Social Human-Robot Interaction: A New Cognitive and Affective Interaction-Oriented Architecture

Carole Adam, Wafa Johal, Damien Pellier$^{(\boxtimes)}$, Humbert Fiorino, and Sylvie Pesty

Univ. Grenoble-Alps, LIG, 38000 Grenoble, France
{carole.adam,wafa.johal,damien.pellier,humbert.fiorino,
sylvie.pesty}@imag.fr

Abstract. In this paper, we present CAIO, a Cognitive and Affective Interaction-Oriented architecture for social human-robot interactions (HRI), allowing robots to reason on mental states (including emotions), and to act physically, emotionally and verbally. We also present a short scenario and implementation on a Nao robot.

1 Introduction

Robots are more and more present in daily life, in roles such as assistive robots, pedagogical robots, companion robots for children or for the elderly, etc., where they must have a closer interaction with their user. By close, we mean that robots must share not only the same physical space but also goals and beliefs to achieve a common task through their interactions. They should also interact intuitively and easily through speech, gestures, and facial expressions. In spite of the numerous contributions in the field of cognitive architectures for agents and for robots (*e.g.* [5,10,15,16]), designing a cognitive architecture dealing with the complexity of human-robot interactions (HRI) remains a real challenge.

In this paper, we present a new architecture: CAIO (Cognitive and Affective Interaction-Oriented architecture) for social Human-Robot Interaction that aims to contribute on the following aspects essential to HRI: managing emotions (non-verbal aspects of interaction), sensorimotor and deliberative levels (fast (emotional) answer versus slower and more deliberate answer), explicit manipulation of mental states (to enable self-explanation) and handling both physical and verbal actions.

2 Related Works

Cognitive architectures have been subject to research for a long time, and good reviews exist (see for example [11,30]). They mostly fall in three categories: biologically-inspired, philosophically-inspired, and Artificial Intelligence architectures. We illustrate these categories with some of the major and well-known architectures.

Biologically inspired architectures: ACT-R is a well-known cognitive architecture (Adaptive Control of Thought-Rational), stemming from the progressive refinement of Anderson's model of human cognition [5], originating in his Human Associative Memory model [6]. The main assumption is the separation between two types of knowledge: declarative (chunks) and procedural (rules); the system is only aware of knowledge with sufficient activation.

CLARION (Connectionist Learning with Adaptive Rule Induction ON-line) [29] is also well-known, based on neural networks. It mainly focuses on the distinction between implicit and explicit processes, and the interactions between them. It has been used to simulate processes in cognitive or social psychology, and to implement AI applications.

ASMO (Attentive Self-MOdifying) [19] was developed more recently based on a biological theory of attention, to solve the problem of competing, possibly incompatible robot goals. Concretely the attention level determines relative priorities of goals, with the most critical ones being treated as reflexes. It is being implemented in a social bear robot interacting with humans, and in Nao for soccer competitions.

Problem-Solving Artificial intelligence architectures: SOAR (State, Operator And Result) [15] is a pure AI symbolic architecture focused on learning and problem solving. It has short-term working memory and long-term memory (procedural, semantic and episodic). Reinforcement learning is triggered when knowledge is inadequate to make a decision. SOAR was extended with emotions that affect learning [14].

ICARUS [17] is grounded in cognitive psychology and AI, and aims at unifying reactive and deliberative problem-solving, as well as symbolic and numeric reasoning. The goal with highest priority that is not satisfied yet takes focus: the skills allowing to achieve it are brought from long-term to short-term memory, or means-end analysis is used to decompose it into subgoals and learn new skills.

Philosophically inspired architectures: Bratman's philosophical action theory [9] models human behaviour as a perception-decision-action cycle. He claimed that the intention to perform an action is adopted from beliefs and desires *via practical reasoning* that makes us rational. BDI logics (*Belief, Desire, Intention*, [12,24]) were then proposed to formalise these three mental states.

