Biol Cybern
DOI 10.1007/s00422-008-0254-9

ORIGINAL PAPER

Biological Cybernetics

Interacting with an artificial partner: modeling the role of emotional aspects

Isabella Cattinelli · Massimiliano Goldwurm · N. Alberto Borghese

Received: 26 September 2007 / Accepted: 8 August 2008

Abstract In this paper we introduce a simple model based on probabilistic finite state automata to describe an emotional interaction between a robot and a human user, or between simulated agents. Based on the agent's *personality*, *attitude*, and *nature*, and on the emotional inputs it receives, the model will determine the next emotional state displayed by the agent itself. The probabilistic and time-varying nature of the model yields rich and dynamic interactions, and an autonomous adaptation to the interlocutor. In addition, a reinforcement learning technique is applied to have one agent drive its partner's behavior toward desired states. The model may also be used as a tool for behavior analysis, by extracting high probability patterns of interaction and by resorting to the ergodic properties of Markov chains.

1 Introduction

Even though machines are increasingly spreading in every sector of our society, becoming indispensable tools able to solve everyday tasks, there are still many typical human abilities which cannot be reproduced by electronic devices: on the one hand human higher cognitive functions, such as language production and understanding, and, on the other hand, *emotional* functions: recognizing other people's emotions, reacting emotionally to situations, establishing relationships with an emotional content and so forth. The introduction of emotional components in computers can appear, at first, as pointless: the machine must be intelligent, not be able to feel emotions. But, if we aim to approach the ideal model of human intelligence, the emotional component cannot be ignored. Emotions are thought to be part of our decisional processes (Damasio 1994), drive our learning, help self-preservation (fear for a dangerous phenomenon makes us move away from it), and are at the basis of human relationships. *Emotional intelligence* is an important part of all our intellective faculties.

The recently arisen effort in designing *emotional machines* which could understand, analyze, and synthesize emotions, derives from the acknowledged importance of emotions in human life. In the 1990s a new interdisciplinary research field (collecting contributions from computer science, neuroscience, psychology, sociology, and so forth) was proposed: *affective Computing*, defined as "*computing that relates to, arises from, or deliberately influences emotions*" (Picard 1997). The research field is, therefore, wide, but we can point out the following main themes:

– Implementation of modules for human emotion recognition, based on physiological parameters (heart-beat rate, skin conductance, respiration, etc.) (Picard et al. 2001) or on *non-verbal communication* (Argyle 1975) (facial expressions (Anderson and McOwan 2006), posture (Silva and Bianchi-Berthouze 2004), gestures (Silva et al. 2006), voice tone (Ciota 2005)).
– Design of systems for simulating emotional states, which could communicate emotions readable by the human user (emotional avatars (Fabri et al. 1999; Lester et al. 2000)).

An early stage part of this work was presented at the 11th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems (KES 2007).

I. Cattinelli · M. Goldwurm · N. A. Borghese (✉)
Applied Intelligent Systems Laboratory (AIS-Lab),
Department of Computer Science, Università degli Studi di Milano,
Via Comelico 39/41, 20135 Milan, Italy
e-mail: borghese@dsi.unimi.it

I. Cattinelli
e-mail: cattinelli@dsi.unimi.it

M. Goldwurm
e-mail: goldwurm@dsi.unimi.it

– Attempts at modeling emotional dynamics, to explain in formal terms how human emotional intelligence works and to reproduce this faculty in machines (Doshi and Gmytrasiewicz 2004).

Combining modules for emotion recognition and production, emotional machines would be obtained, being able to interact with users not only by a limited set of standard commands, but also using emotion exchange as a more direct and natural communication channel. As a consequence, human-machine interaction would result easier and more effective, because it would be based on similar mechanisms as human-human relationships. Much work still remains to be done, though, since results in human emotion recognition are still quite unsatisfactory, due to the complexity of the task itself, and more generally to the difficulty in defining precisely what an emotion truly is. The same applies, of course, to emotion modeling.

In fact, emotions are still largely "ununderstood", and different views have been proposed. According to (Scherer 1984), for instance, emotions are complex processes of which the external display is just one (and not necessarily the most important) of the components. Indeed, emotional expressions can be faked (this is the case for actors) or masked at some degree, so that they are not always valid cues for inferring someone's inner emotional state. However, external expressions are generally the only information we have on someone's emotional state, when interacting with them: blood pressure or skin conductance could be more faithful predictors of the actual experienced emotion, but measuring such data during the interaction would make it far less natural. Moreover, talking about emotions when referring to non-living entities like robots can be misleading, since the physiological changes naturally associated with the rise of an emotion cannot be reproduced (and this is particularly true for non-embodied agents); this is not of secondary importance, if we think that one traditional emotion theory (James 1884) even states that bodily changes *cause* emotions to arise. Thus, in artificial intelligence or robotics emotional states are considered to be basically abstract, properly labeled (e.g. *happiness_state*) structures, upon which a set of behavioral responses is built (e.g. Kuhnlenz and Buss 2004). Being our work focused on modeling emotional interaction in a robotic context, rather than human emotions per se, in what follows we will use the terms *emotion* and *emotional state* without pretense to address the whole complexity of these phenomena. Rather, we will focus on exterior emotional expressions and on general emotional categories which can be associated to them (in a very common-sense approach: "he is smiling"—facial expression—"then he must be happy"—inference about emotional state).

This work is, therefore, focused on analyzing and synthesizing emotional *behaviors*, rather than human emotions as complex psychological processes. In this sense, related work can be traced in the wider field of behavior robotics. The work by (Chernova and Arkin 2007) proposes a model for behavior selection in a QRIO robot, based on the robot's internal state and external inputs. To this purpose an activation level (AL) is associated to each behavior, defined as a weighted sum of four components. These describe, respectively, the robot's motivation and expected satisfaction, and the baseline activation and self-excitation associated to that behavior, plus a random noise parameter which adds variability in behavior selection. The AL formulation includes also a basis for robot's *personality* by means of a pair of weights, which can be set to facilitate self-centered or extroverted behaviors. Furthermore, the past history of performed behaviors is considered as a basis for autonomously learning routine sequential tasks. In (Inamura et al. 2004) robots learn a set of measurement-action pairs in an initially unknown environment; for instance, in an obstacle avoidance task, an adequate movement — right, left, forward — is found depending on the distance value reported by sensors. The measurement-action pairs are represented in a conditional probability table (CPT), which is updated through the interaction with the environment with the help of the information given by the user.

These works are mainly focused on action selection tasks, possibly helped by the interaction with users, rather than on human-robot interaction as such. The problem of modeling an emotional interaction which evolves through time may be better addressed through the use of a finite state automaton (FSA) (see, for instance, Hopcroft and Ullman 1979). This model consists of states (which represent *emotional states*, such as happiness, or anger), inputs (events or information coming from the outside that are able to modify the emotional state) and a transition function, which describes the rules which transform the current state and the current input into a next state. Moreover, a "personality" of the agent could be defined and associated to the transition function, making agents with different personalities respond differently to the same stimulus. In fact, modeling emotional interactions requires taking into account individual variability, i.e. differences in characters and personalities which can affect the outcome of an interaction. Each individual has characteristic traits that should somehow be modeled in order to describe a likely interaction.

Since deterministic FSAs tend to produce stereotyped behaviors, a stochastic version of FSA, termed probabilistic finite state automaton, PFSA, has been recently introduced in emotional interaction modeling. In a PFSA the transition function is stochastic: that is, given the current state and the current input, there are many possible next states, each entered with a given probability (Rabin 1963; Paz 1971). Indeed, the introduction of a stochastic component in an emotional interaction model leaves space for unexpected behavior

(Chittaro and Serra 2004; Kopecek 2003; Kuhnlenz and Buss 2004; Nomura 1996): we do not expect that our interlocutor will always react the same way to the same situation, even if her personality remains unchanged.

In (Chittaro and Serra 2004) an agent has a goal which can be accomplished by selecting different sequences of actions. Which action to perform is decided based on the agent's personality that, in turn, determines the probability of each action: this selection process is represented as a PFSA. In (Kopecek 2003) a PFSA is used as a personality model to describe the dynamics of a dialogue. Here, the entire automaton (that is, the ensemble of its inputs, outputs, states, and transition function) is referred to as the personality of the agent. By analyzing the dialogues, the possible automaton which generated them is searched. Both these models are prevalently static: that is, transition probabilities do not change according to the history of the interaction.

Another stochastic model, proposed as an emotional core for a robot, is a hidden Markov model (HMM), equipped in addition with input control (Kuhnlenz and Buss 2004). Here, emotional states are the HMM's hidden states, to which different observable expressions are associated with different probabilities. Inner emotional dynamics is determined by a matrix of transition probabilities (not depending upon external stimuli), whose entries can be defined in order to design different robot personalities. The impact of different inputs on state transitions is coded into another matrix, which models emotional response to external events (perceptive information coming from sensors), as well as internal ones (produced by cognitive processes). Moreover, a forgetting filter is introduced which keeps a progressively decaying trace of past inputs, so that the probability for a state transition at a given time step is dependent not only on the events occurring at that time but also on past events, whose contributions are weighted decreasingly with the passing of time.

An interesting model for producing inter-individual relationships through conversations between individuals endowed with emotions and personality is proposed in (Nomura 1996). Here, each individual is represented as an automaton, where inputs and outputs are actions (e.g. cooperation, or disregard), states are emotional states and personality is a parameter which determines the probability of each output as a function of the current emotional state. An additional parameter, the attitude, is also introduced, which modifies the probability of each next state, depending on the current input-output pair. So, both the transition and the output functions are parameterized, by attitude and personality, respectively. Attitude is a time-varying parameter: it is subject to updates based on the emotional state and the personality of the individual. While we share, as better explained in Sect. 2, the same keywords of attitude and personality, their meaning in (Nomura 1996) and in our work is quite different (for a discussion, see Sect. 7).

