An Open Architecture for Affective Traits in a BDI Agent

Bexy Alfonso, Emilio Vivancos and Vicente J. Botti
Department of Informatic Systems and Computing, Universidad Politécnica de Valencia, Valencia, Spain

Keywords: Agents, Emotions, Personality, Mood, Architecture, BDI, Human Behavior.

Abstract: Recently an increasing amount of research focuses on improving agents believability by adding affective features to the traditional agent-based modeling. This is probably due to the demand of reaching ever more realistic behaviors on agent-based simulations which extends to several and diverse application fields. The present work proposes O3A: an Open Affective Agent Architecture, which extends a traditional BDI agent architecture improving a practical reasoning with more "human" characteristics. This architecture tries to address disperse definitions combining the main elements of supporting psychological and neurological theories.

1 INTRODUCTION

Artificial intelligence constantly evolves. New methods, algorithms and techniques are created or improved in order to achieve more sophisticated solutions. The agents field is not far behind. With the vision of a computational agent as a reactive and proactive entity, with its own goals, desires, sensing and planning mechanisms, more steps are taken to simulate human behavior and human interactions. Nevertheless for a simulation that truly reflects how humans behave, it is necessary to model also the affective side.

Neuroscience methods have found significant evidence that emotions are associated to regions in the brain in charge of controlling the related functions. They have also demonstrated that these functions are necessary for the individual because they act as internal heuristics guiding decisions in uncertain situations. Psychological and cognitive sciences have also made important contributions to further research on *emotional computing*, considered a "computing that relates to, arises from, or influences emotions" (Picard, 1997). One of the challenges to deal with when addressing issues in emotional computing is to effectively combine results of several and varied sciences. Specifically cognitive science has received special interest by affective computing researchers due to its suitability for creating computational models. Among the psychological perspectives of personality and emotion, the cognitive is the most widely studied because, in some degree, it is contained in the other perspectives. Results of research in affective computing have been applied in fields like education, training, therapies and the simulation of disaster situations.

The aim of this work is to present O3A (an **O**pen **A**ffective **A**gent **A**rchitecture): general enough to consider aspects of rational agents as well as their affective nature and whose components can be customized or replaced according to the domain requirements. In this article it is shown how this mechanism of emotion can be integrated into a practical reasoning architecture. We take the widely accepted BDI (Beliefs, Desires and Intentions) architecture of agents as starting point and we also endow the agent with the main affective concepts inherited from supporting sciences. We provide a summary of the main concepts extracted from psychological and neurological literature that had shed light over our work and briefly comment some significant related works. Then we present our architecture and its main components, pointing out how it is integrated with a traditional BDI algorithm. Final conclusions provide some annotations of the work performed.

1.1 Motivation

The BDI agent architecture has been widely accepted in the agents community because it has important advantages compared to other agent architectures (logic based, reactive, or layered architectures (Weiss, 1999)). It has been able to effectively reflect the human reasoning process having strong philosophical roots. Besides, many logical and software frameworks based on the BDI architecture have been developed. It offers cognitive processes and compo-

Alfonso B., Vivancos E. and Botti V..
An Open Architecture for Affective Traits in a BDI Agent.
DOI: 10.5220/0005153603200325
In *Proceedings of the International Conference on Evolutionary Computation Theory and Applications* (ECTA-2014), pages 320-325
ISBN: 978-989-758-052-9

nents, meaning the processing of perceptions, beliefs, and goals that are necessary and are directly affected by the emotional internal state (Castelfranchi, 2000). That makes this model to become a suitable alternative in order to represent the practical rational side of an agent. Although many approaches have tried to improve the BDI architecture with the human emotional process, they result disperse and confusing on their definitions, and also sometimes they don't follow an incremental line where one reuses the others results (Marsella et al., 2010).

