Adaptive **Behavior**

Adaptive Behavior
2016, Vol. 24(5) 350–372
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1059712316665451
adb.sagepub.com
SAGE

Combining intention and emotional state inference in a dynamic neural field architecture for human-robot joint action

Rui Silva[1], Luís Louro[1], Tiago Malheiro[1], Wolfram Erlhagen[2] and Estela Bicho[1]

Abstract

We report on our approach towards creating socially intelligent robots, which is heavily inspired by recent experimental findings about the neurocognitive mechanisms underlying action and emotion understanding in humans. Our approach uses neuro-dynamics as a theoretical language to model cognition, emotional states, decision making and action. The control architecture is formalized by a coupled system of dynamic neural fields representing a distributed network of local but connected neural populations. Different pools of neurons encode relevant information in the form of self-sustained activation patterns, which are triggered by input from connected populations and evolve continuously in time. The architecture implements a dynamic and flexible context-dependent mapping from observed hand and facial actions of the human onto adequate complementary behaviors of the robot that take into account the inferred goal and inferred emotional state of the co-actor. The dynamic control architecture was validated in multiple scenarios in which an anthropomorphic robot and a human operator assemble a toy object from its components. The scenarios focus on the robot's capacity to understand the human's actions, and emotional states, detect errors and adapt its behavior accordingly by adjusting its decisions and movements during the execution of the task.

Keywords

Human-robot joint action, goal inference, emotional state inference, error detection, decision making, dynamic neural fields

1 Introduction

A major challenge in modern robotics is the design of socially intelligent robots that can interact or cooperate with people in their daily tasks in a human-like way. Needless to say that non-verbal communication is an essential component for everyday social interactions. We humans continuously monitor the actions and the facial expressions of our partners, interpret them effortlessly regarding their intentions and emotional states, and use these predictions to select adequate complementary behavior. Thus, natural human-robot interaction or joint activity requires that assistant robots are endowed with these (high level) social cognitive skills.

There have been various kinds of interaction studies that have explored the role of emotion/affect in human-robot interaction (HRI) (Breazeal, 2003a, 2003b; Cañamero & Fredslund, 2000; Hegel, Spexard, Wrede, Horstmann, & Vogt, 2006; Kirby, Forlizzi, & Simmons, 2010; Novikova & Watts, 2015). The results of such studies have clearly shown that endowing robots with – the recognition and display of human-like – emotions/affects critically contributes to making the HRI more natural and meaningful, from the perspective of the human interacting with the robot (Breazeal, 2003a, 2003b; Cañamero, 2005; Hegel et al., 2006; Kędzierski, Musznski, Zoll, Oleksy, & Frontkiewicz, 2013). However, in all these interaction experiments the robot and the human were not a team, in that the interactions did not involve joint action tasks. One exception goes to the work reported in Scheutz, Schermerhorn, and Kramer (2006), where the robot and the human were

[1]Department of Industrial Electronics, University of Minho, Portugal
[2]Department of Mathematics and Applications, University of Minho, Portugal

Corresponding author:
Estela Bicho, Department of Industrial Electronics, University of Minho, 4804-533 Guimaraes, Portugal.
Email: estela.bicho@dei.uminho.pt

both needed for the task and neither robot nor human could accomplish the task alone. Their results have shown that expressing affect and responding to human affect with affect expressions can significantly improve team performance in a joint human-robot task. However, the human and the robot interacted solely based on 'natural' language, there was no physical interaction, and the robot was not making autonomous decisions, i.e. the robot always carried out human orders (see also Scheutz, 2011).

The work reported here aims to contribute to filling in this gap. Our approach is motivated by recent research in cognitive psychology and cognitive neuroscience that posits that various kinds of shared emotions can, not only motivate participants to engage and remain engaged in joint actions, but also facilitate processes that are central to the coordination of participants' individual actions within joint action, such as representing other participants' tasks, predicting their behavior, detecting errors and correcting accordingly, monitoring their progress, adjusting movements and signaling (Michael, 2011; Rizzolatti & Sinigaglia, 2008).

In order to combine emotions into the decision making and complementary behavior of an intelligent robot cooperating with a human partner our group relies on the development of control architectures for human-robot interaction that are strongly inspired by the neuro-cognitive mechanisms underlying joint action (Bekkering et al., 2009; Poljac, van Schie, & Bekkering, 2009; van Schie, van Waterschoot, & Bekkering, 2008) and shared emotions in humans (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Iacoboni et al., 2005; Wicker et al., 2003). We believe that implementing a human-like interaction model in an autonomous assistive robot will greatly increase the user's acceptance to work with the artificial agent since the co-actors will become more predictable for each other (see also Fong, Nourbakhsh, and Dautenhahn (2003); Kirby et al. (2010)).

Humans have a remarkable ability to perform fluent organization of joint action, achieved by anticipating the motor intentions of others (Sebanz, Bekkering, & Knoblich, 2006). An impressive range of experimental findings, about the underlying neurocognitive mechanisms, support the notion that a close perception-action linkage provides a basic mechanism for real-time social interactions (Newman-Norlund, van Schie, van Zuijlen, & Bekkering, 2007; Wilson & Knoblich, 2005). A key idea is that action observation leads to an automatic activation of motor representations that are associated with the execution of the observed action. It has been advanced that this motor resonance system supports an action understanding capability (Blakemore & Decety, 2001; Fogassi et al., 2005; Fogassi & Rizzolatti, 2013). By internally simulating action consequences using their own motor repertoire the observer may predict the consequences of others' actions. Direct physiological evidence for such perception-action systems came with the discovery of the so-called mirror neurons in the premotor cortex of the macaque monkey (for a review see Rizzolatti and Craighero (2004)). These neurons are a particular class of visuomotor neuron that are active during the observation of goal-directed actions (such as reaching, grasping holding or placing an object) and communicative actions, and during execution of the same class of actions (Ferrari, Gallese, Rizzolatti, & Fogassi, 2003; Rizzolatti, Fogassi, & Gallese, 2001). Later, Fogassi et al. (2005) discovered mirror neurons in the area PF/PFG that code the (ultimate) goal of an observed action sequence, e.g. 'reaching-grasping-placing'. A detailed review and discussion regarding the anatomical and functional organization of the premotor and parietal areas of monkeys and humans, and also, how the mirror neuron mechanism is involved in understanding the action and intention of others in imitative behavior can be found in Rizzolatti, Cattaneo, Fabbri-Destro, and Rozzi (2014).

More recently, Bekkering et al. (2009) have investigated the role of the human mirror neuron system in joint action. Specifically, they have assessed through neuroimaging and behavioral studies, the role of the mirror neuron system while participants prepared to execute complementary actions, and compared with imitative actions. They have shown that the human mirror neuron system may be more active during the preparation of complementary actions than during imitative actions (Newman-Norlund et al., 2007), suggesting that it may be essential in dynamically coupling action observation on to (complementary) action execution, and that this mapping is much more flexible than previously thought (Poljac et al., 2009; van Schie et al., 2008).

There is also good evidence in neuroscience studies that a facial expressions mirroring system exists. The work by Leslie, Johnson-Frey, and Grafton (2004) shows results that are consistent with the existence of a face mirroring system located in the right hemisphere (RH) part of the brain, which is also associated with emotional understanding (Ochsner & Gross, 2005). Specifically, the right hemisphere premotor cortex may play a role in both the generation and the perception of emotionally expressive faces, consistent with a motor theory of empathy (Leslie et al., 2004). That mirror neuron activation is associated with facial emotion processing has also been supported in a more recent study by Enticott, Johnston, Herring, Hoy, and Fitzgerald (2008). van der Gaag, Minderaa, and Keysers (2007) present a more in-depth study on the role of mirror neurons in the perception and production of emotional and neutral facial expressions. The understanding of other people from facial expressions is a combined effort of simulation processes within different systems, where the somatosensory, motor and limbic systems all play an important role. This process might reflect the translation of the motor program, emotions and somatosensory consequences of facial expressions, respectively (Keysers & Gazzola, 2006). The simulation processes in these individual systems have been previously described in the literature (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Keysers et al., 2004; Wicker et al., 2003). Specifically, and at a neuronal level, premotor mirror neurons might resonate the facial movement and its implied intention (Carr et al., 2003; Iacoboni et al., 2005), insula mirror neurons might process the emotional content (Wicker et al., 2003), and somatosensory neurons might resonate proprioceptive information contained in the observed facial movement (Keysers et al., 2004). This process is coherent with current theories of facial expression understanding (Adolphs, 2006; Carr et al., 2003; Leslie et al., 2004), pointing out that different brain systems collaborate during the reading of facial expressions, where the amount and pattern of activation is different depending on the expression being observed.

Current works that take a neuro/bio inspired approach for the integration of emotions into architectures for artificial intelligence focus on more low level aspects of emotions. The work by Talanov, Vallverdu, Distefano, Mazzara, and Delhibabu (2015) explores how to produce basic emotions by simulating neuromodulators in the human brain, and applying it to computational environments for decision making. Lowe, Herrera, Morse, and Ziemke (2007) explore how a dynamical systems perspective can be combined with an approach that views emotions as attentional dispositions.

In previous work, we have developed a cognitive control architecture for human-robot joint action that integrates action simulation, goal inference, error detection and complementary action selection (Bicho, Erlhagen, Louro, & Costa e Silva, 2011; Bicho, Erlhagen, Louro, Costa e Silva, Silva, & Hipólito, 2011), based on the neurocognitive mechanisms underlying human joint action (Bekkering et al., 2009). For the design and implementation, our group takes a neurodynamics approach based on the theoretical framework of Dynamic Neural Fields (DNFs) (Erlhagen & Bicho, 2006, 2014; Schöner, 2008). The robot is able to successfully collaborate with a human partner in joint tasks (e.g. construction tasks, assisting to drink), but thus far has paid attention only to hand actions and to the task itself.

This work extends the cognitive architecture by endowing the robot with the ability to detect and interpret facial expressions of the human co-actor, in order to infer his emotional state. The focus is on – free floating – basic emotions (e.g. happiness, sadness, neutral, anger-irritation, fear) that function as rapid appraisals of situations in relation to goals, actions and their consequences (Oatley & Johnson-Laird, 1987, 2014). From the integration of reading motor intentions and emotional states into the robot's control architecture, we are endowing the robot with the required high level cognitive skills to be a more intelligent socially aware partner.

The results illustrate how the human emotional state influences various aspects of the robot behavior. We show how it influences the decisions that the robot makes, e.g. the same goal directed hand action in the same context but with a different emotional state has a bias on the robot's decisions. We show how the emotional state can have a role in the robot's error handling capabilities, specifically, how the same error is treated in different ways. Also, how the robot can use its emotional expressive capabilities to deal with a human partner persisting in error. And finally, how the human's emotional state can influence the time it takes for the team to complete the joint construction task.