In the following, we propose a novel, more complete model for emotional interaction between a human and an agent, or between two simulated agents, based on PFSAs: for each agent, states are its own emotional states, inputs are the emotions displayed by the interlocutor, while the transitions among states depend upon inputs and the agent's personality and attitude. Some transitions will be more probable in a friendly personality than in a crusty one, for instance. Moreover, transition probabilities are constantly updated throughout the emotional interaction depending on the agent's nature. A basic version of this model has been implemented in a real human-robot interaction. It is also shown that an adequate attitude can be acquired by an agent, simply through its emotional interaction with other entities. The probabilistic features of the model, and especially the capability to adapt to the interlocutor (through the basic update rule based on the agent's nature or by reinforcement learning) can help improve interactions quality: the agent will change its attitude toward the interlocutor, in a dynamic way, depending on the input received, thus endowing the interaction with a more lifelike appearance. Moreover, we introduce here Markov chains to derive quantitative measurements of the expected behavior of two agents; as far as we know, this is the first attempt of a quantitative analysis of emotional interaction.

The paper is organized as follows: in Sect. 2 we introduce the basic model that can be applied to a human-robot interaction context. Section 3 proposes a reinforcement learning approach to obtain interactions aimed at particular goals. In Sects. 4 and 5, we describe the implementation of the discussed ideas and the results obtained, and Sect. 6 develops a set of tools for a quantitative analysis of the expected behavior of the interacting agents. Lastly, Sects. 7 and 8 summarize our work, further discussing the major features of the model.

2 Interaction model

Our interaction model, describing the agent's emotional dynamics (for instance, let us consider a robot), is based on a probabilistic finite state automaton whose transition probabilities may change at each step. Formally, this is defined as a four-tuple $\langle S, U, P, s(0)\rangle$, where:

- $S = \{s_1, s_2, \dots, s_N\}$ is the (finite) set of emotional states (e.g. happy, sad, angry, etc.) for the robot.
- $U = \{u_1, u_2, \dots, u_M\}$ is the (finite) set of input, that is the emotions of the user (again, e.g. happy, sad, angry, etc.).
- $P = \{P_0, P_1, \dots\}$ is the sequence of probabilistic transition functions:
  $P_t : S \times U \times S \to [0, 1]$ for $t = 0, 1, \dots$; and
- $s(0)$ is the initial state.

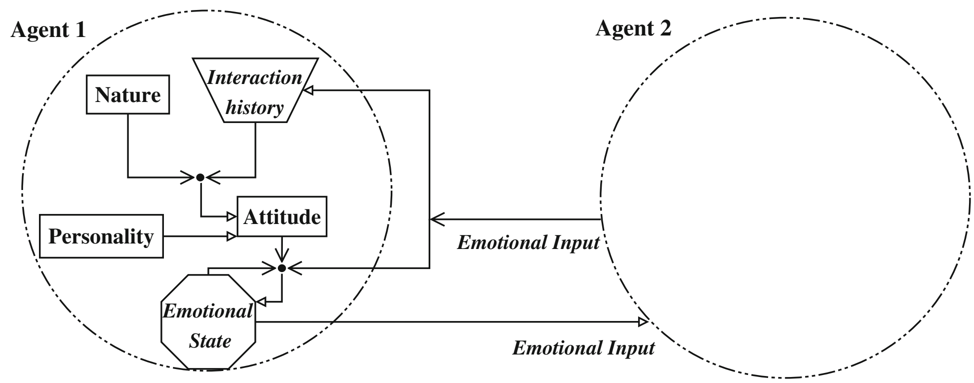


Fig. 1 The key elements driving an emotional interaction in our model are shown. Here we consider two generic agents interacting: the same schema can be applied both to human-robot interaction, and to interaction between two synthetic agents. *Arrows* show dependences among the different parts of the model. For each agent, the emotional input coming from the other agent, together with its own current emotional state and attitude, determine the next emotional state, which is then output as an emotional input for the interacting partner. Attitude is initially coincident with personality, and successively modified during the interaction according to the input history and the agent's nature

We explicitly notice that $\sum_{s' \in S} P_t(s, u, s') = 1$, for every t and every $(s, u) \in S \times U$. The sets of the robot and user's emotional states can be defined freely, and they can consist of the same or of different elements. The only constraint is that the robot is able to reliably detect the user's emotional states u_j.

The robot reads the user's emotional state (for instance, by processing the video of her facial expressions), which becomes the input for its PFSA. At time t, based on the input, u_j, and on the current emotional state of the robot, the transition function P_t outputs the probability of entering any possible next emotional state. P_0 can be regarded as the robot's *personality*. We compiled several personality files, containing the probability for each triple (s, u, s') – where s is the current state, u the user input and s' the next state – to occur: robots with different personalities will tend to react differently to the same emotional stimulus.

The transition function changes as a function of time: P_t, called the robot's *attitude*, is updated depending on the robot's *nature*, which represents the "easiness" to reach certain subsets of emotional states. Nature is defined as follows. First, inputs are clustered in K different categories, c_k: nice, sad and bad inputs, for instance. For each category, a set of one or more target states is defined: $\mathrm{TS}(c_k) = \{\mathrm{ts}_j\}$. Moreover, each category is associated with an eligibility trace, which summarizes the inputs history (Sutton and Barto 1998):

$$e_t(c_k) = \begin{cases} \alpha e_{t-1}(c_k) + h(c_k, u_j) & \text{if the current input is clustered in category } c_k \\ \alpha e_{t-1}(c_k) & \text{otherwise} \end{cases} \quad (1)$$

where α is the decay parameter and $h(c_k, u_j)$ represents the affinity between the current input, u_j, and the category, c_k: some inputs may be more representative of their category than others, and thus they will give a higher contribution to the relative eligibility trace.

When the trace associated with a category (say, c_k) reaches a predefined threshold value, the probability of entering all the target states for that category is incremented by Δ. Thus, for every target state $\mathrm{ts} \in \mathrm{TS}(c_k)$:

$$P_{t+1}(s, u, \mathrm{ts}) = P_t(s, u, \mathrm{ts}) + \Delta \quad (2)$$

The probability of entering the remaining states is decremented such that $\sum_{s' \in S} P(s, u, s') = 1$ for every $s \in S, u \in U$; this means that:

$$P_{t+1}(s, u, s') = P_t(s, u, s') - \frac{\Delta \cdot N_{\mathrm{TS}}}{N - N_{\mathrm{TS}}}, \quad \forall s' \in (S \backslash \mathrm{TS}(c_k)) \quad (3)$$

where N_{TS} is the number of target states for category c_k. So, if a robot has an imitative nature, the transition function will be changed so that the robot's behavior will tend to conform to that of the user: for instance, if the user has provided many positive inputs, the robot will more likely enter positive states; on the contrary, if the robot assumes a compensatory nature, its behavior will eventually diverge from that of the user.

The resulting model is a complete formalism for determining the agent's emotional response to the user's emotions, according to its key parameters of personality, attitude, and nature. In Fig. 1 the main elements composing the model and the existing relationships among them are illustrated: here a more generic scenario is considered, where both interacting partners are artificial agents (as better explained in Sect. 4).

3 Learning attitudes: a reinforcement learning approach

Let us now suppose that our agent has a goal, for instance to make its user (frequently) happy: it needs to learn a behavior that allows it to reach such goal. This problem can be reformulated as a typical *reinforcement learning* problem (Sutton and Barto 1998), where the agent learns a *policy*, that is a transition function that maximizes the long-term reward obtained from the environment, represented here by the user. That is, the agent's behavior should lead the user most frequently into the subset of the desired states.

At each time t, the environment assumes an observable state s_t, which is, for our application, the user's emotional state. The agent chooses an action, a_t, among the possible ones, to be exerted on the environment. The action is the emotional state that the agent chooses to display to the user and it is a function of the actual state of the environment, s_t; such a function represents the agent's *policy* and it is defined by the stochastic function $\pi(s, a)$. Each action has a different effect on the user, who, in turns, changes her own emotional state to s_{t+1} and gives an instantaneous reward, r_{t+1}, to the agent. This reward can be positive or negative, according to whether s_{t+1} is or it is not a useful state in reaching the predefined goal; that is, the instantaneous reward will be positive if s_{t+1} belongs to the set of desired states (for instance, joyful states if the goal is to make the interacting partner happy).

The agent's optimal policy is the one that maximizes the long-term reward, R_t (expected discounted return, Sutton and Barto 1998), that is

$$R_t = \sum_{k=0}^{T} \gamma^k r_{t+k+1} \tag{4}$$

where γ is a discount rate and T is the final step of learning (which goes to infinite in case of infinite horizon problems, like in the present case).

One of the most effective techniques for learning the optimal policy is *Q-learning* (Watkins 1989), where the agent learns an *action value function*, $Q(s, a)$, that gives the expected long-term return starting from state s, executing action a and, from that on, following the given policy, $\pi(s, a)$. For every step of each learning episode, the function $Q(s, a)$, is updated according to

$$Q(s, a) = Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \tag{5}$$

This technique allows the agent to learn the optimal value function and, at the same time, to learn the optimal policy for the given goal.

4 Implementation

The described model was implemented in an emotional interaction between a human and a robot and between two agents.