The BDI model has been at the root of a number of architectures for artificial agents (*e.g.* the *Procedural Reasoning System* - PRS [33]). As we will explain later, the CAIO architecture is also in line with this tradition but we introduce new mental states, in particular emotions that are essential for an expressive social robot since they play a major role in interaction and reasoning.

3   Previous Work

3.1   Complex Emotions and Multimodal Conversational Language

Guiraud et al. [13] proposed a new modal logic (BIGRE logic) derived from BDI for the formal representation of five agent's mental states (B, I, G, R, E) expressed by an agent during a conversation with another agent or a human:

– *Belief* (B) $Bel_i\varphi$: the robot i believes that φ,
– *Ideal* (I) $Ideal_i\varphi$: ideally for robot i, φ should hold (social and moral norms of the robot[1]),
– *Goal* (G) $Goal_i\varphi$: the robot i wants that φ holds,
– *Responsibility* (R) $Resp_i\varphi$: the robot i is responsible for φ (arising from complex reasoning about norms and responsibility of its own actions and those of others).
– *Complex emotion* (E) (*e.g.* gratitude, admiration, reproach, etc.) result from reasoning on *Responsibility*.

Complex emotions are of primary importance in human dialogue and are mainly conveyed through language. They differ from basic emotions built from beliefs and goals, and often expressed by prototypical facial expressions. Eight complex emotions (*regret, disappointment, guilt, reproach, moral satisfaction, admiration, rejoicing* and *gratitude*) and four basic emotions (*joy, sadness, approval, disapproval*) have been formalized in terms of the B, I, G and R operators (BIGR → E).

To ensure that a robot is able to express its mental states in a credible manner, Riviére *et al.* [26] defined a conversational language based on Searle's Speech Acts Theory [28], and in line with previous mentalistic Agent-Communication Languages (ACL) such as FIPA [23]. This language is called *Multimodal Conversational Language (MCL)* because it closely links verbal (the utterance) and non-verbal (*e.g.* underlying emotion of expressive speech acts) aspects in order to improve the expressivity of the robot. The MCL consists of 38 *Multimodal Conversational Acts* (MCA) divided in four classes:

– assertive acts: *to inform, to affirm, to deny*, etc.
– directive acts: *to ask, to suggest, to require*, etc.
– commissive acts: *to promise, to accept, to offer*, etc.
– expressive acts: *to apologize, to rejoice, to reproach, to thank*, etc.

For each MCA there is a formalisation in the BIGRE logic of:

– its preconditions, that the robot has to satisfy before performing this act, ensuring its sincerity in the sense of Searle's Speech Acts Theory (sincerity conditions);
– its sending effects, on the robot performing it;
– its reception effects, on the robot receiving this act performed by the interlocutor.

For example, Table 1 shows the formalisation of the conversational act *to rejoice*. This explicit formal representation has the advantage of allowing the robot to manipulate and reason about the conversational acts: update its mental states when receiving or sending one, and using them in its plan of action.

The interested reader is referred to [13] for detailed semantics and axiomatics of the BIGRE logic.

---

[1] For instance, a moral obligation to help someone in danger, or a social norm to pay one's taxes, etc.

Table 1. Example: the *to rejoice* MCA from agent *a*'s point of view in a dialogue with a human *h*.

Rejoice	$Exp_{a,h,H}(Rejoicing_a\varphi) \equiv Exp_{a,h,H}(Goal_a\varphi \wedge Bel_aResp_a\varphi)$
Preconditions	$Goal_a\varphi \wedge Bel_aResp_a\varphi \overset{d\acute{e}f}{=} Rejoicing_a\varphi$
	Robot *a* "feels" rejoicing; it believes it is responsible for having achieved its goal φ
Sending effects	$Bel_aBel_hRejoicing_a\varphi$
	Robot *a* believes that human *h* believes that *a* is rejoicing about φ
Reception effects	$Bel_aGoal_h\varphi \wedge Bel_aBel_hResp_h\varphi$
	Robot *a* believes that human *h* expressed his rejoicing about φ. Therefore, *a* believes that *h* has achieved its goal φ and believes himself to be responsible for this

3.2 PLEIAD Reasoning Engine

PLEIAD (ProLog Emotional Intelligent Agent Designer) [1] is originally a SWI-Prolog reasoning engine for BDI-like agents. It provides agents with generic reasoning capabilities and emotions. Concretely, it enables the implementation of various logical models of emotions, such as the OCC theory [2] or theories about shame [3]. It has also been recently extended with coping strategies [4] and some personality traits.