The rest of the paper is organized as follows: In the next section, we present an overview of the cognitive control architecture that integrates emotions to modulate the distributed decision making process of an intelligent robot cooperating with a human in joint tasks. In the model details section, we show how the theoretical framework of dynamical neural fields was used to implement the described control architecture. Next, the joint task that will be carried out by the human-robot team and details on the anthropomorphic robot ARoS utilized in the experiments are presented. The effects of the human partner's emotional states in the robot's behavior are presented and described in the results section. The paper ends with a discussion of the presented results and an outlook for future work. The supplemental material provides additional model details, which also includes a list of all parameter values.

2 Cognitive architecture for human-robot joint action modulated by emotional states

Figure 1 presents a sketch of the multi-layered dynamic neural field architecture for joint action consisting of various neural populations. It reflects the neurocognitive mechanisms that are believed to support human joint action (Bekkering et al., 2009) and shared emotional facial expressions (Carr et al., 2003; Iacoboni et al., 2005; Wicker et al., 2003).Every neural population can receive input from multiple connected populations that may be located in different layers.

Ultimately, the architecture implements a context-dependent mapping between observed action and executed action (Erlhagen, Mukovskiy, & Bicho, 2006a; Poljac et al., 2009; van Schie et al., 2008). The fundamental idea is that the mapping takes place on the level of abstract motor primitives defined as whole object-directed motor acts like reaching, grasping, placing,

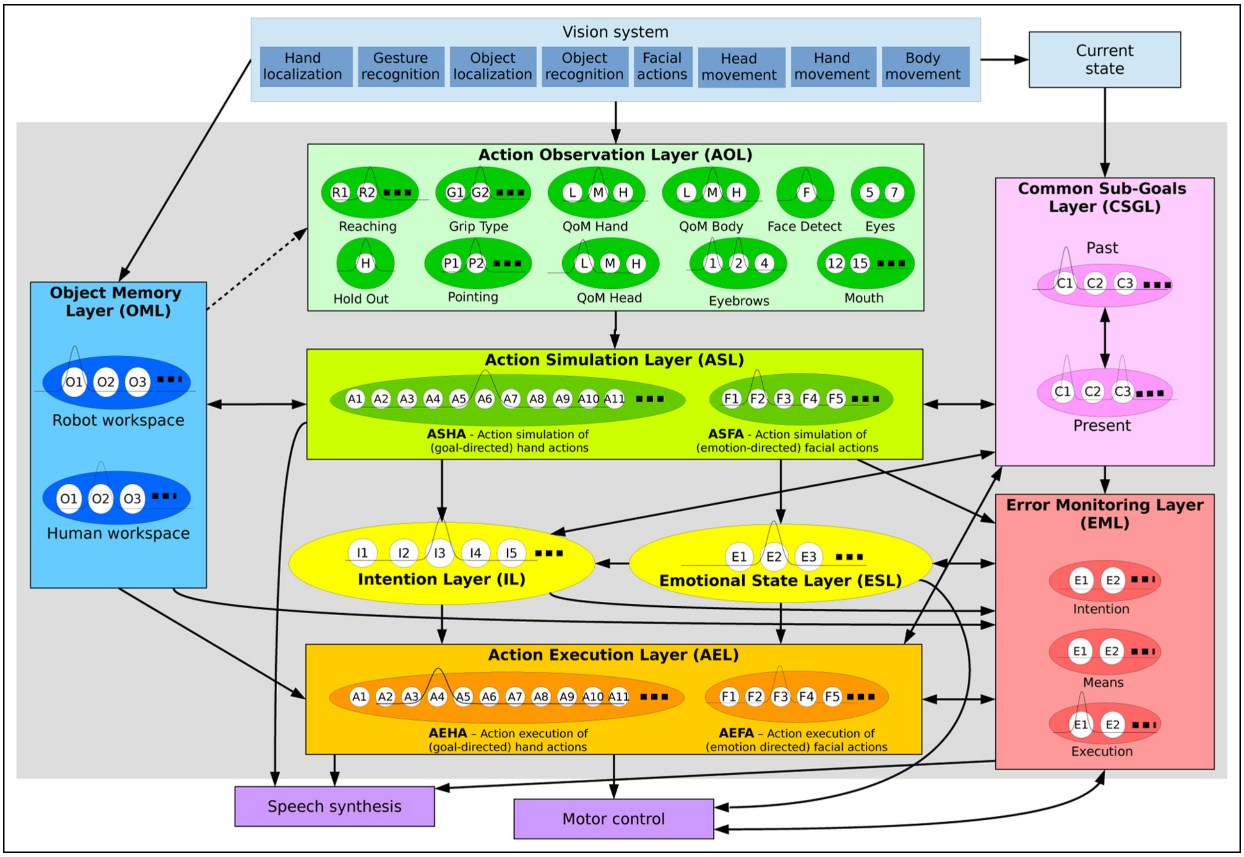

Vision system
Hand localization
Gesture recognition
Object localization
Object recognition
Facial actions
Head movement
Hand movement
Body movement
Current state
Action Observation Layer (AOL)
R1 R2
Reaching
G1 G2
Grip Type
L M H
QoM Hand
L M H
QoM Body
F
Face Detect
5 7
Eyes
H
Hold Out
P1 P2
Pointing
L M H
QoM Head
1 2 4
Eyebrows
12 15
Mouth
Object Memory Layer (OML)
O1 O2 O3
Robot workspace
O1 O2 O3
Human workspace
Action Simulation Layer (ASL)
A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11
F1 F2 F3 F4 F5
ASHA - Action simulation of (goal-directed) hand actions
ASFA - Action simulation of (emotion-directed) facial actions
Common Sub-Goals Layer (CSGL)
Past
C1 C2 C3
C1 C2 C3
Present
Error Monitoring Layer (EML)
E1 E2
Intention
E1 E2
Means
E1 E2
Execution
I1 I2 I3 I4 I5
Intention Layer (IL)
E1 E2 E3
Emotional State Layer (ESL)
Action Execution Layer (AEL)
A1 A2 A3 A4 A5 A6 A7 A8 A9 A10 A11
F1 F2 F3 F4 F5
AEHA – Action execution of (goal-directed) hand actions
AEFA – Action execution of (emotion directed) facial actions
Speech synthesis
Motor control

Figure 1. Schematic view of the cognitive architecture for joint action. It implements a flexible and dynamic mapping from observed hand and facial actions (AOL) onto complementary actions and emotional expressive faces (AEL) taking into account the inferred motor intention (IL), the inferred emotional state of the partner (ESL), detected errors (EML), contextual cues (OML) and shared task knowledge (CSGL). The goal/intention inference and emotional state inference capabilities are based on motor simulation (ASL). In the supplemental material a picture containing all the synaptic links between the pools of neurons can be seen.

attaching or plugging. These primitives encode the motor act in terms of an observable end state or goal rather than in terms of a detailed description of the movement kinematics (Rizzolatti & Craighero, 2004; Schaal, 1999). Also, there is evidence of premotor mirror neurons that might resonate to the facial movement and its implied intention (Carr et al., 2003; Iacoboni et al., 2005).

The cognitive architecture used in this work has its core in the work presented in Bicho, Erlhagen, Louro, and Costa e Silva (2011); Bicho, Erlhagen, Louro, Costa e Silva, Silva, and Hipólito (2011), where only hand actions have been considered. In the work reported here, additional layers have been added to reflect the extra information (for example, observed facial actions) used by the robot in its distributed decision making process. That is, the inferred partner's emotional state, inferred goal and selection of an adequate complementary behavior. The latter includes selection of an appropriate goal-directed hand-action and facial-action set to be performed and displayed by the robot.

An observed hand movement that is recognized by the vision system as a particular primitive (e.g. reach, grasp with top grip or side grip) is represented in the Action Observation Layer (AOL). This layer also incorporates neural populations that code facial actions (e.g. raise inner part of eyebrows, lip corners down) identified by the vision system, as well as a qualitative quantification of the movement of the hand, head and body.

The Action Simulation Layer (ASL) implements the idea that by automatically matching the co-actor's hand and facial actions onto its own sensorimotor representations without executing them, the robot may simulate the ongoing action and facial expression and their consequences. ASL consists of two DNFs layers. One DNF with neural populations representing entire chains of hand action primitives that are in the motor repertoire of the robot (e.g. reaching-grasping-placing or reaching-grasping-holding out) – named Action Simulation of Hand Actions (ASHA) layer. The other DNF with neural populations representing facial action sets (e.g. lift eyebrows – open mouth – express surprise) – named the Action Simulation of Facial Actions (ASFA) layer.

In the case of goal-directed hand actions, the chains are linked to neural representations of specific goals or end states (e.g. attach wheel to base) that are represented by neural populations in the Intention Layer (IL). Facial action sets are linked to specific emotional states represented in the Emotional State Layer (ESL). This layer influences the IL, since an emotional state can play a role in identifying an intention. If a chain (in ASL) is activated by observation of its first motor act, the observer may be able to predict future motor behavior and the consequences of the whole action sequence before its complete execution, effectively inferring the partner's motor intention ahead of time. However, in some situations the observation of the first motor act per see, might not be enough if the motor act being observed is part of multiple chains. Likewise, a single facial action unit may by part of several different facial expressions. In order to disambiguate, additional contextual information is required to be integrated into the inference process (Erlhagen, Mukovskiy, Chersi, & Bicho, 2007). The Object Memory Layer (OML) that represents the robot's memorized knowledge about the location of the different objects in the two working areas, plays a key role. Another important source of information, vital to the success of the task is the shared task knowledge about the possible sequences of sub-tasks (e.g. assembly steps in a joint assembly task). This information is provided by the Common Subgoals Layer (CSGL), which contains neural populations representing the subgoals of the task (e.g. individual assembly steps) that are currently available for the team. For example, in the case of an assembly task, the subgoals are continuously updated in accordance with the assembly plan based on visual feedback about the state of the construction and the inferred goal of the co-actor (represented in the IL). Neurophysiological evidence suggests that in sequential tasks distinct populations in Pre-Frontal Cortex (PFC) represent already achieved subgoals and subgoals that are still to be accomplished (Genovesio, Brasted, & Wise, 2006). In line with this finding, CSGL contains two connected DNF layers with population representations of past and future events. The connections linking the neural populations in one DNF to the other DNF encode the different serial order of subgoals of the task (see Sousa, Erlhagen, Ferreira, and Bicho (2015) for how these can be learned by tutor demonstration and feedback).