In the first case, an emotional interaction between an AIBO™ robot[1] and its master has been developed. The set of inputs, U, consists of the six universal emotions according to (Ekman 1992): joy, sadness, surprise, anger, fear, and disgust, plus the neutral emotion. S is restricted to four states: neutral, joy, sadness, and anger, for sake of simplicity, and the neutral state is chosen as initial state, $s(0)$. The output of the robot, at each step of interaction, is a predefined sequence of body movements, sounds and patterns of lights that represent the current emotional state of the robot.

The master's emotions are detected through the analysis of her facial expressions. This has been accomplished processing the video stream transmitted by AIBO's camera to the on-board processor (MIPS R7000, 576 Mhz). Basic image processing techniques, such as color segmentation, border extraction and block matching, have been implemented in order to meet the real-time response requirement (Campadelli and Lanzarotti 2002). The image processing module identifies a set of expressive features (e.g. mouth corner or inner portion of the eyebrow), which are mapped onto action units. These are the elementary facial movements defined by the *facial action coding system* (FACS (Ekman and Friesen 1978), see also (Magnenat-Thalmann et al. 1988) for a similar approach), and are then mapped to emotion expressions, through a fuzzy-like system of recognition scores. A four-step interaction with AIBO is shown in Fig. 2: in the left panels, AIBO's master displays her emotional state to AIBO through an adequate and well-defined facial expression. This is input to the emotional interaction model of AIBO, which, in turns, produces a new emotional state of AIBO, displayed through an adequate behavior as shown in the right panels.

In order to study more extensively the proposed interaction model, we have applied it to the interaction between two stochastic agents, where each interacting agent can be represented by a PFSA. In this case the state of the first automaton, A^1, becomes the input for the second one, A^2, and vice versa. Thus, we have two agents $A^1 = \langle S, U, P^1, s(0)^1\rangle$ and $A^2 = \langle S, U, P^2, s(0)^2\rangle$, where:

– The set of emotional states S is the same for both A^1 and A^2.
– The set of possible inputs, U, is coincident with the possible states, S.

[1] Web site: http://www.sony.net/Products/aibo.

Fig. 2 Four phases of an emotional interaction with AIBO. In the *left panels*, the emotional expression displayed by the master. In the *right panels*, the emotional response of AIBO. In particular, the sad expression is displayed by AIBO lowering its head and playing a sad melody; in the angry expression AIBO moves quickly forward in an aggressive fashion and growls, while in the happy expression the robot wags its tail and barks happily. In the neutral expression AIBO stands still, looking around as if it is uninterested in the user

– The probabilistic transition functions, P_0^1 and P_0^2, are different at start, that is the two agents have different personalities.
– The initial states $s(0)^1$ and $s(0)^2$ are different.

We have extended the set of possible emotional states for the two automata, S, by including, for each basic emotion, three different levels of intensity. Thus, the emotion of anger, for instance, is now represented by three different states, corresponding to low, medium, and high anger intensity: ANNOYED, ANGRY, and FURIOUS, respectively. A total of $N = 19$ states results, including the neutral emotional state.

Emotional inputs are clustered into M categories. In the following six categories are considered, each associated with a different basic emotion (so we have, for instance, the joy category, which is associated with all joyful inputs). Each level of an emotion contributes differently to the corresponding category eligibility trace ($h(c_k, u_j)$ in Eq. 1), so that probability update can be triggered by few intense inputs or many consecutive low-level inputs.

We first used a simplified version of the above-described model to study the interaction cycles with the aim to determine the most frequent behavioral patterns for the personality of the two agents. To the scope, we assumed that A^2 can be described as a deterministic stationary automaton and we analyzed the behavior of A^1 under the hypothesis that its transition probabilities, which define its personality, are not modified during the emotional interaction, making the automaton A^1 stochastic but stationary. Under these assumptions, we could adopt an algorithm based on depth-first search on the computation tree (Cormen et al. 2001) to extract the probability of each interaction cycle, that is a sequence of emotional states of A^1 which starts and ends in the same state.

Afterwards, we used the full model to analyze if a successful emotional relationship can be discovered by an agent, without any a priori information. To this end, let us regard A^1 as a probabilistic stationary environment for A^2: the emotional states output by A^1 are directly observable by the learning agent, A^2. A^2 has to learn, through reinforcement, a policy such that it obtains the maximum possible reward from A^1; that is, it has to learn a policy that outputs a set of actions, which let A^1 enter the predefined goal states most often (see Sect. 5). We have chosen to give the same amount of reward (namely, $r = 1$) whenever one of the goal states is reached; therefore, if the goal is, for instance, to make A^1 sad, each time it enters any state in the goal set, i.e. {MELANCHOLIC, SAD, IN_DESPAIR}, the same reward r will be delivered to A^2. After learning has been completed, the final policy defines a set of transition probabilities, which represent the attitude of A^2 after it has been adapted to the personality of A^1. Figure 3 summarizes the implemented reinforcement learning scenario.

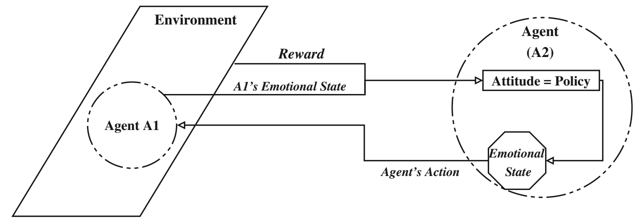


Fig. 3 The reinforcement learning approach applied to our interaction model. Here, one agent (namely, A^2) acts as the learning agent, while the other one (namely, A^1) embodies the environment. Actions exerted by A^2, that is its own emotional states as externally displayed, cause A^1 to enter a new emotional state. If this is a goal state, a positive reward is delivered to the learning agent. The reinforcement learning algorithm will then update the agent's policy (which, in our terminology, is its attitude), which in turn will provide a new action to be executed on the environment

5 Results

In the case of interaction between two agents, we could observe very different behavioral patterns, depending on their personality and nature, and possibly on the goal set. Let us start by briefly discussing how the key concepts of personality, attitude, and nature contribute to shaping an interaction.

In the simplest case, agents do not experiment any attitude evolution over time; therefore, the interaction depends only on the personality of the two agents. An evaluation of an agent's personality can be attained by analyzing its transition matrix, that is the three-dimensional matrix consisting of the transition function value for each possible triple (s, u, s'). Personalities can be designed by stressing, by means of a high probability value, the relevance of particular transitions of interest. For instance, a friendly personality can be characterized by high probability of entering positive emotional states (i.e., joyful), mixed with mirror behaviors (e.g. being sad if the partner is sad) as a sign of emotional involvement.

When the agents' attitude is allowed to evolve with time, the history of past inputs drives their interaction in a direction that is determined by the agents' nature. To illustrate this point, let us consider an interaction setting where both agents have a friendly personality (i.e., their transition matrices were crafted to define a friendly personality). If both agents are endowed with an imitative nature, their emotional relationship quickly converges to a sequence of positive emotional states, since the inputs that each agent is observing are most of the time positive. On the contrary, by setting one of the two agents' nature to compensatory, negative emotional states do occur as the emotional interaction goes on: upon receiving mostly positive inputs, the compensatory agent will, in fact, increase its probability of entering negative emotional states.

Having agents with quite different personalities interact also helps in obtaining more dynamic interactions. For instance, in Fig. 4, we had agent A^1, endowed with a friendly personality, interact with A^2, whose personality was obtained by linearly combining, with equal weights, the transition probability matrix for the friendly personality with one, randomly generated, describing deterministic transitions. In this case more complex interaction patterns tend to emerge, with a variety of experienced emotional states. Therefore, by carefully tuning personalities and natures, one can obtain interactions with the desired characteristics.

Alternatively, the overall trend of an interaction can be predetermined by giving a goal and letting the agent learn by itself the most adequate policy to reach that goal, during the interaction with another agent. Reinforcement learning is used here to this scope. A few results are now presented.

Let us suppose that at a certain time, t, the goal of making its friendly partner, A^1, angry is assigned to agent A^2. This means setting the states ANNOYED, ANGRY, and FURIOUS as

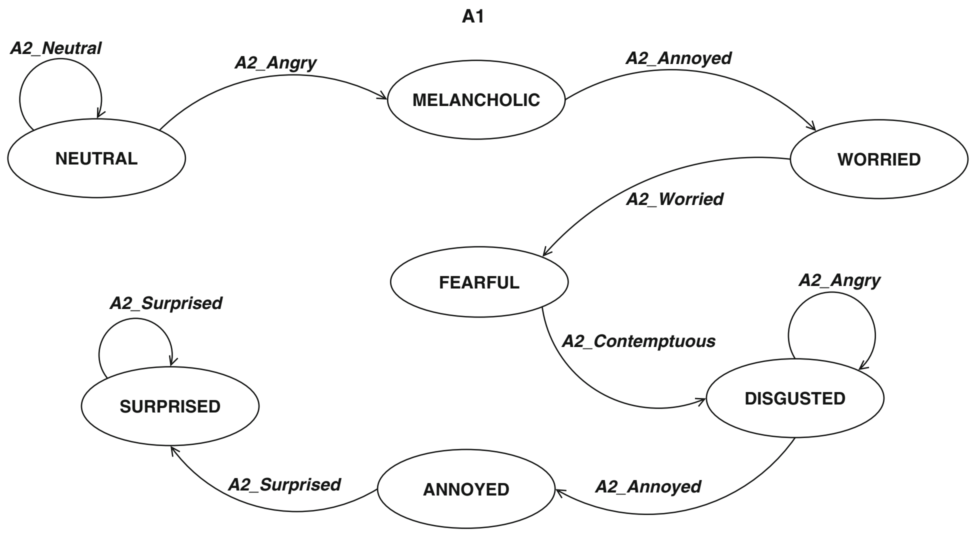


Fig. 4 This state transition graph shows 10 steps of an interaction between a friendly agent A^1 and an agent A^2, whose personality is obtained from the friendly one perturbing it with random traits. Only A^1 transitions are shown, while A^2 transitions can be easily derived from the arc labels. This interaction is rather dynamic, as it can be seen by the variety of states entered by the two agents; there is not a quick convergence to positive emotional states as we would expect if both agents had the same personality