This reasoning engine has been used to implement the BIGRE logical model and is at the core of the CAIO architecture, especially for the *Deliberation module* and the *Emotional appraisal module* (see details in Sect. 4.2).

4 CAIO Architecture

4.1 Overview of the Architecture

The CAIO architecture (see Fig. 1) consists of two fundamental loops: a **Deliberative loop** used to reason on BIGRE mental states and produce plans of actions, and a **Sensorimotor loop** to immediately and continuously trigger emotion expressions. Each loop takes as inputs the result of the multimodal perception of the environment.

During the **Deliberative loop**: the *Cognitive part of the Emotional Appraisal module* deduces complex emotions from the mental states; the *Deliberation module* deduces the robot's *Communicative Intentions* from its mental states, and selects the most appropriate one; then the *Planning module* produces a plan to achieve the selected intention (*i.e.* a set of ordered actions, conversational acts and/or physical actions), and schedules the robot's next action; finally the *Emotional Multimodal Action Renderer module* executes this scheduled action. The modules can provide feedback to each other: the planning module informs the deliberation module of the feasibility of the selected intention; the renderer informs the planner of the success or failure of action performance.
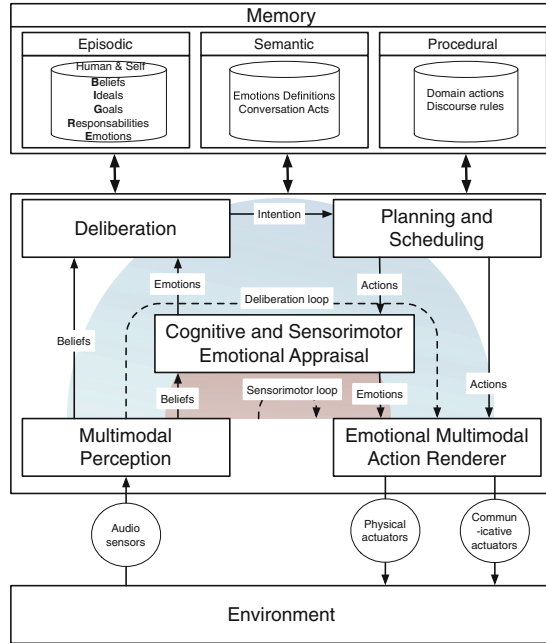
Fig. 1. The CAIO architecture.

Simultaneously, during the **Sensorimotor loop**: the *Sensorimotor part of the Emotional Appraisal module* evaluates the input according to criteria (Scherer's SEC - Stimulus Evaluation Checks); the *Emotional Multimodal Action Renderer module* then dynamically renders the corresponding robot's non-verbal (facial and gestural) expression.

4.2   The 6 Main Modules of the CAIO Architecture

Memory module. The robot's memory is divided into three parts in accordance with the state of the art (see Sect. 2). The *episodic* memory contains BIGRE-based knowledge about the self and the human interlocutor. The *semantic* memory contains the definitions of emotions and conversational acts. The *procedural* memory deals with domain actions (how-to) and discourse rules (*i.e.* when asked a question, one should reply).

The memory is dynamically updated in three steps: first, new beliefs deduced from the perception of the world or of the interaction (reception effects of the user's recognised MCA, and sending effects of the robot's own MCA) are *added*; then inference rules are applied to *update* the robot's BIGRE mental states, possibly deducing new mental states.