The Action Execution Layer (AEL) contains populations representing the same goal-directed action sequences and facial actions sets that are present in the ASL. Hence, all the goal-directed action sequences and facial actions sets that the robot is able to identify (populations present in the ASL), are the same actions that the robot is able to execute (populations in AEL). This implements a mirror neuron mechanism, where the robot understands a goal-directed action or a set of facial actions because it also knows how to execute them. Each population in AEL integrates inputs from the IL, ESL, OML and CSGL to select among all possible actions the most appropriate complementary behavior. Specifically, the ESL (representing the inferred co-actor's emotional state) contributes to the selected emotional state to be expressed by the robot. The mapping from ESL to AEL implements some aspects of shared emotions in joint action (Michael, 2011). For example, if the human is in a positive state (Happy) the robot expresses also a Happy expression. This effect is known as emotion contagion and occurs when one person's perception of another person's emotional expression can have effects that are relevant to an interaction, if the perceiver thereby enters into an affective state of the same type (Michael, 2011). In fact, one important way in which emotion contagion can function as a coordination smoother within joint action is by means of alignment. A key benefit of alignment is manifested by the likelihood of the increase in the participants' motivation to act jointly, since people tend to find other people with similar moods to be warmer and more cooperative, and prefer to interact with them (Locke & Horowitz, 1990).

The implemented context-sensitive mapping from observed actions on to-be executed complementary actions guarantees a fluent team performance if no errors occur (Bekkering et al., 2009). However, if an unexpected or erroneous behavior of the partner occurs, neural populations in the Error Monitoring Layer (EML) are sensitive to a mismatch on the goal level, on the level of action means to achieve a valid sub-goal, and on the level of motor execution. This allows the robot to detect errors in user's intention and/or action means to achieve a subgoal, and execution errors (e.g. a piece the robots was moving falls down), and thus allows the robot to efficiently cope with such situations. The ESL also plays a role in influencing the EML, implementing some aspects of shared emotions in joint action. Michael (2011) talks about the various types of shared emotions present in joint action tasks. One of the types of shared emotions used in our work is the emotion detection, which can facilitate prediction and monitoring of the partner's actions, and can also act as a signaling function. For example, a positive emotional expression, such as a smile, may signal approval of another participant's action or proposed action (Michael, 2011). This way in our joint task, if the human partner is in a positive (e.g. Happy) emotional state, this might mean she/he is committed and engaged in the task, and thus it is not the probable that partner will make errors. In this situation, the processing of the DNFs detecting errors in action means and intention are disabled, since this allows to decrease the computational efforts the robot's decision making processes, and hence the time it takes to select a complementary action is accelerated. In addition, if the human is in a positive emotional state, it means that she/he is comfortable with the robot and therefore one can increase the robot's movement velocity. Altogether this allows for the joint task to be completed in less time. Conversely, if the robot infers the human is in a negative emotional state (e.g. Sad), then it might be that the human is (also) not fully committed in the task and hence can be more prone to errors. The detection of a negative emotional state is used as a signal to activate the full processing of the EML. This is consistent with the modeling study by Grecucci, Cooper, and Rumiati (2007), who proposed a computation model of action resonance and its modulation by emotional stimulation, based on the assumption that aversive emotional states enhance the processing of events. This way, the robot is fully alert to all types of errors that can occur during the execution of the task, being able to anticipate them, and act before they occur. This is fundamental for efficient team behavior.

Through direct connections to the AEL, population activity in the EML may bias the robot's planning and decision process by inhibiting the representations of complementary actions normally linked to the inferred goal and exciting the representations of a corrective response. In order to efficiently communicate detected errors to the human partner a corrective response may consist of a manual gesture like pointing or a verbal comment to attract the partners' attention (Bicho, Louro, & Erlhagen, 2010).

Finally, it is important to highlight the connections from the ESL to both the AEL and motor control. These connections implement the idea that perceived emotions play an important role not only in an early stage, during decision making and action preparation (AEL layer) of a complementary action, but also the latter may affect the execution at the kinematics level (motor control). This is motivated by recent studies in neuroscience by Ferri, Campione, Dalla Volta, Gianelli, and Gentilucci (2010); Ferri, Stoianov, et al. (2010), having investigated the link between emotion perception and action planning & execution within a social context. In summary, they have demonstrated that assisting an actor with a fearful expression requires more smooth/slow movements, compared to assisting an actor with a positive emotional (e.g. Happy) state.

3 Dynamical neural fields as a theoretical framework for the implementation

Dynamical Neural Fields (DNFs) provide a theoretical framework to endow artificial agents with cognitive capacities like memory, decision making or prediction (Erlhagen & Bicho, 2006; Schöner, 2008). DNFs are based on dynamic representations that are consistent with fundamental principles of cortical information processing, implementing the idea that task-relevant information about action goals, action primitives or context is encoded by means of activation patterns of local populations of neurons.

Each layer of the model is formalized by one or more DNFs. The basic units present in these models are local neural populations with strong recurrent interactions that cause non-trivial dynamic behavior of the population activity. One important property that can be observed, is that population activity initiated by time-dependent external signals may become self-sustained in the absence of any external input. This property of the population dynamics behaves like an attractor state and is thought to be essential for organizing goal-directed behavior in complex dynamic situations, they allow the nervous system to compensate for temporally missing sensory information or to anticipate future environmental inputs.

The presented DNF based architecture for joint action is built as a complex dynamic system in which activation patterns of neural populations in the various layers can appear and disappear continuously in time as a consequence of input from connected populations and external sources to the network (e.g. vision, speech) and as defined by field dynamics.

A particular form of DNF first analyzed by Amari (1977), was used for modeling. In each layer i, the activity $u_i(x,t)$ at time t of a neuron at field location x is described in equation (1) (for mathematical details see Erlhagen & Bicho, 2014)

$$\tau_i \frac{\delta u_i(x,t)}{\delta t} = -u_i(x,t) + S_i(x,t) + \int w_i(x-x')f_i(u_i(x',t))dx' - h_i \tag{1}$$

where the parameter $\tau_i > 0$ defines the time scale and $h_i > 0$ the resting level of the field dynamics. The integral term describes the intra-field interactions defined to be of lateral inhibition type described by equation (2)

$$w_i(x) = A_i \exp\left(\frac{-x^2}{2\sigma_i^2}\right) - w_{\text{inhib},i} \tag{2}$$

where $A_i > 0$ describes the amplitude, $\sigma_i > 0$ the standard deviation of the Gaussian. The inhibition ($w_{\text{inhib},i} > 0$) is assumed to be constant, only sufficiently activated neurons contribute to interaction. The threshold $f_i(u)$ is a sigmoidal function with slope parameter β and threshold u_0, described in equation (3)

$$f_i(u_i) = \frac{1}{1 + \exp[-\beta(u_i - u_0)]} \tag{3}$$

The model parameters are adjusted to ensure that the field dynamics are bi-stable (Amari, 1977), allowing the attractor state of a self-stabilized activation pattern to coexist with a stable homogeneous activation distribution, that represents the absence of specific information (resting level $-h_i$). When the input ($S_i(x,t)$), to a local population is sufficiently strong, the homogeneous state loses stability and a localized pattern in the dynamic field evolves, however, weaker external signals lead to a subthreshold, input-driven activation pattern in which the contribution of the interactions is negligible.

DNFs enable us to also implement a working memory function through the existence of self-stabilized activation patterns. The existence of a single, self-stabilized pattern of activation in a dynamic field is also closely linked to decision making. In the different layers of the architecture subpopulations – encoding different hand action chains (ASHA), facial action sets (ASFA), goals (IL), complementary goal directed hand actions (AEHA), complementary facial actions (AEFA) and detected errors (EML) – interact through lateral inhibition. These inhibitory interactions lead to the suppression of activity below resting level in competing neural pools whenever a certain subpopulation becomes activated above the threshold. The population for which the summed input from connected populations is highest wins the competition process.

To represent and memorize simultaneously the location of several objects, and multiple common subgoals, the spatial ranges of the lateral interactions in layers OML and CSGL were adapted to avoid a direct competition between different populations, enabling these layers to support a multi-peak solution. The updating of the memorized information is performed by defining proper dynamics for the inhibition parameter, h_i, of the population dynamics (Bicho, Mallet, & Schöner, 2000).

The summed input from connected fields u_l is given as $S_i(x,t) = k\sum_l S_l(x,t)$. The parameter k scales the total input to a certain population relative to the threshold for triggering a self-sustained pattern. This guarantees that the inter-field couplings are weak compared to the recurrent interactions that dominate the field dynamics (for details see Erlhagen and Bicho (2006)). The scaling also ensures that missing or delayed input from one or more connected populations will lead to a subthreshold activity distribution only. The input from each connected field u_l is modeled by a Gaussian function described in equation (4)

$$S_l(x,t) = \sum_m \sum_j a_{mj} c_{lj}(t) \exp\left(\frac{-(x-x_m)^2}{2\sigma^2}\right) \qquad (4)$$

where $c_{lj}(t)$ is a function that signals the existence or evolution of a self-stabilized activation pattern in u_l centered at position y_j, and a_{mj} is the inter-field synaptic connection between subpopulation j in u_l to subpopulation m in u_i. Inputs from external sources (e.g. vision) are also modeled as Gaussians. As an example, Figure 2 shows the input from a connected population j in layer u_l connected to a target population m in layer u_i, modeled by a Gaussian function. This input is applied whenever the activation in population j is above the threshold for a self-stabilized activation peak.

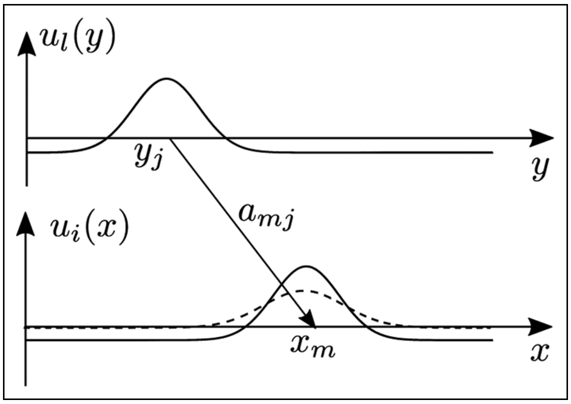


Figure 2. Schematic view of two connected DNFs. For simplicity only one inter-field connection is shown. The activation pattern in field u_l centered at y_j (representing the center of the population j) propagates through inter-field synaptic link a_{mj} to subpopulation m in field u_i and creates a Gaussian input (dashed-line) as defined by equation (4).