According to Castelfranchi (Castelfranchi, 2000) the basic elements constituting emotions are: beliefs, evaluations, goals, arousal, and the "tendency towards action". *Beliefs* are individual representations of the world that activate emotions with a level of *arousal*, and motivate the conduct. Then emotions come from the interpretation of facts and sensations ("recognition" of emotions). On the other hand individuals tend to avoid or maybe to pursue some *goal* if this leads to a desirable emotional state; in this sense emotions itself can be considered as goals. But also emotions can monitor the goals offering guidance about the goals consequences and besides they can activate new goals. Castelfranchi highlighted the difference between the two kinds of appreciation of events' *evaluations*: 1) adaptive and non rational, which are automatic, intuitive and unconscious orientations to what can be wrong or bad, called also *primary* in the literature (Ortony and Turner, 1990; Damásio, 2005; Becker-Asano and Wachsmuth, 2010), and 2) declarative or explicit, which is an evaluation based on reasoning, can be explained and is closely related to goals called also *secondary*. These and other ideas have laid the ground for structuring O3A.

2 BACKGROUND AND RELATED WORK

The psychological literature related to affective human characteristics talks about cognitive concepts like emotions, moods, feelings, and personality (Frijda, 1987; Castelfranchi, 2000; Ryckman, 2007). They generally agree in that *emotions* are reactions as a consequence of agents, other actions and/or objects (Ortony et al., 1988). *Mood*, as emotions, is considered to be an experiential component too but it is not necessarily associated with a cause, lasts longer and has less intensity than emotions (Mehrabian, 1997). On the other hand, *personality* is seen as a set of individual characteristics which generally influence motivations and behaviors of the agent (John and Srivastava, 1999; Ryckman, 2007). Among the perspectives that address personality and emotions, the cognitive perspective has special relevance for affective computing due to its suitability to be used in computational applications. Moreover, in the neurological field we found important works which have laid the foundations for future applications in artificial intelligence and human-computer interaction areas. LeDoux and Damásio made important contributions in this area (LeDoux, 1998; Damásio, 2005). They found evidences of the relationship between emotions and the way in which the brain works.

Based on some interdisciplinary works (Mehrabian and Russell, 1974; Ortony et al., 1988), several approaches have tried to embody agents or virtual characters with affective traits and expressive functions (Gebhard, 2005; Becker-Asano and Wachsmuth, 2010; Neto and Silva, 2012). Many of them are also based on the BDI architecture. For example an interesting work is the one proposed in (Parunak et al., 2006) where authors propose the DETT (Disposition, Emotion, Trigger, Tendency) model for situated agents. It is a domain specific approach that aims to model agents whose goal is to anticipate the actions of an enemy in a combat scenario. It proposes a reasoning for the agents suitable to perform fast actions, and it takes the features of the OCC model (Ortony et al., 1988) for extending a BDI architecture. In this model emotions influence *perception* and *analysis*. The *disposition* element modulates the appraisal, and so, the way that emotions are triggered from beliefs. On the other hand, a *tendency* is imposed to intentions in that analysis is modulated by emotions. The analysis process together with the agent desires produce the intentions. The EBDI architecture for emotional agents (Jiang et al., 2007) is another similar work. The author points out that the way that changes in the environment affect emotions and how these changes influence human behavior differ even individually, so he separates the practical reasoning from the emotion mechanism. But he doesn't use any psychological concept to represent such individual differences. In his work a distinction between *primary* and *secondary* emotions is made. *Primary* emotions are considered reactive responses of the brain and *secondary* emotions appear later and can be caused by *primary* ones or by more complex chains of thinking. Similarly in (Neto and Silva, 2012) an architecture for emotional agents is presented. This architecture includes a *personality* and a *mood* component and it describes how affective characteristics influence perception, motivation, memory and the decision making in a BDI architecture. Its emotional component uses the results of the ALMA project (Gebhard, 2005), and therefore integrates personality, emotions and mood components which are modeled through the Five Factor Model (FFM) of personality (McCrae and John, 1992; Goldberg et al., 1990), the OCC model (Ortony et al., 1988) and the PAD model (Mehrabian and Russell, 1974) respectively. This approach focuses on the cognitive state. The coping actions (Marsella and Gratch, 2009) are linked to “filters” that select each time the percepts or facts that are aligned with the agent emotional state or plans that doesn't lead the agent to an undesirable emotional state. An approach that integrates the majority of issues of previous approaches is the one proposed in (Becker-Asano and Wachsmuth, 2010). The authors propose an architecture suitable for virtual characters. They include primary and secondary emotions where the former is directly linked to expressive capabilities like facial expressions and the latter comes as the result of reasoning about current events, and by considering expectations and past experiences. In this approach mood values are in a bipolar scale and move from positive to negative. Emotions mix theories from P. Ekman (Ekman, 1999) and OCC. This architecture also included a *memory* component which is used to generate the character expectations.