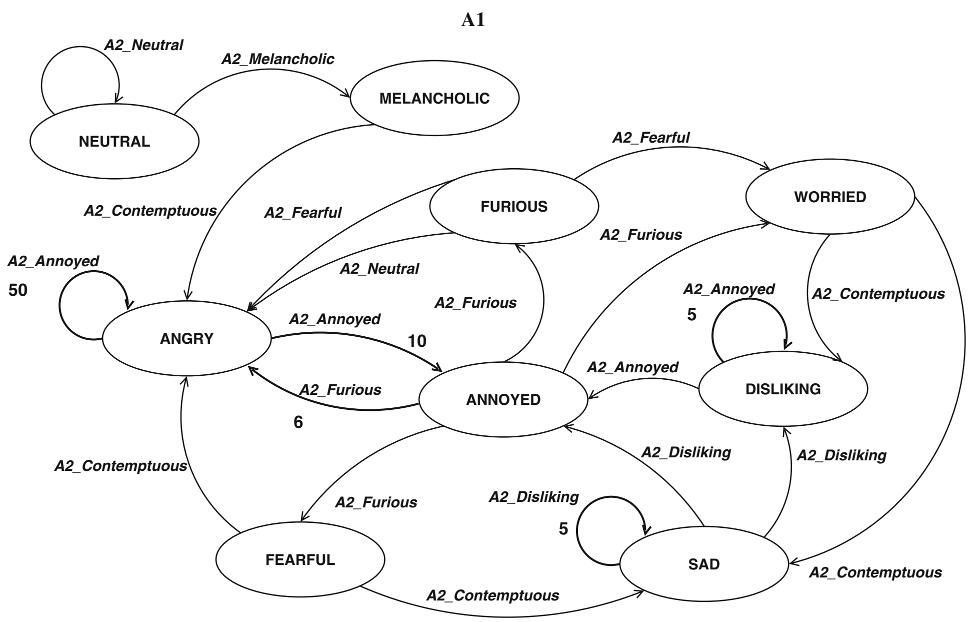


Fig. 5 The interaction between a stochastic agent, A^1, with a friendly personality and a second stochastic learning agent, A^2, is described in this state transition graph. The attitude of A^2 was modified with the goal to make A^1 angry most of the time. A total of 100 interaction steps are reported after learning has been completed. The states visited by A^1 are written inside the graph nodes, while the arcs represent the transitions to next states; each transition was induced by the action chosen by A^2, which is written besides the arc. Notice that, due to the stochastic nature of A^1, different next states can be reached from the same actual state with the same action (e.g. SAD/A2_DISLIKING). The starting emotional state is NEUTRAL. The number of occurrence of the most frequent transitions is reported in *bold* besides the corresponding arc

goal states. Through Q-learning (Watkins 1989) A^2 does learn a new policy (that is, it changes its attitude), to accomplish this goal, as shown in Fig. 5, where 100 steps of emotional interaction between A^1 and A^2 are reported, after learning has been completed. As it can be appreciated, 78% of the states reached by A^1 are goal states, showing that A^2 did learn a policy effective for the goal. Few state transitions occurred frequently during the interaction: in particular, the cycle on the ANGRY state was repeated 50 times over 100. This experimental observation is confirmed by the theoretical analysis of the cycle probability carried out on the 3D transition probability matrix: for instance, the four-step-length cycle ANGRY-ANGRY-ANGRY-ANGRY has, alone, a high probability of occurring (0.58) in this particular setting.

The history of a 100-step interaction between the same agents A^1 and A^2, when the goal for A^2 was to make A^1 surprised (the target states are now WONDERING, SURPRISED, ASTONISHED) is reported in the graph of Fig. 6. In

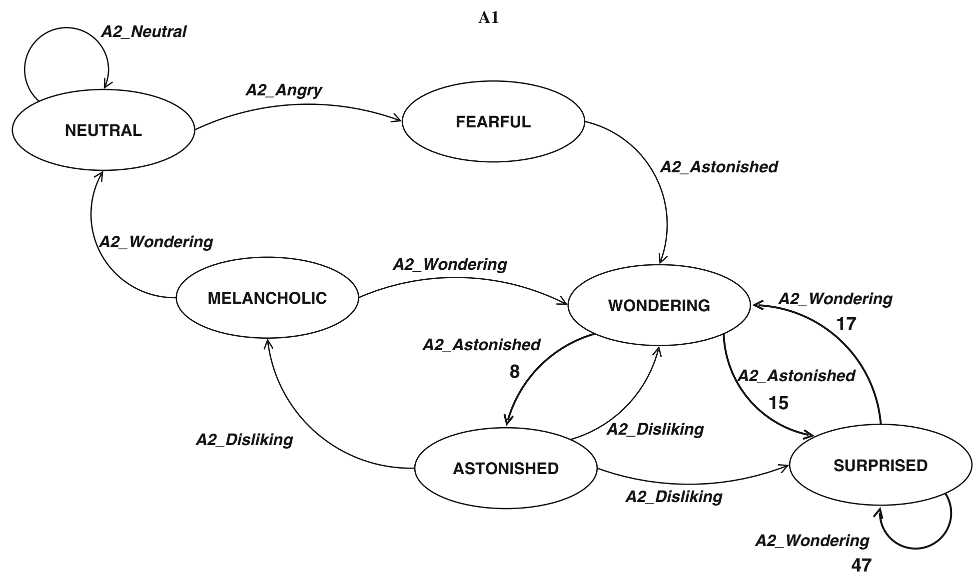


Fig. 6 The interaction between the same friendly stochastic agent, A^1, of Fig. 5 and a second stochastic agent, A^2, with a different goal, that is to make A^1 surprised. A total of 100 interaction steps are reported after learning has been completed. The number of occurrence of the most frequent transitions is reported in *bold* besides the corresponding arc

this case, the rate of goal states entered by A^1 was 95%, with the most frequently occurred cycle on the SURPRISED state.

Even though, in the two examples shown, A^1 has the same personality (and therefore, identical transition probabilities), the actions performed by A^2 according to the two different learned policies are effective in driving the emotional interaction with A^1 to very different groups of states as prescribed by the different goals set, thus producing very dissimilar interaction patterns. Given the same environment, having different goals necessarily leads to different policies.

On the other hand, given the same goal, but a different environment to act in, different policies will be developed by the learning agent. For instance, we considered a scenario where A^1's personality was suddenly changed, while A^2 maintained the attitude that it had previously learned, with the goal of making A^1 angry. In this situation, A^2's policy became less effective in reaching the goal set, since that policy was learned based on a different environment (i.e., a different A^1). After A^1's personality was changed, the interaction showed many different transitions, each of which occurred infrequently (Fig. 7), rather than the few transitions occurring rather frequently shown in Figs. 5 and 6.

Figure 8 shows the rate of goal states, p_{goal}, reached by A^1 over time, referred to blocks of 1,000 Q-learning iterations. Initially, when A^2 explores the state-action pairs, p_{goal} oscillates. It starts to increase when the agent discovers an effective policy, around the 40th block, to plateau at 70% at the 100th block, when the policy of A^2 does not receive any meaningful update anymore. At this stage, the agent ceases to explore, through random actions, the state-action space and just follows the learned policy. Using this policy, A^2 was

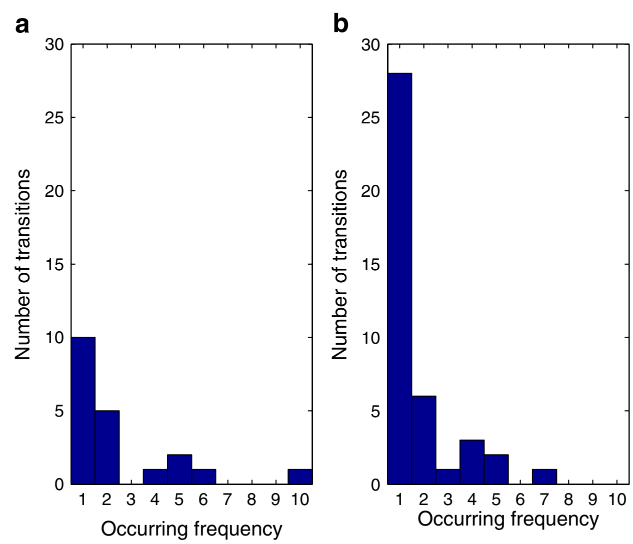


Fig. 7 **a** Shows the histogram of the state transitions, that is of the triples (s, u, s'), for the 100 emotional interaction steps, whose state transition graph in shown in Fig. 5. The cycle on the ANGRY state alone occurred 50 times and it is not shown in the histogram; as it can be seen all the other state transitions occurred rather infrequently. **b** Shows the histogram of the state transitions for 100 steps occurred just after the personality of A^1 was changed, by introducing some random traits in the friendly personality. The policy of A^2 remained that previously learned with the goal to make A^1, with the previous personality, angry. In this case, excluding one state transition (permanence in state DISLIKING on input ANNOYED) which occurred 28 times (also not shown in the histogram), all the other transitions occurred a low number of times (mainly, only once)

able to obtain consistently the desired behavior from A^1, as shown by the high rate of goal states reached.