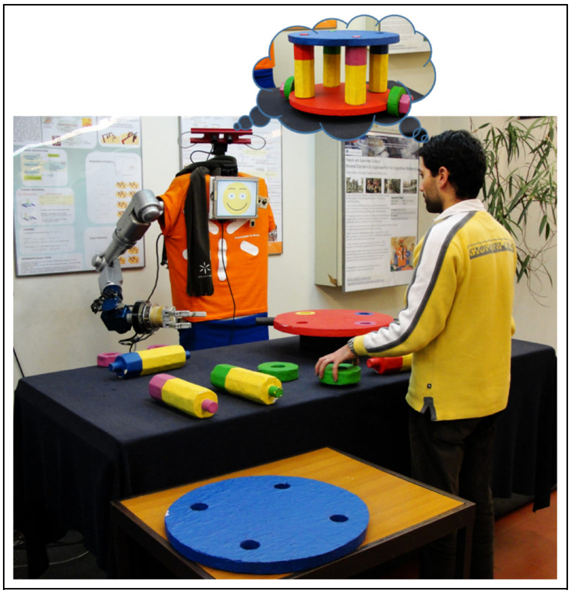


Figure 3. Anthropomorphic robot ARoS and the scenario for the joint construction task.

4 Setup of the human-robot experiments

To test the dynamic neural field architecture for human-robot collaboration we have chosen a joint assembly paradigm in which the team has to construct a toy 'vehicle' from parts that are initially distributed on a table (see Figure 3). The toy 'vehicle' is composed of three sections. The lower section consists of a round platform with an axle on which two wheels have to be attached and then each fixed with a nut. In the middle section, four columns that differ in their color have to be plugged into specific holes in the platform. Finally, at the top section, the placing of another round object on top of the columns finishes the task. The parts have been designed to facilitate the workload of the vision and the motor system of the robot.

The working areas of the human and the robot do not overlap, the spatial distribution of the parts on the table obliges the team to coordinate handing-over sequences. It is assumed that each team mate is responsible for assembling one side of the toy, although, some assembly steps may require that one actor helps the other by holding a part still in a certain position. Both the human and the robot can perform the same assembly actions.

It is assumed that both partners know the construction plan and keep track of the subtasks that have been already completed by the team. The prior knowledge about the sequential execution of the assembly work is represented in layer CSGL of the DNF-architecture, by connections between populations encoding subsequent assembly steps (for how these connections could have been established through learning by demonstration and tutor's feedback see Sousa et al. (2015)). Since the desired end state does not uniquely define the logical order of the construction, at each stage of the construction the execution of several subtasks may be simultaneously possible. The main challenge for the team is thus to efficiently coordinate in space and time the decision about actions to be performed by each of the team mates. The task is complex enough to show the impact of goal inference, emotional state inference, action understanding and error monitoring on complementary action selection.

The robot ARoS used in the experiments has been built in our lab (Silva, Bicho, & Erlhagen, 2008). The robot consists of a stationary torus, on which a 7 DOFs AMTEC arm (Schunk GmbH) with a 3-fingers dexterous gripper (Barrett Technology Inc.), a stereo camera rig mounted on a pan-tilt unit, a PS3Eye camera with an adapted lens, are mounted. In addition, the robot has a monitor located on the chest, which is used to produce expressive faces in order to improve interaction with the human. The expressive faces the robot is able to produce, are performed using the same facial action primitives (Action Units) that can be recognized by the vision system. A speech synthesizer (Microsoft Speech SDK 5.1) allows the robot to communicate the result of its reasoning to the human user.

The vision system is composed of two independent systems, that provide distinct information. The first system is a stereo camera rig mounted on a pan-tilt unit and provides information about objects (class and pose), hands (position, velocity, and classification of (static) hand gestures, such as grasping and communicative gestures like pointing) and the state of the construction task. The information about the objects combines color based search algorithms with stereo data to extract the desired information. Concerning the human hands, the vision system combines a color based search algorithm with invariant moments (Hu, Ming-Kuei, 1962) to distinguish the different gestures. The second system is composed of a single camera (PS3Eye) with an adapted lens dedicated to the human face. It uses the faceAPI library from SeeingMachines to extract information from the face in the form of Action Units. The system uses the Facial Action Coding System created by Ekman and Friesen (1978), as a coding system to describe facial actions.

For the control of the arm-hand system we applied a global planning method in posture space that allows us to integrate optimization principles derived from experiments with humans (Costa e Silva, Costa, Bicho, & Erlhagen, 2011). The goal is to guarantee collision free robot motion that is perceived by the human user as smooth and goal-directed.

5 Results

To validate the dynamic neural field architecture we designed and conducted real-time human-robot experiments in scenarios of the joint construction task described above.

For better understanding we divided the construction task in three logic stages, lower section (wheels and nuts), middle section (columns) and top section (Top Floor).

The focus is on showing and explaining how decision making and error detection are affected by the human partner's emotional state. In all cases, the initial spatial distribution of parts forces both actors to demand and hand-over parts. There is no verbal communication from the human to the robot. This obliges the robot to continuously monitor and interpret the actions of its co-worker. Both the human and the robot can manipulate the parts (e.g. plug a wheel on the axle). The robot uses speech to communicate to the human partner the outcome of the goal inference and decision making processes implemented in the dynamic neural field model. As our studies with naive users show, this basic form of verbal communication facilitates natural and fluent interaction with the robot (Bicho et al., 2010).

To validate the high level cognitive control architecture, five different experiments were designed. Each experiment addresses a specific feature with different scenarios, in order to better understand how the partner's emotional state can affect the robot's behavior. Experiment 1 explores how the robot's decisions can be influenced by the partner's emotional state. Experiment 2 shows how the inferred user's emotional state can influence how the robot detects and handles errors during task execution. Experiment 3 shows how the robot, by expressing emotional facial expressions, deals with a human persisting in an error. Experiment 4 presents a comparison of the influence of the user's emotional state in the time the task takes to be performed. Finally, Experiment 5 shows the dynamic nature of the architecture in a longer interaction, where the robot adjusts its behavior in real time, in response to the change of the human emotional state.

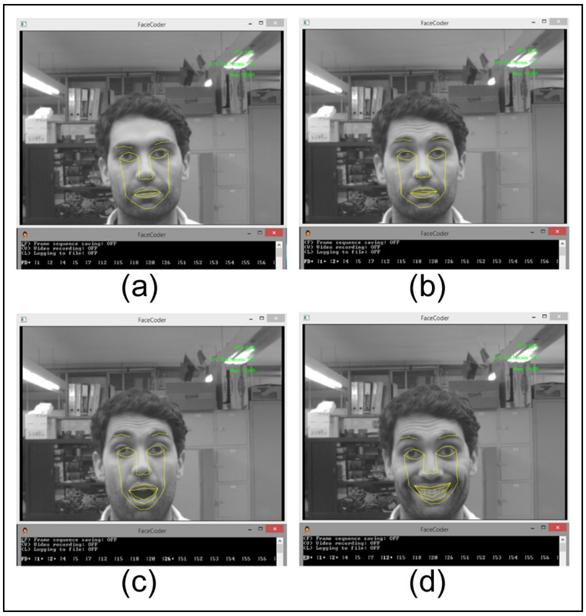
The graphics presented for each scenario, show the time evolution of fields activity in some layers of the control architecture. It would be impractical to show the evolution in all layers of the architecture, hence, only key layers for each scenario will be presented.

The main contribution of this work is the integration of emotions into the robot's cognitive architecture. Hence, before presenting the interaction results, we provide details on how the information acquired by the vision system regarding the human face is handled. Figure 4 presents snapshots of the analysis performed by the system developed for this robot. A dedicated camera placed on the robot acquires an image of the face, which is then processed by combining the library faceAPI from SeeingMachines with some post-processing algorithms implemented using the OpenCV library. The system is continuously processing (at 60 fps) and coding the face according to the FACS (Ekman, Friesen, & Hager, 2002), resulting in a real-time description of the face with Action Units (AUs) (see Appendix Table 1).

The entry point in the architecture for the information provided by the vision system is the Action Observation Layer (AOL). Three DNFs in this layer are responsible for representing information about detected facial muscle movements that are associated to the eyes, eyebrows and mouth. Figure 5 shows the time evolution of the DNFs involved in the processing of this visual information, and the simulation and inference of the user's emotional state (layers AOL, ASFA and ESL respectively).

On top, and regarding AOL, one can see a DNF, $u_{\mathrm{AOL\_FaceDetect}}(x,t)$, that codes the presence (or absence) of a human face and the three DNFs responsible for representing AUs related to the eyebrows ($u_{\mathrm{AOL\_Eyebrows}}(x,t)$), mouth ($u_{\mathrm{AOL\_Mouth}}(x,t)$) and eyes ($u_{\mathrm{AOL\_Eyes}}(x,t)$). These fields provide input $S_{\mathrm{ASFA}}(x,t)$ to the DNF $u_{\mathrm{ASFA}}(x,t)$ that contains neural populations that respond or not to the presence of the several AUs detected. The field activity $u_{\mathrm{ASFA}}(x,t)$ provides the input to the DNF in ESL, $u_{\mathrm{ESL}}(x,t)$, which depending on the initial active populations and other dynamic factors, such as time and quantity of (head/hand) movement, produces an activation at the correspondent inferred emotional state. In Appendix Table 3 which combinations of AUs and human movements cause the inference of which emotional state is being shown.

The example presented in Figure 4, starts with a facial expression where no AUs are present. Hence from times T1 to T2, the activity in $u_{\mathrm{AOL\_Eyebrows}}(x,t)$), mouth ($u_{\mathrm{AOL\_Mouth}}(x,t)$) and eyes ($u_{\mathrm{AOL\_Eyes}}(x,t)$ code absence of AUs, while the bump of activity in field $u_{\mathrm{AOL\_Face}}(x,t)$) represents the presence of the human face. During this time interval only this input arrives to $u_{\mathrm{ASFA}}(x,t)$ which produces a pattern of activation that represents solely 'face detected' and thus activity in $u_{\mathrm{ESL}}(x,t)$ produces a bump of activity centered at the emotional state 'Neutral'.



```
(a)
<F> Frame sequence saving: OFF
<V> Video recording: OFF
<L> Logging to file: OFF

FD• !1  !2  !4  !5  !7  !12  !15  !18  !20  !26  !51  !52  !53  !54  !55  !56

(b)
<F> Frame sequence saving: OFF
<V> Video recording: OFF
<L> Logging to file: OFF

FD• !1• !2• !4  !5  !7  !12  !15  !18  !20  !26  !51  !52  !53  !54  !55  !56

(c)
<F> Frame sequence saving: OFF
<V> Video recording: OFF
<L> Logging to file: OFF

FD• !1• !2• !4  !5  !7  !12  !15  !18  !20  !26• !51  !52  !53  !54  !55  !56

(d)
<F> Frame sequence saving: OFF
<V> Video recording: OFF
<L> Logging to file: OFF

FD• !1• !2• !4  !5  !7  !12• !15  !18  !20  !26  !51  !52  !53  !54  !55  !56
```

Figure 4. Face analysis by the vision system.

Next, from times T2 to T3, the human raises its eyebrows (see Figure 4b) producing an activation in $u_{AOL\_Eyebrows}(x,t)$ representing the detection of AUs 1 and 2. As a consequence of the spread of field activation from AOL to ASFA, a bump of activity in $u_{\mathrm{ASFA}}(x,t)$ emerges centered in the population 'Raise eyebrows', which in turn leads to a bump of activity in $u_{\mathrm{ESL}}(x,t)$ representing an inferred emotional state of 'Surprise'.