Marsella et. al. present a general computational appraisal model (Marsella et al., 2010), which tries to cope with the main issues associated with emotions and their impact on the cognitive processes and state of the agent. This architecture is composed of three elements: the person-environment relationship, the appraisal variables, and the emotion or affect component. These components are linked in such a way that links have associated a transformation model and the component in the pointing side of the link needs elements from the previous component in the link. The *person-environment relationship* represents the relationship between the entities in the social environment, beliefs, desires or intentions, and the external events. Other works like (Dastani and Meyer, 2006; Steunebrink et al., 2012) also propose the syntax and semantic of a logic-based agent language and the logical formulation of the OCC model of emotions respectively. They specify how external stimuli and the current state of the agent beliefs and goals may derive in emotions. Specifically in (Dastani and Meyer, 2006), the authors propose transition rules for the execution of actions according the current emotions, so emotion evaluation, as well as action tendency is described through programming constructs. These approaches do not contradict the proposed architecture. Conversely they could support some of its components (as detailed in section 3.1).

Generally existing approaches offer specific structures for particular models of emotions where primary and secondary emotions are often treated through the same mechanism. Besides, the concept of a central core affect (or mood) found in the psychological literature as well as the individual differences like those determined by the *personality* are not always considered. In section 3 we try to address this issues by presenting our affective agent architecture called O3A (an **O**pen **A**ffective **A**gent **A**rchitecture).

3 THE O3A ARCHITECTURE

The O3A architecture addresses mental, cognitive and motivational components of emotions, what, according to (Castelfranchi, 2000), makes them a complex, hybrid subjective state of mind. O3A is based on the appraisal theory that is currently the most accepted computational model of emotions. O3A is general because its components are integrated into a BDI agent-based architecture aiming to provide a model useful to build believable agents behaviors in any domain.

3.1 Main Components

The O3A architecture (Figure 1), consists of four main components in charge of controlling the agents emotional issues. These components have a well defined interface that allows to easily re-implement any component if other emotional approach is used in the future.

The Appraisal Component is in charge of deriving emotions from perceptions and from the agent state. Two sub-components make this task. The ***Emotion reactive component*** takes what has been observed in the environment to obtain the *Primary emotions*. The function of this component is based on the idea of the non-rational, automatic, unconscious and adaptive evaluation of events stated in (Castelfranchi, 2000), that derives in primary emotions. In order to implement this component we have labeled each percept with a set of “more probable emotions” to be experimented by the agent after perceiving the event. Our taxonomy of emotions is based on the cognitive perspective of emotions proposed by Clore & Collins in 1988 (Ortony et al., 1988), that was later improved by Ortony in 2003 (Ortony, 2003). This model of emotions, popularly know as OCC, consist of 22 emotions types with their specifications and attributes. These emotions are linked to eliciting conditions, and their intensity is affected by a set of variables. ***The emotion deliberative component*** is in charge of deriving *Secondary emotions*, and corresponds to the declarative and explicit evaluation of events that can be explained and argumented upon (Castelfranchi, 2000). As these emotions are the result of more complex chains of thinking it is necessary to check current beliefs, options available, as well as information from previous results (e.g. how successful or ineffective has been an action or plan, or the emotional experience after a plan in previous executions). We use the ideas proposed in (Marsella and Gratch, 2009) for this component, where emotions emerge from evaluating appraisal variables of propositions. A work like (Steunebrink et al., 2012) is also suitable to be used in this component of the architecture. The authors make logical formalizations to derive emotions starting from the agent mental attitudes.

The Beliefs Component determines how *Current Mood* influences the percepts before they become the agent's *Beliefs*. It starts from the idea that mood can intensify or blur perceptions and hence generate different perceptions for each agent (Neto and Silva, 2012; Niedenthal and Setterlund, 1994).

