (P). In each reasoning cycle the agent current mood can be modified. Mood (Mo) is defined by a tuple $\langle mP, mA, mD \rangle$ according to the dimensional theory of A. Mehrabian [13]. We represent the agent personality (P) using a tuple $\langle pO, pC, pE, pA, pN \rangle$, following the Five Factor Model of personality [11]. The initial agent mood is determined by the agent personality (following the mapping offered in [12]). The personality is also used to define a "equilibrium mood" of the agent.

Primary and secondary emotions change in each cycle of the BDI algorithm, so we have added two new components to the tuple that represents the temporary information (T). The temporary information (T) is represented by the tuple $\langle R, Ap, \iota, \varepsilon, \rho, PEM, SEM \rangle$, where R and Ap are the sets of relevant and applicable plans respectively[3]. ι, ε, and ρ are used to record a particular intention, event, and applicable plan. PEM and SEM represent primary and secondary emotions. Both sets of emotions can contain any emotion defined in the OCC model. We have also added to the O3A architecture the "surprise" emotion that can be derived as a primary or secondary emotion.

O3A has added new steps to the Jason reasoning cycle and thus the corresponding transitions rules. The resulting reasoning cycle is shown in figure 1. New steps are colored while new and modified transitions are represented by arrows with dashed lines. Although the Jason formalization considers that an agent can perceive new information from the environment, to the best of our knowledge, there is no explicit step in the reasoning cycle for this task. We have decided to make this step explicit as a initial step (`Perceive`). The next step is `DerivePEM` step, in charge of deriving primary emotions. It is also a task of the `DerivePEM` step to endow beliefs (derived from percepts) of reactive emotions associated to the agent emotional and cognitive state (the function of the *Beliefs component* of the O3A architecture). The next two steps are `DeriveSEM`, in charge of deriving secondary emotions, and `UpMood` which updates the mood based on the new appraised emotions. `DeriveSEM` follows `RelPl` (which determines the relevant plans), since the derivation of secondary emotions considers the relevant plans for a triggering event. The current mood is updated using the new primary and secondary emotions. If no new emotions are apprised, the O3A reasoning cycle goes on with the `ApplPl` step, which determines the applicable plans. In O3A the agent current mood is also used to select the next applicable plan affecting the `SelAppl` step. Finally the `MoodDecay` step determines the mood tendency to return to its equilibrium state.

In the operational semantic of Jason [4, 24] the elements denoted by T_R, T_{Ap}, T_ι, T_ε, and T_ρ are used to represent the current set of relevant plans T_R, the current set of applicable plans T_{Ap}, the current intention T_ι, event T_ε, and applicable plan T_ρ. We have added to this notation the T_{PEM} and T_{SEM} elements which indicates the PEM (primary emotions) and SEM (secundary emotions) components of the of temporary information tuple (T).

[3] In Jason the relevant plans are those that are candidate to be executed as a consequence of the activation of an event, while applicable plans are those relevant plans whose condition regarding the state of the world is satisfied.

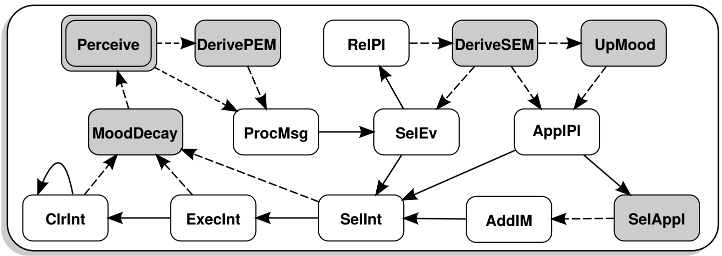

Fig. 1. Extension of the reasoning cycle of AgentSpeak

We have also defined new functions that are used by the agent interpreter on some steps of the reasoning cycle. These functions can be customized by the agent programmer: $IniMood(P)$ that is used to determine the agent "equilibrium mood"; $SEMDer(ag_{bs}, T_\varepsilon, T_R)$ is used to derive secondary emotions; $UpMo(Mo, T_{PEM}, T_{SEM})$ used to update the current mood; $MoodDc(Mo, P, \delta)$ determines how mood will decay in each reasoning cycle; $NewP(PercSet, ag_{bs})$ determines new percepts; and $RemP(PercSet, ag_{bs})$, which determines what percepts are not longer detected in the environment.