Afterward, from times T3 to T4, the human then opens the mouth by dropping its jaw, getting coded by the vision system as AUs 1 + 2 + 26 (Figure 4c). This gives rise to several inputs, $S_{\mathrm{ASFA}}(x,t)$, competing for a decision in $u_{\mathrm{ASFA}}(x,t)$. The population representing 'raise eyebrows & mouth open' wins the competition. However, the inferred emotional state, represented in $u_{\mathrm{ESL}}(x,t)$, remains as 'Surprise'. This demonstrates the ability to detect the same emotional state in more than one way.

Finally, in the time interval T4-T5, the human smiles maintaining the eyebrows raised, the resulting expression is coded with AUs 1 + 2 + 12 (Figure 4d). The disappearance of AU 26 and presence of AU 12 changes the competition in $u_{\mathrm{ASFA}}(x,t)$, and ultimately, the winning population in this field then triggers in $u_{\mathrm{ESL}}(x,t)$ a different inferred emotional state, i.e. 'Happy'.

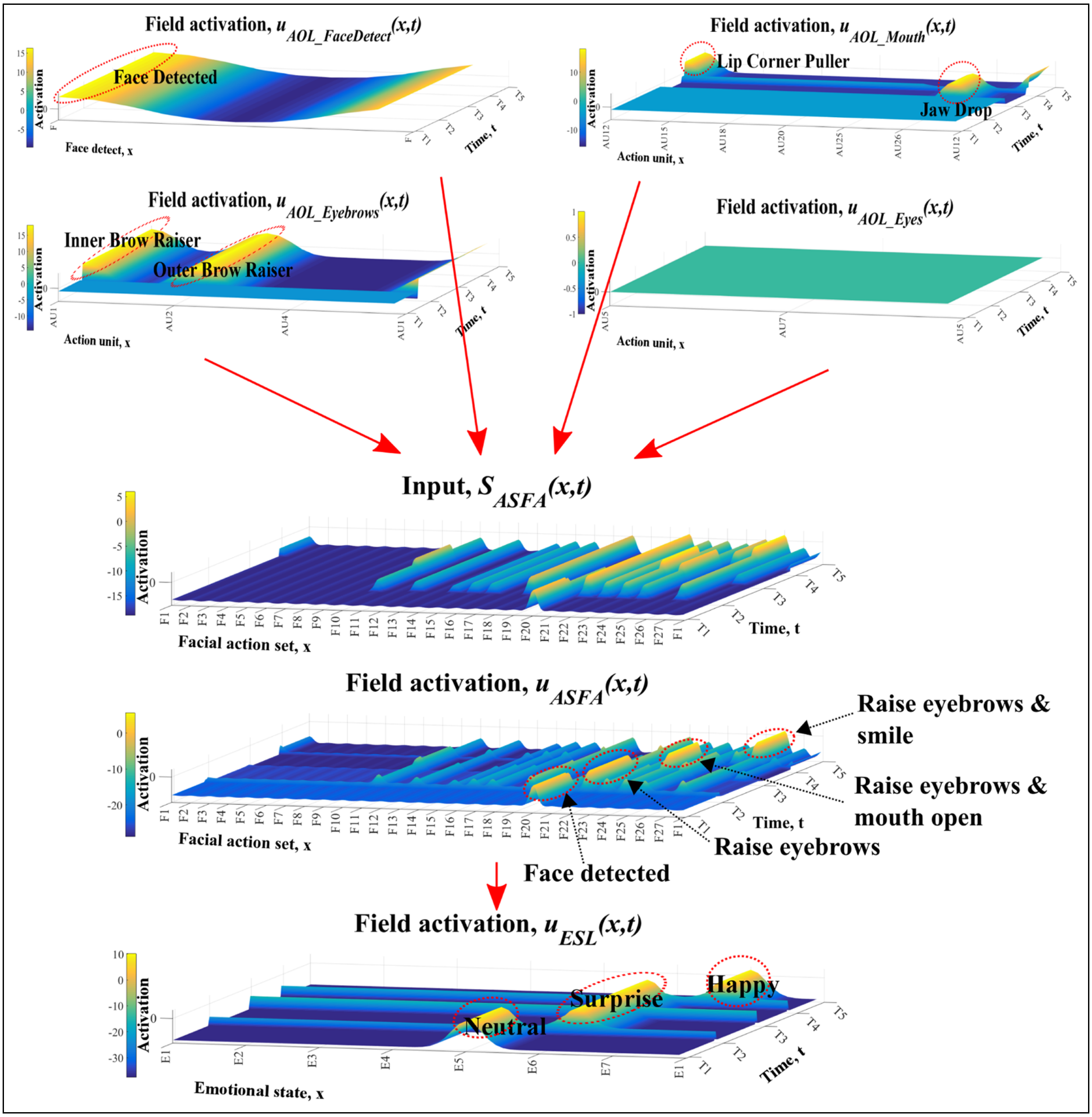


Figure 5. Field activities in layers AOL, ASFA and ESL, in response to the information provided by the vision system regarding the user's facial expressions depicted in Figure 4.

Next, we focus on scenarios addressing several aspects of human-robot joint action.

5.1 Experiment 1: Influence of the human's emotional state in the robot's decisions

Experiment 1 is composed of two scenarios, 1-1 and 1-2, and explores how the same action being performed by the human in the same context of the task, but carried out with a different emotional state, can trigger in the robot different decisions for the complementary action. We used only the construction of the lower section of the task, attach the wheels and fix them with nuts.

The objects disposition for the current experiment is the following:

- robot's workspace: 2x Nut;
- human's workspace: 2x Wheel, Column 1, Column 2, Column 3, Column 4, Top Floor.

For Scenario 1-2, we added a Nut in the human's workspace but hidden from the robot's view. Video snapshots of the human-robot join action in Scenario

1-1 and Scenario 1-2 are shown in Figures 6 and 7 respectively.

In both scenarios the human starts with grasping a wheel (Figures 6(a) and 7(a)) and inserting it (Figures 6(c) and 7(c)). When the human grasps the wheel, the robot infers that he will insert it and decides to handover a nut to the human partner because it is the part he will need next.

The difference in the two scenarios happens here. While in Scenario 1-1 the human continues to display a neutral face (Figure 6(f)), the robot hands over the nut (Figure 6(e)), the human accepts and inserts it (Figure 6(g)). In Scenario 1-2, when the robot verbalizes its decision to handover a nut the human expresses anger (Figure 7(f)). This makes the robot understand that the human does not want the nut (Figure 7(e)), and as a consequence the robot changes its decision and asks the human to hand over a wheel (Figure 7(g)), so that it can insert a wheel on its side of the construction.

In Scenario 1-1, the human working with the robot exhibits a neutral emotional state during the entire interaction, and so, all the decisions made by the robot incorporate no positive nor negative emotions from its human partner. The field activity in the Emotional State Layer (ESL) codes the inferred human's emotional state. Figure 8(a) shows the field activation $u_{\mathrm{ESL}}(x, t)$ in this layer, which always has a bump of activity centered in in the same position ('Neutral') throughout the duration of the task. The change in the inferred emotional state of the human during interaction Scenario 1-2 is presented in Figure 8(b). As can be seen, in the time interval T2-T3, a shift in the bump of activation from 'Neutral' to 'Anger' occurs.

The influence of the human emotional state in the robot's decisions regarding its complementary behavior is clearly demonstrated by analyzing the DNF $u_{\mathrm{AEHA}}(x, t)$ in the Action Execution Layer (Figure 9). This field selects an adequate complementary goal-directed hand action. In Scenario 1-1, after the human grasped the wheel, the robot selected the action of handing over a nut (Figure 9(a): see bump of activation coding 'Give nut'). In Scenario 1-2, the robot initially makes the same decision (Figure 9(b): Field activation, times T1 to T2), but in response to the anger expressed by the human, the robot changes its decision to 'Request a wheel' (Figure 9(b): Field activation, times T2 to T3). The preshaping present in Figures 9(a) and 9(b), of the populations coding the actions 'Point to wheel' and 'Request wheel', means alternative actions the robot could in principle select.

5.2 Experiment 2: Influence of the human's emotional state in the robot's error detection and handling capabilities

Experiment 2 contains two scenarios, 2-1 and 2-2, and explores how the robot deals with errors in reaction to different inferred emotional states. While in Scenario 2-1 the human is displaying a happy expression (Figure 10(b)), in Scenario 2-2 the human has a fearful expression (Figure 11(b)). We show how the same error being committed during the construction task is detected in different ways, influenced by the human emotional state.

The two scenarios start with the lower section of the toy robot assembled, i.e. the Wheels and Nuts are already inserted in the Base. Thus, the next assembly steps consist of mounting the four columns. We impose a specific serial order for plugging the columns: Column 1 → Column 2 → Column 3 → Column 4. The different columns are identified by their color patterns. Given the reachable workspace of the two agents, it happens that Column 1 and Column 4 can only be mounted by the robot, while Column 2 and Column 3 can only be mounted by the human partner.

The object disposition is: robot's workspace: Wheel (inserted), Nut (inserted), Column 4; human's workspace: Wheel (inserted), Nut (inserted), Column 1, Column 2, Column 3, Top Floor.

Both scenarios start in the same way, with the robot requesting the human to handover Column 1 (See Figures 10(a) and 11(a)). However, the human ignores the robot's request and instead grasps Column 3 with the intention to insert it (Figures 10(c) and 11(c)). This is an error because Column 3 cannot yet be mounted.

When the human operator is in a positive emotional state the (expected) probability that he will commit errors is low because this signals that he is engaged in the joint task. In Scenario 2-1, the fact that the human is displaying since the beginning a happy facial expression, has made the robot disable the processing of the DNFs in the Error Monitoring Layer (EML) responsible for detecting user's errors in intention and errors in the means. Thus, although the robot is able to infer, at the moment of grasping, that the intention of the human is to insert Column 1, it is not able to predict that the user's intention/goal is wrong. The human advances and inserts Column 3 (Figure 10(e)). The robot detects that this was error only after the column was plugged (error in execution) and orders the human to correct the error he has made (Figure 10(g)).

In Scenario 2-2, the human is in a negative emotional state, this causes the robot to enable the processing of all the error detection components in the EML. As a consequence, as soon as the human grasps Column 3 to insert, the robot interprets this as an error in intention and prevents the error from occurring (Figure 11(e)).