Multimodal perception module. We focus on language perception. Concretely, this module first recognises text from speech (using Google Speech). It

then extracts the human's MCA from the recognised text utterance[2]. This MCA then generates new beliefs (its reception effects) that enter the 2 (sensorimotor and deliberative) loops of processing. As the aim of this module is to merge multimodal inputs to generate new beliefs on the user's mental states, future works will consider facial expression and para-linguistic signals.

Appraisal module. The appraisal module is in two parts. The cognitive part is an extension of PLEIAD and takes as input the robot's perceptions and mental states to trigger the corresponding emotions from their logical definition in terms of mental states. For example, the emotion of gratitude is triggered when a robot has the goal φ and believes that the human is responsible for φ, *i.e.* when the robot i's has a mental states $Goal_i\varphi \wedge Bel_i Resp_j\varphi$). The emotion intensity is derived from the priority of the goal or the ideal included in its definition.

The sensorimotor part assesses all MCA perceived or sent by the robot *w.r.t.* Scherer's *Stimulus Evaluation Checks* (SEC, [27]) (Novelty, Intrinsic pleasantness, Goal/Need conductivness, Coping, Norm). The results of the SEC evaluation process are then sent out to the renderer for their facial and bodily expression. Figure 2 shows an example of a SEC sequence corresponding to a reproach expression.

Novelty	Intrinsic Pleasantness	Goal/Need Conductivness	Coping (Control/Power)	Norm Compatibility
Novel	Unpleasant	Unmet	Low Control	Norm violated
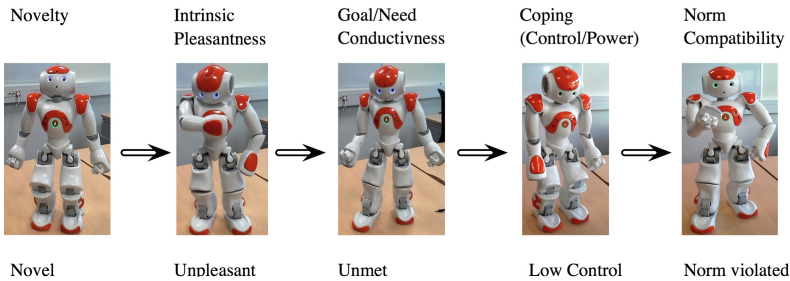
Fig. 2. SEC sequence corresponding to a reproach expression.

Deliberation module. Deliberation is the process of selecting the robot's next intention to achieve, *via* practical reasoning [9] from its mental states and a set of priority rules. It deals in particular with three kinds of *intention. Emotional intentions* are intentions to express the robot's emotions. In order for the robot to be sincere, affective and expressive, we assume that all emotions felt during the interaction lead to an emotional intention to express them, which participate in the local regulation of dialogue by enabling a more natural robot-human interaction [8]. *Obligation-based intentions* also contribute to the local regulation of dialogue [7]. They are adopted from a set of *discourse obligation rules* defined

---

[2] Natural Language Understanding is a complex research field of its own, we do not tackle this problem here, and instead use an ad-hoc grammar specifically designed for our scenario.

by Traum and Allen [31] to represent social norms guiding the robot's behaviour and making it reactive at the discourse level. Concretely the robot always adopts the intention to fulfill its obligation deduced by these rules. Finally, the *global intention* gives the global direction of dialogue and defines its type (*e.g.* deliberation, persuasion... [32]). It is adopted when the robot has committed to achieve the corresponding goal, either publicly (by performing a commissive MCA such as *Promise* or *Accept*) or privately (via practical reasoning on its beliefs and plans).

Planning module. It is in charge of finding a way of achieving the selected intention according to a plan-based approach of dialogue [22]. It is based on the planning approach proposed by [21] and on the PDDL4J Java library [20]. The plans produced contain MCA and/or physical actions, whose preconditions and effects are formalised in the classical Planning Domain Description Language (PDDL), making most existing planners compatible with CAIO.