The Mood Component feeds on the agents *Personality* to establish the agent's initial mood and to update the *Current Mood*. This is based on the idea that individuals differentiate from each other in the way their mood changes depending on their personality traits. An explosive individual may reach a mood with high levels of arousal more easily than one that has a less neurotic personality. The Mood component updates the *Current Mood* also considering the *primary* and *secondary* emotions elicited. We use the dimensional representation of the core affect made by A. Mehrabian and A. Russell (Mehrabian and Russell, 1974; Russell and Mehrabian, 1977) in order to describe mood. A three dimensional space whose dimensions are Pleassure, Arousal and Dominance (PAD) describes any emotional state of the agent. This component also deals with the concurrence of "evaluation and appraisal about the same entity/event", which can "give rise to convergence and enhancement of the valence, or to conflicts" (Castelfranchi, 2000). Another issue that is addressed in this component is the duration and the return to an "equilibrium" state of the mood.

The Coping Component decides if the changes experimented in the current mood deserve to take actions in the cognitive processes of the agent, determining in this case the way intentions are selected to be achieved. As stated in (Castelfranchi, 2000) the responses of the agents to external events having the same knowledge, desires, and abilities will be different depending on each agent internal state. We use a plans prioritization strategy but elements of previ-

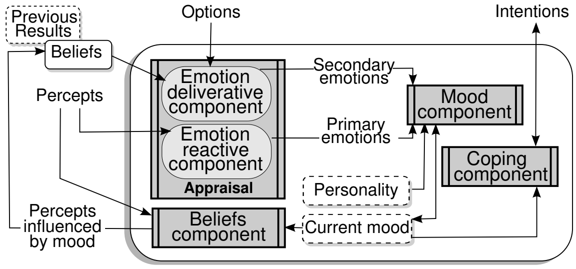

Figure 1: Main components of the O3A architecture.

ous works can also be used, such as the programming constructs offered in (Dastani and Meyer, 2006), that offer transition rules for the execution of actions according to the current emotions[1].

Just as the mood can be modeled through a dimensional representation, there are models like the Five Factor Model of personality (McCrae and John, 1992) that are useful to build *Personality* profiles. This model uses a set of five dimensions to describe each individual personality quite accurately.

3.2 Integration into a BDI Architecture

Starting from the control cycle of a BDI practical reasoning agent offered in (Bordini et al., 2007), we propose the main cycle for agents under O3A. The algorithm of Figure 2 integrates a traditional BDI agent architecture with the emotional components proposed in O3A. Note that this architecture is independent of the internal implementation of these emotional components that can be substituted in the future. In lines 1-3 the initial values for Beliefs (B_0), Intentions (I_0) and Current Mood (M_0) are respectively set. In particular, Current Mood is initialized according to the Personality profile (P). We use the proposal made in (Mehrabian, 1996) to establish the correlation between the PAD space (Mehrabian and Russell, 1974) and the Five Factor Model (McCrae and John, 1992). Lines 4-31 represent the basic control loop. The main actions performed in this loop are: observe, execute, and update options. In line 5 the next percept (ρ) is observed from the environment. This percept may have associated a set of the "more probable emotions" to experiment with this percept. In line 6 Beliefs are updated considering the agent's current Beliefs (B), the new percept (ρ), and the Current Mood (M). In this algorithm, Desires are considered "candidate options", and are determined in line 7 on the basis of the current Beliefs and Intentions. In lines 8 and 9