The main difference in Scenarios 2-1 and 2-2 is due to the expressed emotional state by the human, whose inferred state by the robot is coded in activation of the DNF $u_{\mathrm{ESL}}(x, t)$ in ESL. Figure 12(a) shows a bump of activation representing 'Happy' throughout the duration of Scenario 2-1, while Figure 12(b) shows a bump

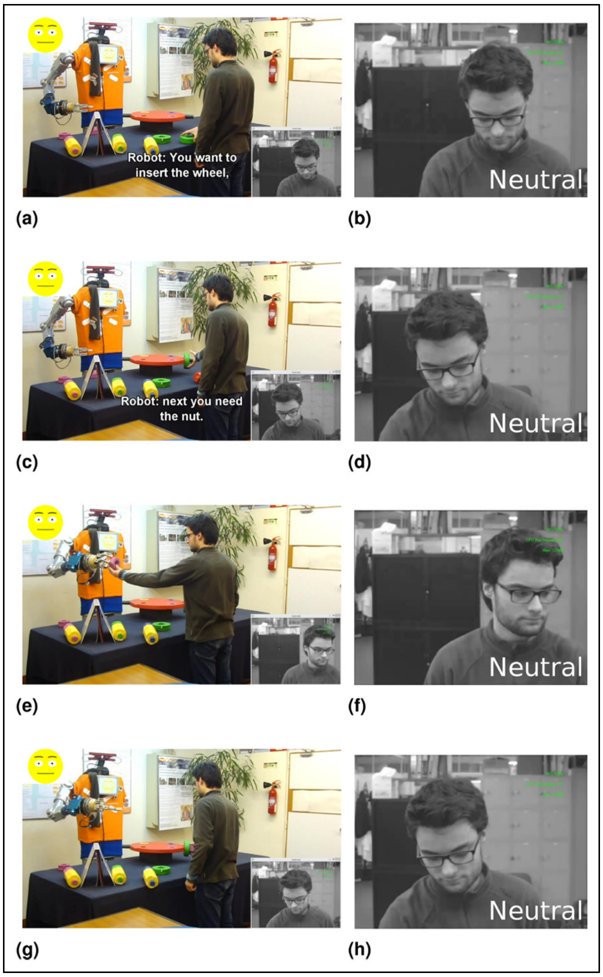

Robot: You want to insert the wheel,
Neutral
(a)
(b)
Robot: next you need the nut.
Neutral
(c)
(d)
Neutral
(e)
(f)
Neutral
(g)
(h)

Figure 6. Video snapshots for scenario 1-1.
Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp1-Scen1_1.html

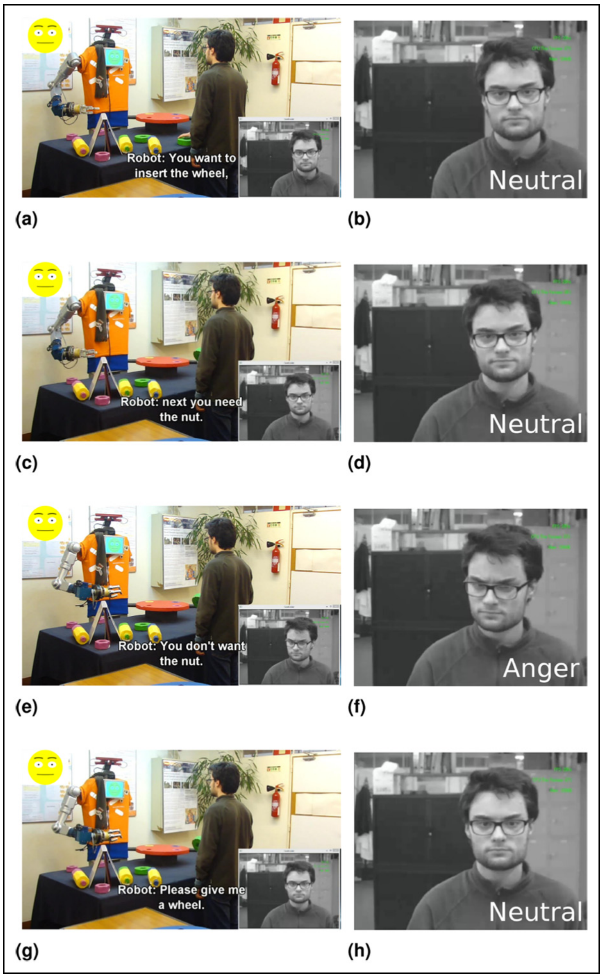

Robot: You want to insert the wheel,
Neutral
(a)
(b)
Robot: next you need the nut.
Neutral
(c)
(d)
Robot: You don't want the nut.
Anger
(e)
(f)
Robot: Please give me a wheel.
Neutral
(g)
(h)

Figure 7. Video snapshots for scenario 1-2.
Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp1-Scen1_2.html

of activation in a different location representing ‘Fear’ during Scenario 2-2.

The influence of the emotional state in the robot’s error detection capabilities can be observed in the EML (Figure 13). While in Scenario 2-1 the robot detected the error ‘Insert Column 3’ as an Execution Error (Figure 13(a)), in Scenario 2-2, the same error was anticipated and detected as an Error in Intention (Figure 13(b)).

The fact that the human was in a happy emotional state prevented the robot from anticipating the error. When the human displays a happy emotional state the robot assumes the construction is going well and disables the detection of errors in intention and errors in means, this way it can accelerate the processing and make decisions faster, with the downside of the robot being unable to anticipate errors the human can commit. However if an error is actually performed, the robot will be able to detect it and issue a warning or corrective order to this fact.

5.3 *Experiment 3: Reaction of the robot to the human’s persistence in error*

In the interaction Scenarios 2-1 and 2-2 described in the previous section, the human partner has accepted the warnings and corrective orders issued by the robot. The robot has never displayed a negative emotional state toward the human partner. Experiment 3 will explore how the robot, by producing expressive faces when required, can react to a stubborn human, and thus induce a change of his behavior/attitude (see video snapshots in Figures 14 and 15).

The situation is the same as the previous Scenario 2-2, but this time the negative emotional state displayed by the human operator is ‘Anger’. All DNFs in EML are therefore activated (their activation can be seen in Figure 16).

The robot starts by requesting Column 1 to the human (Figure 14(a)). However, the human grasps Column 3 (Figure 14(c)) and the robot infers that he

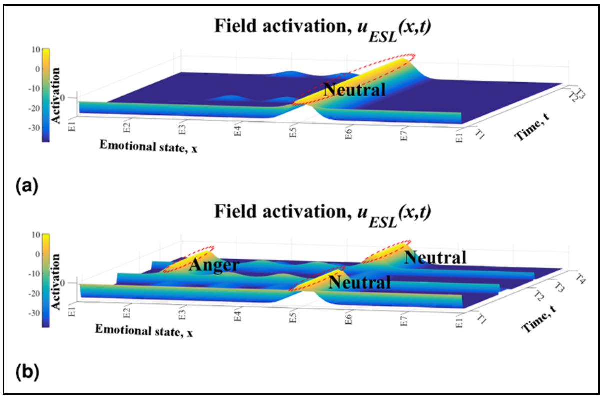


Figure 8. Experiment 1: Emotional state layer. (a) Scenario 1-1: ESL. (b) Scenario 1-2: ESL.

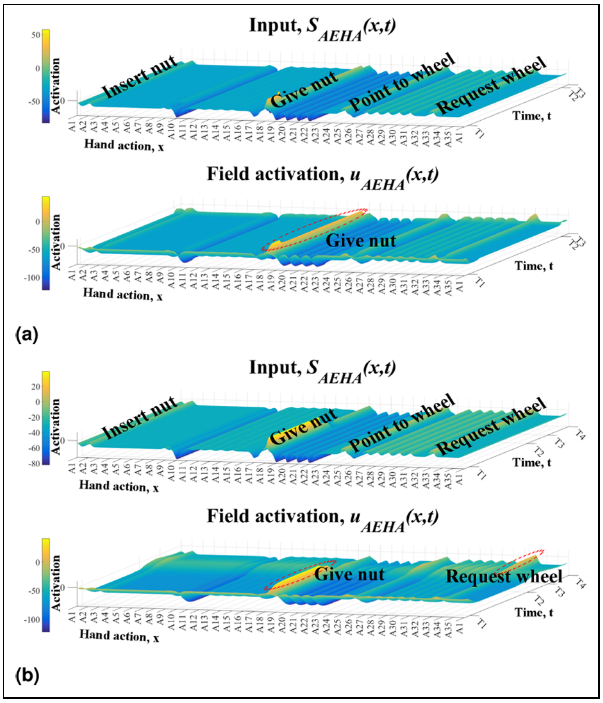


Figure 9. Experiment 1: Action execution layer–goal-directed hand actions. (a) Scenario 1-1: AEHA. (b) Scenario 1-2: AEHA.

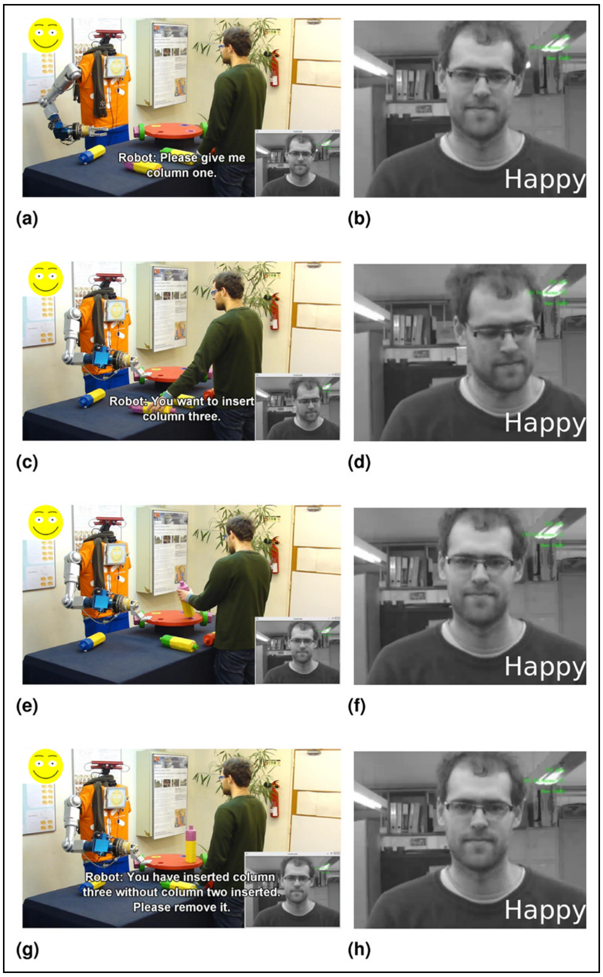

Robot: Please give me column one.
Happy
(a)
(b)
Robot: You want to insert column three.
Happy
(c)
(d)
Happy
(e)
(f)
Robot: You have inserted column three without column two inserted. Please remove it.
Happy
(g)
(h)

Figure 10. Video snapshots for Scenario 2-1.
Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp2-Scen2_1.html

will insert that column (see activation $u_{\mathrm{ASHA}}(x,t)$ in times T2-T3, Figure 17). As before, the robot detects that the human's goal to plug Column 4 is wrong (see activation $u_{\mathrm{EML\_Intention}}(x,t)$ in times T2-T3, Figure 16(a)), and warns that he will commit an error (Figure 14(e)). Despite the warning, the human proceeds to insert Column 3 (Figure 14(g)), and as a consequence the robot now detects it as an execution error and issues a corrective action (see activation $u_{\mathrm{EML\_Exec}}(x,t)$ in times T3-T4, Figure 16(b)). Ignoring the robot, the human persist in the error. In response to this persistence and because the user is in an Angry state (see action $u_{\mathrm{ESL}}(x,t)$ times T1-T5, Figure 19), the robot takes a stand by expressing (also) an angry face (see activation u_{AEFA} in times T5-T6, Figure 18) and explaining again that an error was committed (Figures 14(i) and 15(a)).