At the time corresponding to the 110th block, the personality of A^1 was changed. A random component was added to the state transition matrix, to obtain a different personality of A^1. The new personality was obtained by linearly

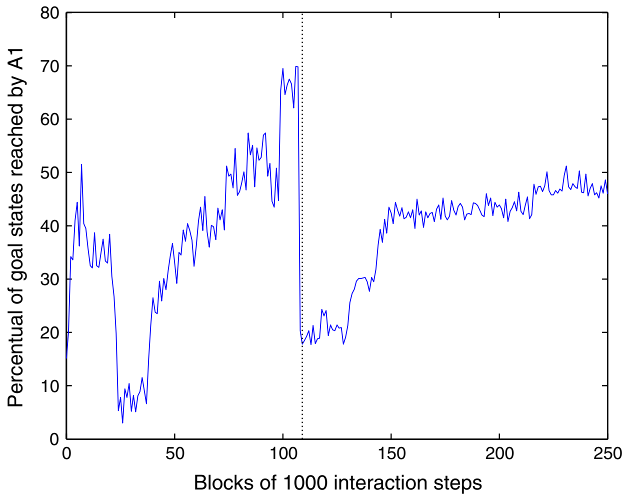
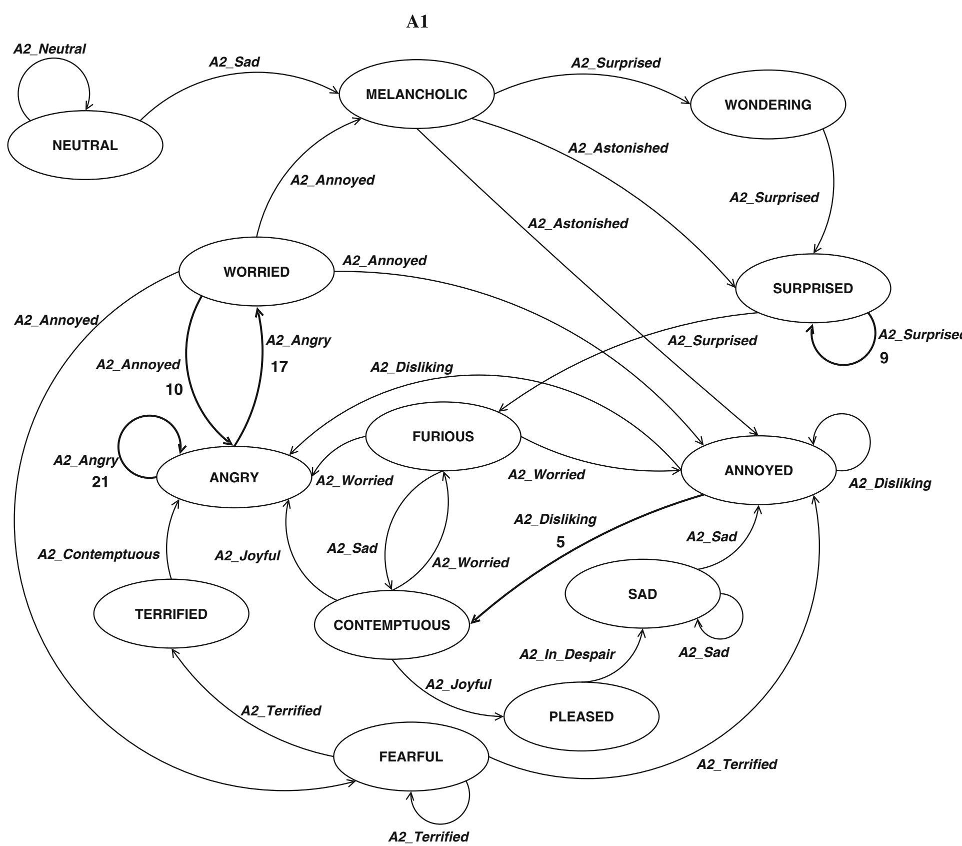


Fig. 8 Rate of goal states entered by A^1 during learning, computed over blocks of 1,000 interaction steps. At the beginning of block 110th A^1's personality is changed, which produces a sharp decrease in the goal rate, until a new effective policy is discovered

combining P_0^1 with a random transition matrix, with blending coefficient equal to 0.5. As a result the personality of the agent remained basically friendly. Nevertheless, the success rate abruptly decreased to around 20% and the agent A^2 had to learn again to cope with the new A^1. At this point, Q-learning was called to operate. As we can see in Fig. 8, after a set of 10 blocks, the goal rate starts increasing to plateau around the 220th block at a success rate of about 51%, which is lower than with the previous A^1's personality. A sequence of 100 interaction steps under these conditions (A^1 has a new personality, A^2 learned a new policy for it) is shown in Fig. 9. The state transition graph shows several frequent transitions which involve non-goal states. Comparing these results with the goal rates obtained for A^1's previous personality, the agent A^2 was less able to interact successfully with the agent A^1, when endowed with this new personality.

As a closing remark, we observe that the actual interaction resulting from the learning process can assume very different shapes: these depend not only on which goal has to be accomplished, but also on the dynamics of the learning process itself. In particular, the end result is influenced by the starting policy (A^2's personality) and by the stochastic nature of the environment. Together, these two aspects determine which regions of the state-action space are explored. In fact, A^2's starting policy determine which actions and how often they will be tried during learning. This means that some actions, though in principle useful in reaching the goal, will produce little or no reward at all, simply because they will hardly be experimented. We can think of someone who, although his goal would probably be accomplished by acting a certain

Fig. 9 An instance of interaction between a stochastic agent A^1, whose personality, initially friendly, was changed during interaction, and a second stochastic agent, A^2. A^2 had to learn a new policy to adapt to the change in personality of A^1, while maintaining the same goal, namely to make A^1 mostly angry. A total of 100 interaction steps are shown



way, still is unwilling to carry out these actions because of his personality. Moreover, since A^1's responses to A^2's actions are not deterministic, different reinforcement learning runs will result in different policies, depending on which state-action pairs will be observed. For instance, in the interaction instance shown in Fig. 6 A2_SURPRISED would be effective in keeping A^1 in the SURPRISED state (thus, a goal state); however, it is never performed since the reinforcement learning process was unable to highlight that particular state-action pair, preferring A1_SURPRISED-A2_WONDERING instead.

In summary, we showed how an emotional interaction can be driven towards predefined directions by using two alternative ways. The first way involves the tuning of interacting personalities and natures, in order to obtain the desired interaction: this could be, for instance, characterized by mostly positive states, or on the contrary by a wider variety of states. Alternatively, reinforcement learning can be applied to have one agent autonomously learn how to get the desired behavior from the interacting partner.

6 Quantitative behavior analysis

In a probabilistic framework like the one described here, each different run of the application will produce a different interaction pattern. Therefore, apparently there is no easy way to predict the outcome of an interaction. However, we show here that, if we take a picture of the two agents' attitude at some point in time (for instance after the adaptation/learning process is concluded), their asymptotic behavior can be described by resorting to Markov chains theory. To this purpose, we first summarize some well-known properties of (homogeneous finite) Markov chains that we are going to apply in our context (see for instance Iosifescu 1980; Häggström 2002).

6.1 Markov chains theory

Given a finite set S and a stochastic matrix $P = [P_{ij}]_{i,j \in S}$, i.e. a matrix satisfying conditions $P_{ij} \geq 0$ and $\sum_{l \in S} P_{il} = 1$ for all $i, j \in S$, let $\{X_n\}_{n \geq 0}$ be a Markov chain taking values on the set of states S with transition matrix P. Moreover, for every $n \in \mathbb{N}$, let $\mu^{(n)}$ be the probability distribution of X_n, which we consider here as a column array with index in S. Then, the values of $\mu^{(n)}$ and the n-steps transition probabilities between states can be computed from the powers of the matrix P: for each n and every $i, j \in S$, we have

$$\Pr(X_n = j \mid X_0 = i) \; = \; (P^n)_{ij} \tag{6}$$

$$\mu_j^{(n)} \; = \; \Pr(X_n = j) \; = \; (\mu^{(0)'} P^n)_j \tag{7}$$

The properties of $\{X_n\}_{n \geq 0}$ are of particular interest in the case when the matrix P is *primitive*, that is $P^k > 0$ for some $k \in \mathbb{N}$ (i.e., $(P^k)_{ij} > 0$ for all $i, j \in S$). This is equivalent to require that P is irreducible (meaning that the corresponding graph of transitions with non-null probability is strongly connected) and aperiodic (i.e., the greatest common divisor of the lengths of cycles is 1). If P is primitive then the following properties hold, which either are classical results or can be easily derived from standard issues:

1. There exists a unique probability distribution π over S such that

   $$\pi' P \; = \pi' \tag{8}$$

   which is called *stationary distribution* of the chain. Note that π' is a left eigenvector of P corresponding to the eigenvalue 1. This means that, if π is the probability distribution of X_0, then $\mu^{(n)} = \pi$ for all n.
2. For every $i, j \in S$

   $$\lim_{n \to +\infty} (P^n)_{ij} \; = \; \lim_{n \to +\infty} \Pr(X_n = j) \; = \; \pi_j \tag{9}$$

   that is, for large n, the probability that X_n equals j approximates π_j independently of the value of X_0, that is the initial state of the chain.
3. The values of π are related to the average waiting time for the first entrance in the states on the chain. More precisely, for every $i \in S$, let τ_j be the random variable defined by $\tau_j = \min\{n > 0 \mid X_n = j\}$. Thus, for all $i, j \in S$, the value $E_i(\tau_j) = E(\tau_j \mid X_0 = i)$ is the mean waiting time for the first entrance in j starting from the state i. Then, it turns out that

   $$E_j(\tau_j) = 1/\pi_j \text{ for each } j \in S \tag{10}$$

4. For $i \neq j$, the values $E_i(\tau_j)$ can be computed as follows. Let $G(z)$ be the matrix of polynomials in the variable z given by $G(z) = I - Pz$ and denote by $r_{ij}(z)$ the entry of indexes i, j of the adjunct of $G(z)$, i.e. $r_{ij}(z) = (-1)^{i+j} \det(G_{ji}(z))$ where $G_{ji}(z)$ is the matrix obtained from $G(z)$ by deleting the j-th row and the i-th column. Then, through the computation of some derivatives, one can prove that

   $$E_i(\tau_j) \; = \; \frac{r'_{ij} r_{jj} - r_{ij} r'_{jj}}{r^2_{jj}} \tag{11}$$

   where $r_{ij} = r_{ij}(1)$, $r_{jj} = r_{jj}(1)$, $r'_{ij} = r'_{ij}(1)$ and $r'_{jj} = r'_{jj}(1)$ (which are all well defined in this case).
5. One can also evaluate the error in the approximation of $\mu^{(n)}$ towards π. To this end, let us recall that the total

variation distance between two probability distributions μ, ν defined on the same finite set S, is given by

$$d_{\mathrm{TV}}(\mu, \nu) \;=\; \frac{1}{2}\sum_{i\in S}|\mu_i - \nu_i| \tag{12}$$

For such a distance the following relation is satisfied: $d_{\mathrm{TV}}(\mu, \nu) = \max\{|\mu(A) - \nu(A)| \mid A \subseteq S\} \leq 1$, and hence it yields a complete evaluation of the difference between two (finite) probability measures. Moreover, for every stochastic matrix T on S, we define the coefficient

$$m(T) \;=\; \frac{1}{2}\max_{i,j\in S}\left\{\sum_{l\in S}|T_{il} - T_{jl}|\right\} \tag{13}$$

which is the maximum (total variation) distance between pairs of rows of T. Then (still assuming P primitive) for every $\varepsilon > 0$ we have

$$d_{\mathrm{TV}}(\mu^{(n)}, \pi) \;\leq\; \varepsilon \tag{14}$$

for all $n \in \mathbb{N}$ such that

$$n \;\geq\; t\left(1 + \frac{\log_2 k - \log_2 \varepsilon - 1}{-\log_2 m(P^t)}\right) \tag{15}$$

where k is the cardinality of S and t is the smallest integer such that $P^t > 0$.

6.2 Markov chains for interaction analysis

Let us now go back to our interaction model and explicitly notice that two interacting agents can be regarded as a single closed system, as transition probabilities do not depend upon external input; here, states are pairs of emotional states (one for each agent). Such a system can be, therefore, described by a single matrix, M, collecting transition probabilities over its states: if S is the set of emotional states for both A^1 and A^2, the set of states over which M is defined will be $S \times S$. In other words, $M(i, j)$ represents the probability for the system to go, in one step, from state $i = (a, b)$ to state $j = (a', b')$, where a, a' are emotional states for agent A^1, and b, b' are emotional states for agent A^2. Matrix M can be easily derived from the two agents' personalities: entry $M(i, j)$ is obtained by multiplying $P^1(a, b, a')$ and $P^2(b, a', b')$.

In our context it often turns out that M is not an irreducible matrix; hence we cannot immediately apply the results presented in Sect. 6.1. M usually consists of more than one strongly connected component, but among these we can focus on *essential* components: these are defined as strongly connected subgraphs that cannot be left once entered. Therefore, at some point in the interaction the system enters one of these components, and afterwards only its states are visited. On the other hand, nonessential components are transient, and with probability 1 they will at some point be abandoned and never visited again; this is the reason why they can be excluded from our analysis.

In all the examples we considered, we found only one essential (strongly connected) component of M, which turned out to be aperiodic, too. Let us call M_{red} the transition matrix for such essential component. Its stationary distribution π can be computed as shown in Eq. 8. In our framework, π_i will thus provide the probability for the system of being in a state i, which represents a pair of emotional states, one for each agent. The corresponding probability for each agent separately can then be derived, by summing the probabilities over the states for the other agent:

$$\pi_1(a) = \sum_{b\in S}\pi_{(a,b)} \qquad \pi_2(b) = \sum_{a\in S}\pi_{(a,b)} \tag{16}$$

We observe that, by using standard methods (Iosifescu 1980; Chap. 3), one can compute the average waiting time required by the system to be absorbed into the essential component (also called time to absorption). In all our experiments, the average time to absorption is smaller than 3. This justifies our choice to study a restriction of M to the set of essential states.

The stationary distribution provides a description of our (reduced) system yielding the limit probability of visiting each state i. By comparing the stationary distribution for two different systems, i.e. two systems consisting of different interacting agents, we can quantify how their behavior differs: in particular, different interaction scenarios will be characterized by different states having maximum probability to be visited. A special case occurs when the two systems being compared are the system at the beginning of the interaction (before any adaptation occurs) and the system resulting from the adaptation process at the end of that interaction. In this particular case, by comparing the stationary distributions found for the two systems we can quantify the results produced by the interaction itself; here we are interested in identifying the most probable states, before and after the evolution of the two agents' attitude.

Let us start with a simplified example, where the set of emotional states includes just the six basic emotions plus the neutral state. In this setting, a *random* agent A^1 was designed, for which every transition (s, u, s') has the same probability. The transition matrix M can be computed assuming A^1 interacts with a friendly agent A^2 without adaptation. Twelve strongly connected components were found, among which only one was identified as an essential component of 38 states (over a total number of 49 states). For this component, we computed the stationary distribution, and we found that the most probable states of the system are (JOYFUL, JOYFUL) ($p = 0.1144$) and (SAD, SAD) ($p = 0.1118$). Focusing on

A^2's states only (A^1 has equal probability of entering each state, by definition), we found this stationary distribution: $\pi_2(\text{SAD}) = 0.3573$, $\pi_2(\text{JOYFUL}) = 0.1906$, $\pi_2(\text{NEUTRAL}) = 0.1857$, $\pi_2(\text{FEARFUL}) = 0.1435$, $\pi_2(\text{ANGRY}) = 0.0660$, $\pi_2(\text{SURPRISED}) = 0.0548$, $\pi_2(\text{DISGUSTED}) = 0.0022$.

We then had A^1 adapt to A^2, setting for it an imitative nature. As a result, we expected that the stationary distribution over A^1's states would somehow approximate that found for A^2 at the beginning of the interaction. This was confirmed by our experimental analysis: after 200 steps of interaction, we found that A^1's stationary distribution was: $\pi_1(\text{SAD}) = 0.3413$, $\pi_1(\text{JOYFUL}) = 0.1547$, $\pi_1(\text{FEARFUL}) = 0.1430$, $\pi_1(\text{ANGRY}) = 0.1080$, $\pi_1(\text{SURPRISED}) = 0.0847$, $\pi_1(\text{DISGUSTED}) = 0.0847$, $\pi_1(\text{NEUTRAL}) = 0.0837$. In particular, we can see that the most probable state is SAD, as it was for A^2 at the beginning of the interaction. This happens because, when an imitative nature is set, the adapting agent increases its probability of entering those states which have been observed more often in the other agent.

Similarly, we studied the asymptotic behavior of the systems depicted in Figs. 5, 6, and 9. For clarity purposes, we will refer to these systems as *System 1*, *System 2*, and *System 3*, respectively. In these examples, the total number of states of the system is $19^2 = 361$.

Figure 5 shows an instance of interaction between a friendly agent and a second agent having learned how to make A^1 angry most of the time. The corresponding matrix M contains only one essential component of 15 states; therefore we could study the reduced matrix M_{red}, of size 15×15. The computation of the stationary distribution for this component showed that the most probable state is (ANGRY, ANNOYED), with $p = 0.5148$, followed by (ANNOYED, FURIOUS), with $p = 0.1548$, and (SAD, DISLIKING), with $p = 0.0973$. This is mirrored by the actual behavior of the system as shown in Fig. 5: for instance, we can see that, over 100 steps, the state pair (ANGRY, ANNOYED) occurred 60 times, thus rather close to the theoretically predicted frequency.

Similarly, when we considered *System 2*, consisting of a friendly agent interacting with an agent trained for making the partner surprised (see Fig. 6), we identified an essential component of 10 states. According to the associated stationary distribution, the most probable states are (SURPRISED, WONDERING) ($p = 0.6286$), (WONDERING, ASTONISHED) ($p = 0.2292$), and (ASTONISHED, DISLIKING) ($p = 0.0917$). As in the previous case, observed frequencies are close to those provided by the stationary distribution.

Similar remarks can be made about *System 3*, for which an instance of interaction is shown in Fig. 9. The only essential component in this system consists of 19 states, and its stationary distribution identifies, as most probable states, (ANGRY, ANGRY), with $p = 0.2930$, (WORRIED, ANNOYED), with $p = 0.1725$, and (ANNOYED, DISLIKING), with $p = 0.1250$. Taken altogether, these data confirm that the learning process was successful in having A^2 acquire an effective policy, since the goal states defined for A^1 are among the most probable states of the system, in each of the considered examples.

A natural question now is to establish how precisely these stationary distributions describe the actual behavior of the systems. To this purpose, we used property 5 in Sect. 6.1 to compute the minimum number of steps required to approximate the stationary distribution with an arbitrary small error ε. For $\varepsilon = 0.001$, we computed this value for the three examples considered above, and found, respectively, $n_1 = 38.12$, $n_2 = 26.62$, and $n_3 = 42.04$. This means that the stationary distribution is a suitable descriptor of the actual behavior of the above systems even after a limited amount of steps. This also explains why the probability values of the stationary distributions are rather close to the frequencies observed in the experiments.