In the case of emotional and obligation-based intentions, the built plan is usually made up of a single MCA (for example the emotional intention to express gratitude can be achieved with *to thank* or *to congratulate* depending on the emotion's intensity). In the case of global intentions, domain-dependent actions may be necessary, whose preconditions and effects are described in the static procedural memory (for instance, to book a train it is necessary to know the time and date of departure and destination). The planner will then produced a plan with both MCA (*e.g. to ask* the relevant information to the user) and domain actions (*e.g.* actually book the train). If no plan can be computed to achieve it, the current intention is discarded, and feedback is sent to the deliberation module that selects a new intention.

Multimodal Action Renderer. It receives as input the action to be executed and the complex emotion computed by the appraisal module, and controls the robot's actuators to execute this action and to dynamically generate the facial expression corresponding to the emotion. In particular for MCA it expresses the underlying complex emotion. Independently from this deliberative expression, this module also receives the SEC values computed by the sensorimotor part of the emotional appraisal module, and dynamically builds the corresponding facial and bodily expression, leading to a sequence of postures (for example Table 1).

5   Discussion and Conclusion

The CAIO architecture was first implemented and evaluated for a virtual character [25]. The current version is based on ROS (Robot Operating System), which is largely used in the robotic community. The modules were implemented in Python in order to allow easy interfacing with SWI-Prolog (PLEIAD engine) via the Pyswip library (a Python library that allows to query SWI-Prolog from Python programs). Below is a short scenario illustrating the ROS nodes encapsulating each process involved in the CAIO architecture (see its UML Sequence Diagram on Fig. 3).
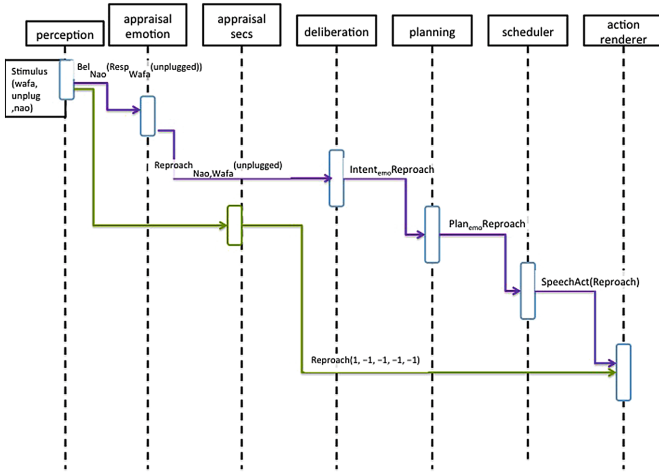
Fig. 3. UML Sequence Diagram

This scenario involves two actors: the human, *Wafa*, and the *Nao robot*, which has a low battery life and requires to be plugged all the time. In its episodic memory, Nao has the ideal of being plugged $Ideal_{Nao}(\neg unplugged)$. Now, Wafa has a party tonight and needs to dry her hair, but Nao is plugged to the only plug near the mirror. She thus tells Nao: "Nao, I am unpluging you".

1. **Multimodal Perception:**
   The *perception* node receives the utterance "Nao, I am unpluging you" and extracts an *to inform* conversational act (i.e. a *to inform* MCA) $Inform(Stimulus(unplugged, wafa, nao))$. The episodic memory is updated: $Bel_{Nao}(unplugged)$ and $Resp_{Wafa}(unplugged)$.

2. **Emotional Appraisal:**

   (a) **Cognitive Appraisal:** The *appraisal_emotion* node deduces the complex emotion $Reproach_{Nao, Wafa}(unplugged)$; the episodic memory is updated. The complex emotion is sent to *deliberative* node.

   (b) **Sensorimotor Appraisal:** The *appraisal_checks* node evaluates the *to inform* MCA $Inform(Stimulus(unplugged, wafa, nao))$ in accordance with the five evaluation criteria (SEC). The result of the SEC sequence is sent to the *action_renderer* node.