3.3 Transitions between the steps of the O3A reasoning cycle

In this section we present the description of the new and updated steps added to the Jason operational semantics. Note that, as described in the previous section, the initial state of the O3A reasoning cycle is denoted by the tuple $\langle ag, C, M, T, Mo, P, \texttt{Perceive}\rangle$. The mood Mo has a initial value (and a "equilibrium mood") determined by the function $IniMood(P)$, that rerceives the agent personality as a parameter. To the best of our knowledge the operational semantics for the perception process has not been defined yet, so we carefully provide a transition rule to this end.

Perception: This is the initial step (**Perceive** in Fig. 1). The agent checks the environment and updates its belief base. The function $NewP(PercSet, ag_{bs})$ is used to add the new perceived beliefs. When removing those percepts that are not longer detected in the environment, O3A uses the function $RemP(PercSet, ag_{bs})$. These functions start from a set of percepts ($PercSet$) that are detected in the environment where the agent is situated. In the "perceptions update rule", if no new percepts are acquired, then the reasoning cycle goes directly to the `ProcMsg` step without deriving primary emotions.

Derivation of primary emotions (DerivePEM): Primary emotions are the result of reactive emotional responses to the agent perceptions. These reactive responses are similar for most individuals, and therefore O3A assumes that percepts are labeled with the most common categories for reactive emotions that people usually experience. The structure as well as what should be these labels is up to the programmer. Primary emotions can also be inferred given the percepts' nature and the current mood of the agent. The function $PEMDer(ag_{bs}, Mo)$

performs the task of deriving primary emotions. In this O3A first approach only explicit labels for emotion categories are considered.

Relevant Plans (RelPl): In the original Jason reasoning cycle there were two transitions starting from the step `RelPl` because there were two cases: when there are relevant plans for the selected event, and when there aren't. In O3A there is only one transition since secondary emotions can be derived even if there are no relevant plans (which is the next step).

Derivation of Secondary Emotions (DeriveSEM): In this step secondary emotions are derived using the function $SEMDer\,(ag_{bs}, T_{\varepsilon}, T_R)$. There are three possible transitions from this step. The cycle can move towards `UpMood` if primary or secondary emotions were appraised (first condition). If no emotions were appraised but there are relevant plans (second condition), the cycle goes on with the `ApplPl` step. If none of these two previous conditions hold, then the next step will be `SelEv`. Although this is not an optional step in the BDI reasoning cycle, it is up to the programmer to decide when and how secondary emotions should be derived.

Update Mood (UpMood): After primary and secondary emotions are appraised, the current mood is updated. The function $UpMo(Mo, T_{PEM}, T_{SEM})$ represents this task.

Selection of an Applicable Plan (SelAppl): This transformation rule is similar to its Jason counterpart, except the S_{Ap} function (which selects one applicable plan) has an additional parameter: the current mood (Mo). The agent current mood influences the process of selecting the action to respond to one event or to reach a goal. This function can be customized by the programmer.

Mood Decay (MoodDecay): This step determines how mood decays. Information related to the agent personality is used by O3A in a function represented by $MoodDc(Mo, P, \delta)$, where δ is the decay rate that determines how mood will decay in each reasoning cycle.

4 Conclusions

In this work we offer an informal semantic description of O3A, an Open Affective Agent Architecture. This formalization integrates the affective characteristics of an O3A agent into the BDI reasoning cycle. We have extended the operational semantic used in Jason. Our architecture has its grounds on widely studied psychological and neurological theories, and it can be easily adapted to the particular emotional requirements of the problem to be solved. Agents that are created according to the proposed reasoning cycle will be able to have a representation of the current state of the world as well as of its own affective state. Therefore the O3A decision making process will be influenced by emotions, mood and personality. Different implementations of each of the proposed functions will modify the way in which the affective changes occur in the agent, so, although we offer a default implementation for the functions, they can also be fitted to particular requirements or theories.

This is a work in progress, and we are currently engaged in completing this formalization. Our immediate aim is to evaluate the O3A architecture and its formalization by enriching previous experiments [1] in order to reach agents' behaviors closer to humans behavior.

Acknowledgments. This work was supported by the Spanish government grant MINECO/FEDER TIN2012-36586-C03-01 and HUMBACE (Human and Social Behavior Models for Agent-Based Computational Economics) project. Thanks also to the Research and Development Support Programme of the Universidad Politécnica de Valencia PAID-06-11 Project.