Properties 3 and 4 of Sect. 6.1 can be used to compute mean entrance times into a given state of interest j starting from another state of interest i, for any i, j. To this purpose, one can define a set of starting states, SS, and a set of end states, ES, and compute the minimum, maximum, and average waiting time to go from SS to ES. More formally, let us call $S_{\text{red}} \subseteq S \times S$ the set of states in the essential component we are studying, and let us define $\text{SS} \subseteq S_{\text{red}}$, $\text{ES} \subseteq S_{\text{red}}$. Then we can build a matrix of mean entrance times:

$$\text{MET}_{\text{SS,ES}} = \{E_{(a,b)}(\tau_{(c,d)}) \mid (a,b) \in \text{SS}, (c,d) \in \text{ES}\} \tag{17}$$

Depending on how we define SS and ES, we can study the mean waiting time for the system to go, for instance, from (JOYFUL, SAD) to (SAD, JOYFUL). Alternatively, we can focus on a single agent's states, to check, for instance, how many steps are required, on average, for A^1 to go from the JOYFUL state to the SAD one; the same holds for agent A^2.

At first, we applied the analysis of mean entrance times to the simplified interaction scenario we introduced earlier in this section, where a random agent interacts with a friendly partner and no adaptation occurs. We set $\text{SS} = \{(a,b) \mid b = \{\text{ANGRY}\},\ a \in S\}$, and $\text{ES} = \{(a,b) \mid b = \{\text{JOYFUL}\},\ a \in S\}$. In other words, we were interested in studying how many steps are required for the friendly agent A^2 to go from the ANGRY state to the JOYFUL one, regardless of A^1's state. We computed $\text{MET}_{\text{SS, ES}}$, and found a minimum value of 9.27 and a maximum value of 374.43 (mean 191.53). Therefore, whereas in the best case the JOYFUL state is reached quite quickly, in the worst case scenario reverting the emotional state of A^2, from a negative to a positive one, can take very long. This is due both to the random nature of agent A^1 and to the absence of a strategy in A^1's attitude aimed at making A^2 JOYFUL.

When considering a reinforcement learning scenario, where A^2 interacts with A^1 by adopting the policy it learned for driving A^1 towards some given goal states, it is natural to ask how many interaction steps, on average, will be required for leading A^1 to a goal state, regardless of A^2's states: $\mathrm{ES} = \{(a, b) \mid a$ is a goal state, $b \in S\}$. The computation of mean entrance times, in this case, provides a measure of effectiveness of the learning process, in terms of how quickly a goal is reached. Since, by default, all our interactions were started from the initial state (NEUTRAL, NEUTRAL), a natural choice for SS would be $\mathrm{SS} = \{(\text{NEUTRAL}, b) \mid b \in S\}$.

We, therefore, analyzed mean entrance times for the three systems previously considered, starting with *System 1*. Since, in this example, A^2 learned a policy for making A^1 angry most of the time, the set of ending states was defined as $\mathrm{ES} = \{(a, b) \mid a = \{\text{ANNOYED}, \text{ANGRY}, \text{FURIOUS}\},\ b \in S\}$. Note that in this example no state of the form (NEUTRAL, b) belongs to the essential component, and hence we could not use it as starting state. A natural choice of starting state in this case is (MELANCHOLIC, CONTEMPTUOUS), which seems to be rather far away from the states in ES. Under these assumptions, we computed $\mathrm{MET}_{\mathrm{SS,ES}}$ and found a minimum of 5.91 and a maximum of 213.10 steps, on average, for going from states in SS to states in ES (mean 77.98).

In *System 2*, A^2's aim was to make A^1 surprised. Therefore we defined $\mathrm{ES} = \{(a, b) \mid a = \{\text{WONDERING}, \text{SURPRISED}, \text{ASTONISHED}\},\ b \in S\}$. Since the (NEUTRAL, ANGRY) state belongs to the essential component, we could choose it as the unique starting state. The computation of $\mathrm{MET}_{\mathrm{SS,\ ES}}$ yielded a minimum of 3.86 and a maximum of 12.43 steps (mean 7.07).

Lastly, we applied the same analysis to *System 3*, where goal states are the same as *System 1*, and therefore: $\mathrm{ES} = \{(a, b) \mid a = \{\text{ANNOYED}, \text{ANGRY}, \text{FURIOUS}\},\ b \in S\}$, $\mathrm{SS} = \{(\text{NEUTRAL}, b) \mid b \in S\}$. Given $\mathrm{MET}_{\mathrm{SS,\ ES}}$, the minimum and the maximum values are 6.58 and 18.14, respectively (mean 11.02).

Therefore, we can conclude that, in the last two examples, goal states are reached very quickly after interaction starts (within 15–20 steps, approximately), which confirms that the learned policies are effective in driving A^1's behavior to the given goals. On the contrary, in order to reach the goal states *System 1* seems to require rather a long time (about 78 steps, on average) with respect to the size of the essential component (15, in this case). However, we observe that this is mainly due to two particular end states ((FURIOUS, FURIOUS) and (FURIOUS, NEUTRAL)) that in general have very low entrance probabilities; we note that the other three goal states in this example can be reached within 30 steps, confirming in any case a good performance of the system.

To summarize our results, we used classical properties in Markov chains theory to extract some statistical information about given interaction scenarios. Through the computation of the stationary distribution of the essential component of the system, we extracted the most probable states. These largely correspond to goal states as defined in the reinforcement learning framework. The accuracy of the approximation of the actual distribution on the system to the stationary distribution was computed, too, and found to be rather good since the earliest interaction steps. Mean waiting times for going from state i to state j were used to establish the number of steps required, on average, for the system to reach a set of states of interest. This analysis showed that goal states, as defined in the reinforcement learning framework, are reached rather quickly after interaction starts, confirming that the learned policies are effective.

7 Discussion

The model is characterized by three key elements: *personality*, *attitude* and *nature*. These terms were chosen to fit the corresponding psychological elements. *Personality* (Ryckman 2003) is related to the psychological structure of an individual; it is believed to be mostly stable and independent from external events. In our model, the agent's personality is represented as the transition probabilities matrix given at time zero, P_0. Personality defines therefore the initial interaction behavior of the agent (which emotional states will be assumed, and how frequently) and acts as a basis upon which the agent's attitude can develop. *Attitude* (Zanna and Rempel 1988) is more related to specific situations and targets (other agents, for instance), shaping the individual's behavior in response to external stimuli. In the presented model, starting from an initial behavior, exclusively defined by the agent's personality, the agent gradually has its behavior changed according to the interaction history, so that the current emotional interaction behavior results from a combination of personality and a sequence of attitude changes. The current attitude towards the interlocutor is therefore defined as the transition probabilities matrix at time t, P_t. Lastly, we defined *nature* as a driving force used for updating attitude according to the past interaction sequence, to favor a subset of the emotional states, which will be entered more frequently in subsequent interaction steps. We considered, for instance, an imitative and a compensatory nature, driving attitude updates towards two opposite directions.

The resulting model allows carrying out implicitly what is called "affective reasoning" (e.g. André et al. 1999), in which "on the basis of the domain knowledge [. . .] the appropriate emotional reaction is determined" (Schroeder 2004). In classical approaches this is achieved by giving to the agent an explicit knowledge of the behavior of the other agent (Ortony et al. 1988). Here, instead, it is the agent itself that discovers the best emotional reactions to the interlocutor, without building any explicit knowledge of it.

In fact, using reinforcement learning it is possible to generate policies for eliciting specific behaviors: the interacting agent will frequently be happy, angry or sad, depending on the goals we have set for the learning phase. Thus, the learning agent will then display ad hoc emotional states for causing its partner to enter desired states. Of course, assuming an emotional state is generally not a matter of decision: it is a spontaneous, often uncontrollable event. Emotional expression, though, can generally be controlled at some level and used to influence the interlocutor. In this sense we believe that reinforcement learning can be applied to emotional interactions in order to define strategies for driving them to desired results. By appropriately setting the interacting personalities, natures and goals, the interaction can be directed toward a general trend, without losing the unpredictable traits of a real emotional interaction.

Emotional states change with continuity and can be viewed as points in a multi-dimensional continuous space (Schlosberg 1954), which is organized along affective dimensions (such as positive/negative). Emotion gradations can be well captured, as any level of emotion can be represented by a point in this space. In our case, instead, emotional states are discretized into a finite number, and for each state a few discrete levels are considered. For instance, the surprise emotion is represented by three different states (WONDERING, SURPRISED, and ASTONISHED) corresponding to increasing levels of this emotion. Thus, the model cannot entirely account for all different emotion gradations. However, the interaction framework does not require to model a full range of gradations, since even during human-human interaction the displayed emotional states are not perceived at the highest precision, but rather clustered to wider categories (e.g. slight as opposed to intense surprise). While this applies to emotional interaction modeling, fuzzy approaches would probably help when considering more complex processes involving a fine representation of emotional gradations.

It is often postulated (Scherer 1984) that emotional states are subject to continuous changes as cognitive appraisal of external stimuli is carried out by the organism. In our model this process has been discretized so that a single emotional state is entered following the evaluation of the interlocutor's emotion and maintained until the next interaction step, when a new emotional input is fed into the model. We can regard this single emotional state, for each interaction step, as being the only one observable by the other agent as the final result of a continuous emotional process, consisting of intermediate steps that, however, are concealed. Since we are not focusing on reproducing the whole process of emotion production, but rather on modeling emotional interaction, we believe that this discretized approach may be suitable for our aims.