3. **Deliberation:**
   The *deliberative* node infers a list of intentions. The one with highest weight is an *emotional intention* to perform a reproach to the user. The list of intentions is sent to the *planning* node.

4. **Planning and Scheduling:**
   The *planning* node picks the most weighted intention from the list of intentions, and publishes a list of plans (here a unique plan consisting of the single *to reproach* MCA).

The plan is received by the *scheduler* node which picks the first action (here the *to reproach* MCA).

5. **Emotional Multimodal Action Renderer:**
   Finally the *action_renderer* node receives both the SEC sequence and the *to reproach* MCA, and plays them on the Nao robot. It thus tells Wafa: "Wafa, you must not unplug me". Nao performs this utterance in a multimodal way.

The ROS version of the CAIO architecture is currently being further validated through real-time interaction with children to verify that the robot clearly conveys its intentions, and is perceived as sincere. In parallel, we have run a more conceptual evaluation of CAIO against Langley *et al.* evaluation criteria for cognitive architectures [18]. This conceptual evaluation shows that the CAIO architecture already provides new contributions regarding the state of the art in cognitive architectures for companion robots.

Further research on the multimodal perception module is however needed to automate the extraction of the user's speech act, and to deal with facial expressions and para-linguistic features (to guarantee better recognition and sincerity). A learning module would also be a nice extension to ensure that the robot can improve during the interaction, and progressively learn to know its user to better adapt to them and engage them.

References

1. Adam, C.: The emotions: from psychological theories to logical formalisation and implementation in a BDI agent. Ph.D. thesis, INP Toulouse, July 2007
2. Adam, C., Herzig, A., Longin, D.: A logical formalization of the OCC theory of emotions. Synthese **168**(2), 201–248 (2009)
3. Adam, C., Longin, D.: Shame, when reasoning and emotions are linked. In: EUMAS 2013 (2013)
4. Adam, C., Lorini, E.: A BDI emotional reasoning engine for an artificial companion. In: Corchado, J.M., Bajo, J., Kozlak, J., Pawlewski, P., Molina, J.M., Gaudou, B., Julian, V., Unland, R., Lopes, F., Hallenborg, K., García Teodoro, P. (eds.) PAAMS 2014. CCIS, vol. 430, pp. 66–78. Springer, Heidelberg (2014). doi:10.1007/978-3-319-07767-3_7
5. Anderson, J.R.: Human symbol manipulation within an integrated cognitive architecture. Cogn. Sci. **29**(3), 313–341 (2005)
6. Anderson, J.R., Bower, G.H.: Human Associative Memory. Winston and Sons, Washington, DC (1973)
7. Baker, M.J.: A model for negotiation in teaching-learning dialogues. J. Artif. Intell. Educ. **5**(2), 199–254 (1994)
8. Bates, J.: The role of emotion in believable agents. Commun. ACM **37**(7), 122–125 (1994)
9. Bratman, M.E.: Intention, Plans, and Practical Reason. Harvard Univ. Press, Cambridge (1987)
10. Brooks, R.: A robust layered control system for a mobile robot. IEEE J. Robot. Autom. **2**(1), 14–23 (1986)
11. Chong, H.Q., Tan, A.H., Ng, G.W.: Integrated cognitive architectures: a survey. Artif. Intell. Rev. **28**(2), 103–130 (2009)

12. Cohen, P.R., Levesque, H.J.: Persistence, intention and commitment. In: Cohen, P., Morgan, J., Pollack, M. (eds.) Intentions in Communication, pp. 33–69. MIT Press, Cambridge (1990)

13. Guiraud, J., Longin, D., Lorini, E., Pesty, S., Rivière, J.: The face of emotions: a logical formalization of expressive acts. In: International Conference on Autonomous Agent and Multiagent Systems, pp. 1031–1038 (2011)

14. Hogewoning, E., Broekens, J., Eggermont, J., Bovenkamp, E.G.P.: Strategies for affect-controlled action-selection in soar-RL. In: Mira, J., Álvarez, J.R. (eds.) IWINAC 2007. LNCS, vol. 4528, pp. 501–510. Springer, Heidelberg (2007). doi:10.1007/978-3-540-73055-2_52