[1]This is valid only if, according to the requirements of the domain, a centralized processing of a "mood" is not considered and the emotional internal state is represented directly by the emotions "evaluated".

```
1:  B ← B_0; {B_0 are initial beliefs}
2:  I ← I_0; {I_0 are initial beliefs}
3:  M_0 = M = initialize_mood(P); {P: personality}
4:  while (true) do
5:      get next percept ρ via sensors;
6:      B ← get_new_beliefs(B, ρ, M);
7:      D ← get_options(B, I);
8:      PEm ← get_primary_Em(ρ);
9:      SEm ← get_secondary_Em(B, M, D);
10:     M ← update_M(PEm, SEm, M, P);
11:     I ← filter(B, D, I, M);
12:     π ← plan(B, I, Ac); {Ac: set of actions}
13:     while not (empty(π) or succeeded(I, B) or
        impossible(I, B)) do
14:         α ← first element of π;
15:         execute(α);
16:         π ← tail of π;
17:         observe environment to get next percept ρ
18:         B ← get_new_beliefs(B, ρ, M);
19:         PEm ← get_primary_Em(ρ);
20:         SEm ← get_secondary_Em(B, M, D);
21:         M ← update_M(PEm, SEm, M, P);
22:         if (reconsider(I, B, M)) then
23:             D ← get_options(B, I);
24:             I ← filter(B, D, I, M);
25:         end if
26:         if not (sound(π, I, B)) then
27:             π ← plan(B, I, Ac);
28:         end if
29:         SuccRate_π ← get_succ_rate(SuccRate_π, π);
30:     end while
31: end while
```

Figure 2: Control loop for an emotional BDI agent.

Primary and Secondary emotions are obtained. Secondary Emotions also consider Desires and Beliefs because they are the product of a more complex deliberative process that may emerge from evaluating the agent options, current Beliefs (including past experiences and/or expectations) and current Mood. As it has been posed in section 3.1, some previous approaches may be used in this last two steps such as (Dastani and Meyer, 2006; Steunebrink et al., 2012). In particular we derive PEm directly from percepts and SEm according to (Marsella and Gratch, 2009). Current Mood is updated taking into account Primary and Secondary Emotions as well as the previous Mood and the agent's personality (line 10). The main goal of this step is to perform a transformation from a set of emotions to a mood in a coherent way. Intentions are updated in the same way in line 11 on the basis of the selected Desires and the Current Mood. Then in line 12 the *plan* function generates a plan for achieving the selected intentions. The actions of the loop in lines 13-30 will be executed as soon as the plan is not empty, succeeded, or impossible. In lines 14-16 the first action of the plan is selected and executed, and the plan is updated with the remaining actions. Lines 18-21 represent a pause that the agent makes to detect changes in the environment (which is verified in line 17), and reconsider its Intentions, deriving again Primary Emotions, Secondary Emotions and Mood as previously in lines 8-10. If it's worth to reconsider and to deliberate (Intentions may suffer changes according to the current state and Mood) the Desires and Intentions are reevaluated (lines 23-24). In lines 26-28 a replanning is made in case the current plan doesn't fit well any more with the current Intentions and Beliefs[2]. Finally, in line 29 a measure of the relation between the times the plan has successfully fulfilled the committed Intentions and the times it has been executed is saved. This indicator is used for future deliberations[3].

4 CONCLUSIONS

The O3A architecture is inspired on the most prominent results of psychological and neurological areas. It offers a general agent structure, with an open component implementation in order to be applied in a wide range of domains. It's integration into a typical BDI architecture allows to combine practical rational elements with more "human" features, that results in believable behaviors for the agents. At the same time it offers an open structure in order to be flexible enough to adapt to several domains and applications. A proposal of design for each component of the architecture is also offered. This approach has many practical uses in human-computer interaction applications like education, pathologies treatments, training, entertainment and human simulation behavior in general. Nevertheless, the main challenge after roughly define how the agent reasoning processes are integrated into the architecture, is the detailed specification of each one of the particular components and how to achieve a coherent behavior aligned with real situations.

This approach is a work in progress. We are currently engaged on implementing a practical application in order to test the strength of the proposal as well as to identify improvements and/or necessary modifications. Specifically it is been tested in classical games of experimental economy where humans take

[2]Note that lines 22-28 are kept from the original algorithm in (Bordini et al., 2007) where the authors point out that reconsideration is performed only if this leads to a change of intentions and a replanning is performed if the plan is not sound any more according to what the agent wants to achieve (intentions) and it thinks is the current state of the world (beliefs).

[3]The intention in this step is above all to keep the necessary records associated to the results of current results of plans executed and/or achievement of goals in order to reuse this information as a "memory" for further deliberations.

decisions that are often biased from the most 'rational' solution.

ACKNOWLEDGMENTS

This work was supported by the Spanish government grant MINECO/FEDER TIN2012-36586-C03-01 and HUMBACE (Human and Social Behaviour 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

Becker-Asano, C. and Wachsmuth, I. (2010). Affective computing with primary and secondary emotions in a virtual human. *AAMAS '10*, 20(1):32–49.