Thus far, the robot had never displayed a negative emotion toward the human partner. Thus he gets surprised (Figure 15(d)) by the robot's anger. See activation in $u_{\mathrm{ESL}}(x,t)$ at time T6 (Figure 19). The human finally accepts the robot's correction and removes the inserted column from the Base (Figure 15(c)). The robot then takes a neutral expression (Figure 18, times T7) and requests again that Column 1 is inserted on its side (Figure 15(e)). But because the human expresses surprise in response to the robot's request, the decision of the robot changes from preparing to receive Column 1 to pointing toward to it (Figure 15(g)). This gesture drives the attention of the human operator to the requested column. The human finally grasps and hands over Column 1 to the robot (Figure 15(i)), and the decision of the robot is to receive it. The temporal evolution of these changes in the selected goal-directed hand

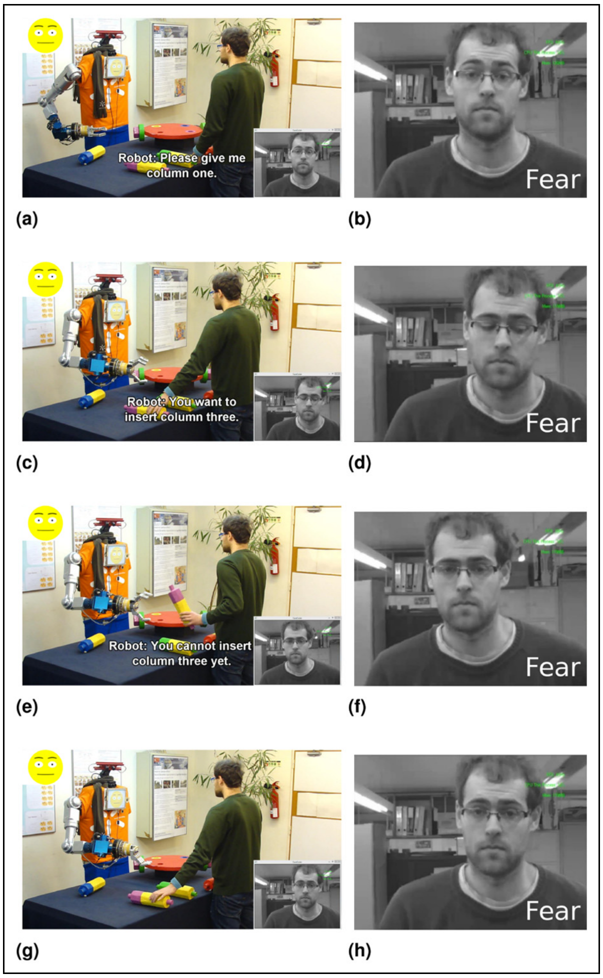

Robot: Please give me column one.
Fear
(a)
(b)
Robot: You want to insert column three.
Fear
(c)
(d)
Robot: You cannot insert column three yet.
Fear
(e)
(f)
Fear
(g)
(h)

Figure 11. Video snapshots for Scenario 2-2.
Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp2-Scen2_2.html

gestures of the robot can be seen in the activation of $u_{\text{AEHA}}(x, t)$, times T6-T8, Figure 20.

5.4 Experiment 4: Influence of the human's emotional state in task time

In Experiment 4 we explore how the human's emotional state might influence the time that it takes to complete the task. We use as a test scenario, the construction of the lower section of the toy vehicle.

Three scenarios were designed, in each scenario the human kept the expression of the same emotional state throughout the duration of the task. In the first scenario the human expressed a negative emotional state (Fear), in the second the human was in a neutral state, and in the third the human displayed a positive emotional state (Happy). In all scenarios, the distribution of the objects in the robot's and human's workspace was the same.

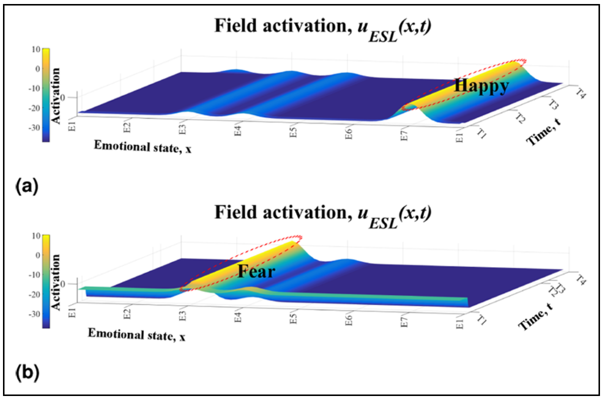


Figure 12 Experiment 2: Emotional state layer. (a) Scenario 2-1: ESL. (b) Scenario 2-2: ESL.

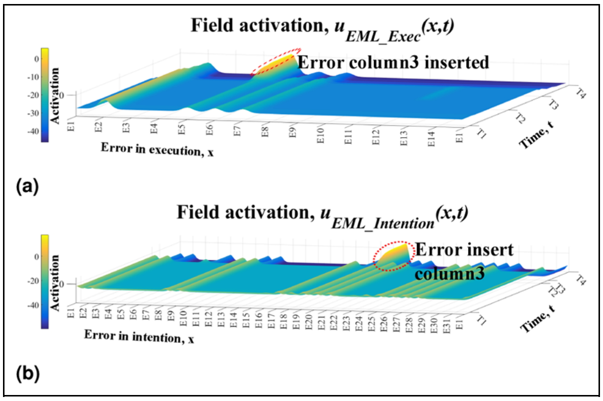


Figure 13. Experiment 2: Error monitoring layer. (a) Scenario 2-1: EML – Error in Execution. (b) Scenario 2-2: EML – Error in Intention.

Table 1. Experiment 4: Time to complete the task as a function of the human emotional state.
Videos online at:
4-1: http://marl.dei.uminho.pt/public/videos/adb/Exp4-Scen4_1.html
4-2: http://marl.dei.uminho.pt/public/videos/adb/Exp4-Scen4_2.html
4-3: http://marl.dei.uminho.pt/public/videos/adb/Exp4-Scen4_3.html

Scenario	Emotional state	Time
4-1	Fear	2 min 55 s
4-2	Neutral	2 min 30 s
4-3	Happy	1 min 50 s

Table 1 shows the results of the three interaction scenarios. When the human is in a fearful state, the robot adjusts the arm movements to be slower and takes more time explaining its actions in order to not startle the human. In a neutral state, the robot uses a medium velocity for the arm movements. When the human displays a happy emotional state, the robot assumes the

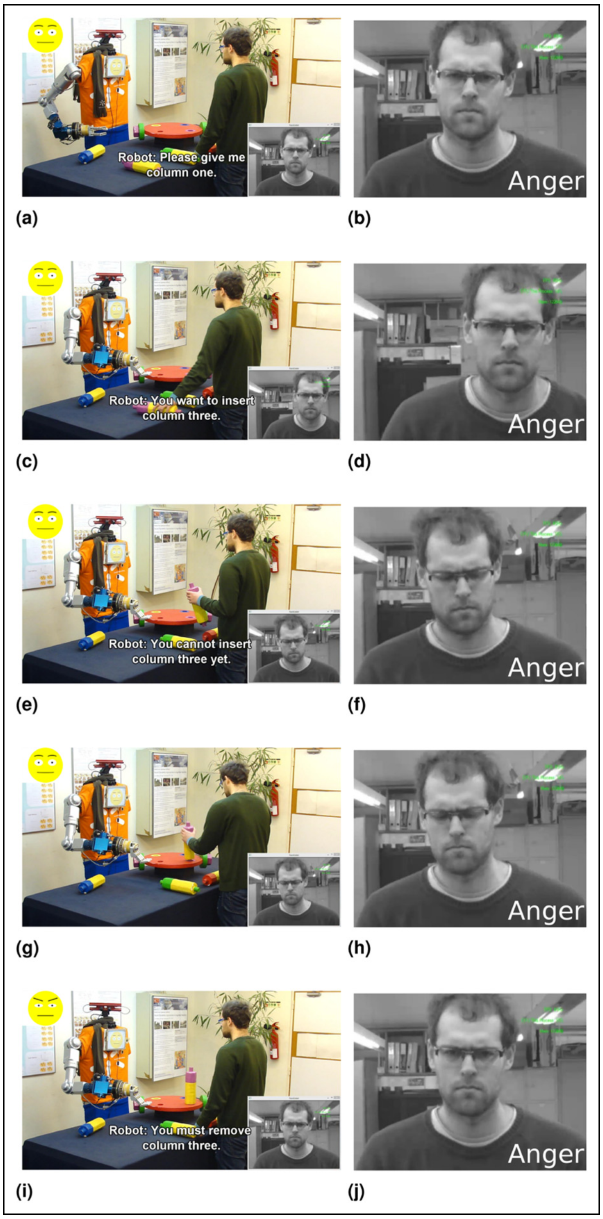

Robot: Please give me column one.
Anger
(a)
(b)
Robot: You want to insert column three.
Anger
(c)
(d)
Robot: You cannot insert column three yet.
Anger
(e)
(f)
Anger
(g)
(h)
Robot: You must remove column three.
Anger
(i)
(j)

Figure 14. Video snapshots for Experiment 3.
Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp3.html