15. Laird, J.E.: The Soar Cognitive Architecture. MIT Press, Cambridge (2012)

16. Lallée, S., Pattacini, U., Lemaignan, S., Lenz, A., Melhuish, C., Natale, L., Skachek, S., Hamann, K., Steinwender, J., Sisbot, E., Metta, G., Guitton, J., Alami, R., Warnier, M., Pipe, T., Warneken, F., Dominey, P.: Towards a platform-independent cooperative human robot interaction system: III an architecture for learning and executing actions and shared plans. IEEE Tran. Auton. Ment. Dev. **4**(3), 239–253 (2012)

17. Langley, P., Choi, D., Rogers, S.: Interleaving learning, problem solving, and execution in the icarus architecture. Technical report, Computational Learning Laboratory, CSLI, Stanford University (2005)

18. Langley, P., Laird, J., Rogers, S.: Cognitive architectures: research issues and challenges. Cogn. Syst. Res. **10**(2), 141–160 (2009)

19. Novianto, R., Johnston, B., Williams, M.A.: Attention in asmo cognitive architecture. In: 1st International Conference on Biologically Inspired Cognitive Architectures (BICA) (2010)

20. Pellier, D.: PDDL4J: Planning Domain Description Language Library for Java (2015)

21. Pellier, D., Fiorino, H., Métivier, M.: Planning when goals change: a moving target search approach. In: International Conference on Advances in Practical Applications of Heterogeneous Multi-Agent Systems (2014)

22. Perrault, C.R., Allen, J.F.: A plan-based analysis of indirect speech acts. Comput. Linguist. **6**(3–4), 167–182 (1980)

23. Poslad, S.: Specifying protocols for multi-agent systems interaction. ACM Trans. Auton. Adapt. Syst. **2**(4) (2007). http://doi.acm.org/10.1145/1293731.1293735

24. Rao, A.S., Georgeff, M.P.: Modeling rational agents within a BDI-architecture. In: International Conference on Principles of Knowledge Representation and Reasoning (KR 1991), pp. 473–484 (1991)

25. Rivière, J.: Interaction affective et expressive Compagnon artificiel - humain. Ph.D. thesis, de Grenoble (2012)

26. Riviere, J., Adam, C., Pesty, S., Pelachaud, C., Guiraud, N., Longin, D., Lorini, E.: Expressive multimodal conversational acts for SAIBA agents. In: Vilhjálmsson, H.H., Kopp, S., Marsella, S., Thórisson, K.R. (eds.) IVA 2011. LNCS (LNAI), vol. 6895, pp. 316–323. Springer, Heidelberg (2011). doi:10.1007/978-3-642-23974-8_34

27. Scherer, K.R.: Appraisal considered as a process of multilevel sequential checking. In: Scherer, K.R., Schorr, A., Johnstone, T. (eds.) Appraisal Processes in Emotion: Theory, Methods, Research, pp. 92–120. Oxford University Press, New York (2001)

28. Searle, J., Vanderveken, D.: Foundations of Illocutionary Logic. Cambridge University Press, Cambridge (1985)

29. Sun, R.: Duality of the Mind: A Bottom-up Approach Toward Cognition. Lawrence Erlbaum Associates, Mahwah (2002)

30. Thórisson, K., Helgasson, H.: Cognitive architectures and autonomy: a comparative review. J. Artif. Gen. Intell. **3**(2), 1–30 (2012)
31. Traum, D.R., Allen, J.F.: Discourse obligations in dialogue processing. In: Proceedings of the 32th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 1–8 (1994)
32. Walton, D., Krabbe, E.: Commitment in Dialogue: Basic Concept of Interpersonal Reasoning. State University of New York Press, Albany (1995)
33. Wooldridge, M.: An Introduction to MultiAgent Systems. Wiley, New York (2009)