Bordini, R. H., Hübner, J. F., and Wooldridge, M. (2007). *Programming multi-agent systems in AgentSpeak using Jason*. Wiley.

Castelfranchi, C. (2000). Affective appraisal versus cognitive evaluation in social emotions and interactions. In *Affective Interactions*, pages 76–106. Springer.

Damásio, A. R. (2005). *Descartes' Error: Emotion, Reason, and the Human Brain*. Penguin Group US.

Dastani, M. and Meyer, J.-J. C. (2006). Programming agents with emotions. In *Proceedings of the 2006 conference on ECAI 2006: 17th European Conference on Artificial Intelligence August 29 – September 1, 2006, Riva del Garda, Italy*, pages 215–219. IOS Press.

Ekman, P. (1999). *Basic Emotions*, pages 45–60. John Wiley & Sons, Ltd.

Frijda, N. H. (1987). *The Emotions*. Studies in Emotion and Social Interaction. Cambridge University Press.

Gebhard, P. (2005). ALMA: a layered model of affect. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 29–36, New York, USA. ACM.

Goldberg, L. R. et al. (1990). An alternative "description of personality": The big-five factor structure. *Journal of personality and Social Psychology*, 59(6):1216–1229.

Jiang, H., Vidal, J. M., and Huhns, M. N. (2007). EBDI: An architecture for emotional agents. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 11:1–11:3. ACM.

John, O. P. and Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In Pervin, L. A. and John, O. P., editors, *Handbook of Personality: Theory and Research*, pages 102–138. Guilford Press, second edition.

LeDoux, J. E. (1998). *The emotional brain: the mysterious underpinnings of emotional life*. A Touchstone book. Simon & Schuster.

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

Marsella, S. C., Gratch, J., and Petta, P. (2010). Computational models of emotion. In *A Blueprint for Affective Computing: A Sourcebook and Manual*, Affective Sciene, chapter 1.2. OUP Oxford.

McCrae, R. R. and John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215.

Mehrabian, A. (1996). Analysis of the big-five personality factors in terms of the pad temperament model. *Australian Journal of Psychology*, 48(2):86–92.

Mehrabian, A. (1997). Comparison of the PAD and panas as models for describing emotions and for differentiating anxiety from depression. *Journal of Psychopathology and Behavioral Assessment*, 19(4):331–357.

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

Neto, A. and Silva, F. (2012). A computer architecture for intelligent agents with personality and emotions. In Zacarias, M. and Oliveira, J. V., editors, *Human-Computer Interaction: The Agency Perspective*, volume 396 of *Studies in Computational Intelligence*, pages 263–285. Springer Berlin Heidelberg.

Niedenthal, P. M. and Setterlund, M. B. (1994). Emotion congruence in perception. *Personality and Social Psychology Bulletin*, 20(4):401–411.

Ortony, A. (2003). On Making Believable Emotional Agents Believable. In Trapple, R. P., Petta, P., and Payer, S., editors, *Emotions in Humans and Artifacts*, chapter 6, pages 189–212. MIT Press.

Ortony, A., Clore, G. L., and Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.

Ortony, A. and Turner, T. J. (1990). What's basic about basic emotions? *Psychological review*, 97(3):315.

Parunak, H. V. D., Bisson, R., Brueckner, S., Matthews, R., and Sauter, J. (2006). A model of emotions for situated agents. In *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, AAMAS '06, pages 993–995, New York, USA. ACM.

Picard, R. W. (1997). *Affective computing*. MIT Press, Cambridge, MA, USA.

Russell, J. A. and Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of research in Personality*, 11(3):273–294.

Ryckman, R. M. (2007). *Theories of Personality*. PSY 235 Theories of Personality Series. Thomson/Wadsworth.

Steunebrink, B. R., Dastani, M., and Meyer, J.-J. C. (2012). A formal model of emotion triggers: an approach for BDI agents. *Synthese*, 185:83–129.

Weiss, G. (1999). *Multiagent systems: a modern approach to distributed artificial intelligence*. MIT press.