References

1. B. Alfonso, E. Vivancos, V. Botti, and P. Hernández. Building Emotional Agents for Strategic Decision Making. In *Proceedings of the 7th ICAART*, pages 390–397, 2015.
2. B. Alfonso, E. Vivancos, and V. J. Botti. An Open Architecture for Affective Traits in a BDI Agent. In *In Proceedings of the 6th ECTA. Part of the 6th IJCCI*, pages 320–325, 2014.
3. C. Becker-Asano and I. Wachsmuth. Affective Computing With Primary and Secondary Emotions in a Virtual Human. *AAMAS '10*, 20(1):32–49, 2010.
4. R. H. Bordini and J. F. Hübner. Semantics for the Jason Variant of AgentSpeak (Plan Failure and Some Internal Actions). In *Proceedings of the 19th Conference ECAI 2010*, pages 635–640, Amsterdam, The Netherlands, 2010. IOS Press.
5. C. Castelfranchi. Affective Appraisal Versus Cognitive Evaluation in Social Emotions and Interactions. In *Affective Interactions*, pages 76–106. Springer, 2000.
6. A. R. Damásio. *Descartes' Error: Emotion, Reason, and the Human Brain.* Quill, 1994.
7. H. A. Demaree, M. A. DeDonno, K. J. Burns, P. Feldman, and D. E. Everhart. Trait dominance predicts risk-taking. *Personality and Individual Differences*, 47(5):419–422, 2009.
8. J. Dias, S. Mascarenhas, and A. Paiva. Fatima Modular: Towards an Agent Architecture With a Generic Appraisal Framework. In *Proceedings of the International Workshop on Standards for Emotion Modeling*, 2011.
9. P. Gebhard. ALMA: A Layered Model of Affect. In *Proceedings of the 4th IFAAMAS*, pages 29–36, NY, USA, 2005. ACM.
10. S. C. Marsella and J. Gratch. EMA: A Process Model of Appraisal Dynamics. *Cognitive Systems Research*, 10(1):70–90, 2009.
11. R. R. McCrae and O. P. John. An Introduction to the Five-Factor Model and its Applications. *Journal of personality*, 60(2):175–215, 1992.
12. A. Mehrabian. Analysis of the Big-Five Personality Factors in Terms of the PAD Temperament Model. *Australian Journal of Psychology*, 48(2):86–92, 1996.
13. A. Mehrabian. Pleasure-Arousal-Dominance: A General Framework for Describing and Measuring Individual Differences in Temperament. *Current Psychology*, 14(4):261–292, 1996.
14. J.-J. C. Meyer. Reasoning About Emotional Agents. *International Journal of Intelligent Systems*, 21(6):601–619, June 2006.

15. J.-J. C. Meyer, W. Van der Hoek, and B. Van Linder. A Logical Approach to the Dynamics of Commitments. *Artificial Intelligence*, 113(1):1–40, 1999.
16. K. Oatley and J. Jenkins. *Understanding Emotions.* John Wiley & Sons, 1996.
17. A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions.* Cambridge University Press, July 1988.
18. D. Pereira, E. Oliveira, and N. Moreira. Formal Modelling of Emotions in BDI Agents. In F. Sadri and K. Satoh, editors, *Computational Logic in Multi-Agent Systems*, volume 5056 of *Lecture Notes in Computer Science*, pages 62–81. Springer, 2008.
19. A. S. Rao and M. P. Georgeff. Decision procedures for BDI logics. *Journal of logic and computation*, 8(3):293–343, 1998.
20. R. M. Ryckman. *Theories of Personality.* PSY 235 Theories of Personality Series. Thomson/Wadsworth, 2007.
21. B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer. A Formal Model of Emotions: Integrating Qualitative and Quantitative Aspects. In G. Mali, C. Spyropoulos, N. Fakotakis, and N. Avouris, editors, *Proceedings of the 18th ECAI'08*, pages 256–260, Greece/Amsterdam, 2008. Patras / IOS Press.
22. B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer. A Formal Model of Emotion-Based Action Tendency for Intelligent Agents. In *Proceedings of the 14th EPIA '09*, pages 174–186, 2009. Springer-Verlag.
23. B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer. A Formal Model of Emotion Triggers: an Approach for BDI Agents. *Synthese*, 185:83–129, 2012.
24. R. Vieira, Á. F. Moreira, M. Wooldridge, and R. H. Bordini. On the Formal Semantics of Speech-Act Based Communication in an Agent-Oriented Programming Language. *J. Artif. Intell. Res.(JAIR)*, 29:221–267, 2007.