The probabilistic model presented here can be considered as an evolution of previously proposed models. In (Chittaro and Serra 2004), the term personality has a similar meaning to ours; however, emotional states are not taken into account and no updating of the transition probabilities is included. The model in (Kopecek 2003) includes inputs and outputs (not just external communications of the emotional state) and no updates. Principally, while our model has a strong time-varying imprint, those frameworks are mainly static. The dependence of transition probabilities on past inputs is considered in (Kuhnlenz and Buss 2004) through the implementation of a digital (forgetting) filter. However, modifications induced by inputs are not permanent: the impact of a stimulus is effective only during the time interval corresponding to the length of the filter response. Therefore this mechanism does not allow for long-term adaptation of the transition function. In the present work, instead, the transition function for an agent does change in order to adapt to the interaction partner. This reflects an interaction scenario where emotional input from the partner has a strong impact on the agent's attitude, actually shaping it. In order to filter out input variability, time filtering is introduced through the trace mechanism in Eq. 1. Our model differs also from that in (Kuhnlenz and Buss 2004) for other architectural details: our concepts of personality and attitude can be seen as unifying the two matrices of the HMM in a simpler structure; moreover, in our model emotional states are not hidden but directly output to the outside.

In (Nomura 1996) two of our keywords, attitude and personality, are used, but in different roles. In our model, personality (at the start) and attitude select the next emotional state according to the current state and to the input, while in (Nomura 1996) personality determines the probability of some output given the current state and attitude provides the probability of each next state given current input and output. In our model, this information is merged into one transition probabilities matrix, P_t, representing the current attitude of the agent built over time, starting from the basis defined by its personality. In both models we have attitude updates, but while in (Nomura 1996) these updates are based on the emotional state and the personality of the individual, in our model attitude is changed according to the interaction history and the agent's nature. Moreover, in our work inputs and outputs are not actions as in (Nomura 1996), but bodily expressions of the current emotional state. Lastly, the intended aims are quite different. In (Nomura 1996) the goal is to study the dynamics of the relationships within a group, registering attitude changes over time: that is, to analyse group dynamics. In our work, instead, the aim is to synthesize and to predict emotional behavior in the context of interaction between a human and a robot (or between two generic agents); our model can also be employed to explore likely interactions between two individuals.

In contrast to the previously discussed works, which lack a robotic implementation, the basic interaction model was implemented on an AIBO robot and therefore experimented also in a real human-robot scenario, thus showing its effectiveness in supporting an emotion-based interaction. Nevertheless, to be applied in more complex human-robot interactions, an accurate detection of human expressions has to be carried out. This is still far beyond reach to actual computer systems as facial movements are small and very fast. In order to have a successful emotional interaction, facial expressions have still to be somehow exaggerated as shown in Fig. 2.

8 Conclusion

The proposed model allows, thanks to its probabilistic and dynamic nature, to model a wide variety of behaviors occurring during emotional interactions. By adopting the reinforcement learning framework, the model is also able to automatically discover behavioral patterns which adapt to the interlocutor, in order to successfully interact with another agent, without needing any a priori knowledge of it.

The described interaction model has a basic structure that can easily be extended and personalized, by adding or modifying emotional states, inputs, personalities and natures. The model can be used as a basis for emotional agents (e.g. in video games) or robots, in an effort to have technology adapt to its user's characteristics. We can imagine a video game where the user's avatar has to interact with different synthetic agents in order to walk successfully through the game. Such an interaction may be based also on emotional cues, with synthetic agents reacting differently to different emotional inputs, and thus leading to different outcomes for the game itself. In social robotic applications, the robot's personality might be carefully designed to best meet the needs, or simply the preferences, of the user; similarly, starting from a basic personality, user-robot interactions would autonomously shape the robot's attitude according to the user's wishes. The capability of the model of adapting to the input trend—where adaptation can be meant in an imitative, or opposite direction, or in learning how to drive the interlocutor's behavior toward desired goals, or also defined anew by the user—improves the interaction quality, providing lifelike features to its dynamics.

On the other hand, the model allows for behavioral dynamics analyses: high probability cycles can be identified as characteristic patterns for the considered emotional context. Moreover, Markov chains theory can be applied to specific instances of interaction for extracting statistical information about the expected overall behavior of the system, for instance for predicting how frequently, or after how many interaction steps, an emotional state will be entered. Our study is here based on the properties of homogeneous Markov chains, where transition matrices do not change with time. It would be interesting to develop a similar analysis by using nonhomogeneous chains, where transition probabilities change as time goes by.

References

Anderson K, McOwan P (2006) A real-time automated system for the recognition of human facial expressions. IEEE Trans Syst Man Cybern B Cybern 36(1):96–105

André E, Klesen M, Gebhard P, Allen S, Rist T (1999) Integrating models of personality and emotions into lifelike characters. In: Proceedings of the workshop on affect in interactions—Towards a new Generation of Interfaces, pp 136–149

Argyle M (1975) Bodily communication. International Universities Press, Inc.

Campadelli P, Lanzarotti R (2002) Localization of facial features and fiducial points. In: Proceedings of the IASTED International Conference Visualisation, Imaging and image (VIIP2002)

Chernova S, Arkin RC (2007) From deliberative to routine behaviors: a cognitively inspired action-selection mechanism for routine behavior capture. Adapt Behav 15:199–216

Chittaro L, Serra M (2004) Behavioral programming of autonomous characters based on probabilistic automata and personality: Research articles. Comput Anim Virtual Worlds 15(3–4): 319–326

Ciota Z (2005) Emotion recognition on the basis of human speech. In: The 18th International Conference on Applied Electromagnetics and Communications

Cormen TH, Leiserson CE, Rivest RL, Stein C (2001) Introduction to Algorithms, 2nd edn. The MIT Press, Cambridge

Damasio A (1994) Descartes' error: emotion, reason, and the human brain. Avon Books

Doshi P, Gmytrasiewicz P (2004) Towards affect-based approximations to rational planning: a decision-theoretic perspective to emotions. In: Working Notes of the Spring Symposium on Architectures for Modeling Emotion: Cross-Disciplinary Foundations

Ekman P (1992) An argument for basic emotions. Cogn Emot 6(3–4): 169–200

Ekman P, Friesen WV (1978) Manual for the Facial Action Coding System. Consulting Psychologists Press, Palo Alto

Fabri M, Moore DJ, Hobbs DJ (1999) The emotional avatar: non-verbal communication between inhabitants of collaborative virtual environments. In: Proceedings of the International gesture workshop on gesture-based communication in human-computer interaction (GW '99), pp 269–273

Häggström O (2002) Finite Markov Chains and Algorithmic Applications. Cambridge University Press, London

Hopcroft JE, Ullman JD (1979) Introduction to automata theory, languages and computation. Addison-Wesley, Reading

Inamura T, Inaba M, Inoue H (2004) Pexis: probabilistic experience representation based adaptive interaction system for personal robots. Syst Comput Jpn 35(6):98–109

Iosifescu M (1980) Finite Markov processes and their applications. Wiley, NY

James W (1884) What is an emotion? Mind 9:188–205

Kopecek I (2003) Constructing personality model from observed communication. In: Proceedings of the 9th International Conference on user modeling; assessing and adapting to user attitides and effect: why, when and how? pp 28–30

Kuhnlenz K, Buss M (2004) Towards an emotion core based on a hidden markov model. In: Proceedings of the 13th IEEE international workshop on robot and human interactive communication (RO-MAN 2004)

Lester JC, Towns SG, Callaway CB, Voerman JL, FitzGerald PJ (2000) Deictic and emotive communication in animated pedagogical agents. In: Embodied conversational agents, MIT Press, Cambridge, pp 123–154

Magnenat-Thalmann N, Primeau E, Thalmann D (1988) Abstract muscle action procedures for human face animation. Vis Comput 3(5):290–297

Nomura T (1996) Generation of relations between individuals based on a stochastic automaton and an analogy from social psychology. In: Proceedings of the ALIFE V Poster Presentations, pp 125–132

Ortony A, Clore GL, Collins A (1988) The cognitive structure of emotion. Cambridge University Press, Cambridge

Paz A (1971) Introduction to probabilistic automata. Academic Press, London

Picard RW (1997) Affective Computing. MIT Press, Cambridge

Picard RW, Vyzas E, Healey J (2001) Toward machine emotional intelligence: analysis of affective physiological state. IEEE Trans Pattern Anal Mach Intell 23(10):1175–1191

Rabin MO (1963) Probabilistic automata. Inf Control 6(3):230–245

Ryckman RM (2003) Theories of personality, 8th edn. Wadsworth Publishing, Belmont

Scherer KR (1984) On the nature and function of emotion: A component process approach. In: Approaches to emotion, Lawrence Erlbaum Associates, pp 293–317

Schlosberg H (1954) Three dimensions of emotion. Psychol Rev 61(2):81–88

Schroeder M (2004) Speech and emotion research: an overview of research frameworks and a dimensional approach to emotional speech synthesis. Dissertation, Saarland University

Silva PRD, Bianchi-Berthouze N (2004) Measuring posture features saliency in expressing affective states. In: Proceedings of the 2004 IEEE international conference on intelligent robots and systems (IROS'04), pp 2003–2008

Silva PRD, Osano M, Marasinghe A, Madurapperuma AP (2006) Towards recognizing emotion with affective dimensions through body gestures. In: Proceedings of the 7th international conference on automatic face and gesture recognition (FGR06), pp 269–274

Sutton RS, Barto AG (1998) Reinforcement Learning: an Introduction (adaptive computation and machine learning). The MIT Press, Cambridge

Watkins C (1989) Learning from delayed rewards. Dissertation, University of Cambridge, England

Zanna M, Rempel JK (1988) Attitudes: a new look at an old concept. In: Bar-Tal D, Kruglanski A (eds) The social psychology of knowledge. Cambridge University Press, Cambridge, pp 315–334