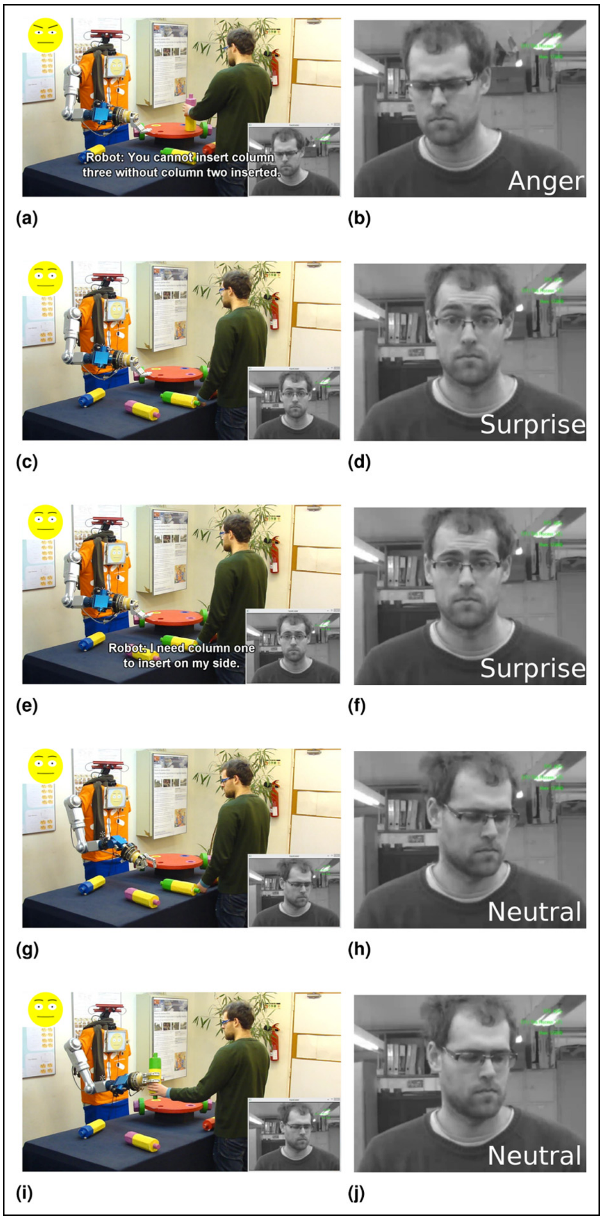

Robot: You cannot insert column three without column two inserted.
Anger
(a)
(b)
Surprise
(c)
(d)
Robot: I need column one to insert on my side.
Surprise
(e)
(f)
Neutral
(g)
(h)
Neutral
(i)
(j)

Figure 15. Video snapshots for Experiment 3 (continued).
Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp3.html

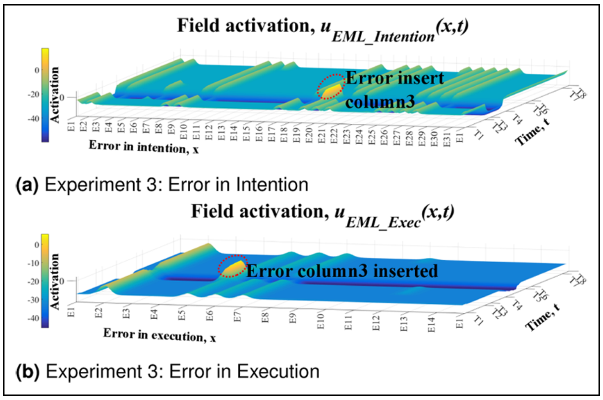
task is running smoothly, increases the velocity for the arm movements, disables the processing of DNFs responsible for the detection of some types of errors, decreasing the time it takes to make decisions.

What the results show in this particular experiment is, the negative expressions impact in the task time by increasing it when compared to a neutral emotional state, 16% in this case. And when in a positive emotional state, the task time is reduced by 27% when compared to the neutral state, but due to disabling the detection of some types of errors, its more prone for errors to occur, since the robot cannot anticipate them.

5.5 Experiment 5: A longer interaction scenario–dynamically adjusting behavior to the expressed human emotional state

As a final interaction scenario, we performed the entire construction task where the human cooperating with the robot shifts the expressed emotional state from negative (Fear) to neutral and then positive (Happy).

The task starts with the human presenting a fearful expression (see Figure 21(b)). The robot adjusts its arm movement velocity to be slower in order to not startle the human, also it takes more time explaining its actions (see Figure 21(a)).



(a) Experiment 3: Error in Intention

(b) Experiment 3: Error in Execution

Figure 16. Experiment 3: Error monitoring layer.

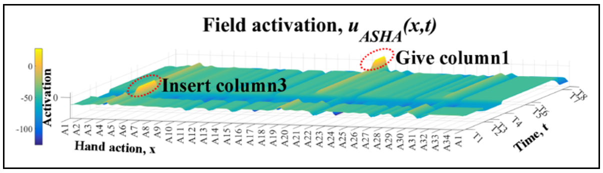


Figure 17. Experiment 3: Action simulation layer–simulation of goal-directed hand actions.

After the wheels are inserted the human presents a neutral expression during the insertion of the nuts (see Figure 21(d)). The robot adjusts the movement velocity to medium and verbalizes less information.

When the middle section is assembled, the human is expressing happiness (see Figure 21(f)), so the robot also smiles and increases the movement velocity for the arm. Here one can see how the robot dynamically and in real time adjusts its behavior – information verbalization and movement velocity – during the execution of the task.

6 Discussion

Decision making refers to the process of selecting a particular action from a set of alternatives. When acting alone, an individual may choose a motor behavior that best serves a certain task based on the integration of sensory evidence and prior task knowledge. In a social context, this process is more complex since the outcome of one's decisions and emotions can be influenced by the decisions and emotions of others. A fundamental building block of social interaction is thus the capacity to predict and understand actions and emotional states of others. This allows an individual to select and prepare an appropriate motor behavior in joint action tasks (Michael, 2011; Sebanz et al., 2006).

Here, we have presented a DNF-architecture that combines the role of emotions in the decision making and movement execution of an autonomous and

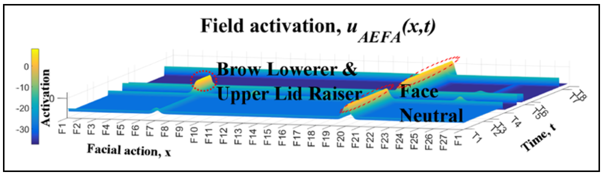


Figure 18. Experiment 3: Action execution layer–facial actions set execution.

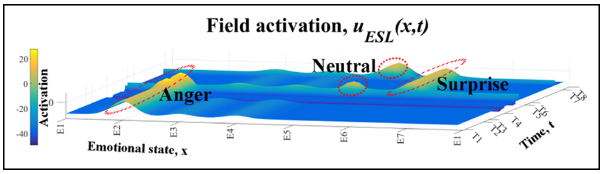


Figure 19. Experiment 3: Emotional state layer.

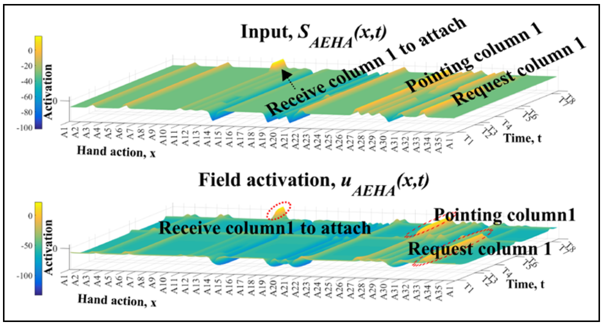


Figure 20. Experiment 3: Action execution layer–goal-directed hand actions.

socially aware robot cooperating with human partners in real-world joint tasks. The proposed architecture is strongly inspired by converging evidence from cognitive and neurophysiological studies suggesting that mirror neurons encoding different levels of abstraction coexist and that there is an automatic but highly context-sensitive mapping from observed on to-be-executed actions as an underlying mechanism (Bekkering et al., 2009; Rizzolatti & Sinigaglia, 2008).

Dynamic neural fields model the emergence of persistent neural activation patterns that allow a cognitive agent to initiate and organize behavior informed by past sensory experience, anticipated future environmental inputs and distal behavioral goals. The DNF-architecture for joint action reflects the notion that cognitive representations, i.e. all items of memory and knowledge, consist of distributed, interactive, and overlapping networks of cortical populations ('cognit' from Fuster (2006)). Network neurons showing suprathreshold activity are participating in the selection of actions, emotional states and their associated consequences. Since the decision-making normally involves multiple, distributed representations of potential

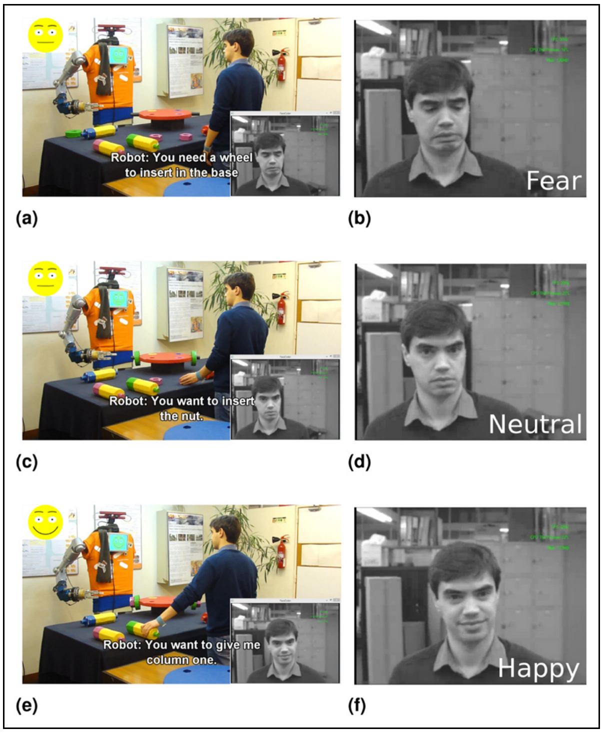

Robot: You need a wheel to insert in the base
(a)
Fear
(b)
Robot: You want to insert the nut.
(c)
Neutral
(d)
Robot: You want to give me column one.
(e)
Happy
(f)

Figure 21. Video snapshots for Experiment 5.
Online at: http://marl.dei.uminho.pt/public/videos/adb/Exp5.html

actions that compete for expression in overt performance, the robot's goal-directed behavior is continuously updated for the current environmental and social context. Important for decision making in a collaborative setting, inferring others' goals and emotional states from their behavior is realized by internal motor simulation based on the activation of the same joint representations of (hand and facial) actions and their environmental effects ('mirror mechanism', Rizzolatti and Sinigaglia (2008); for a recent review see Rizzolatti et al. (2014)). Through this automatic motor resonance process, the observer becomes aligned with the co-actor in terms of actions, emotional states and goals. This alignment allows the robot to dynamically adapt its behavior to that of the human co-actor, without explicit communication (for an integration of verbal communication in the DNF-architecture see Bicho et al. (2010)).

The implementation of aspects of real-time social cognition in a robot based on continuously changing patterns of neuronal activity in a distributed, interactive network strongly contrasts with traditional views of human-like (social) intelligence. These realize the underlying cognitive processes as a manipulation (based on formal logic and formal linguistic systems) of discrete symbols that are qualitatively distinct and entirely separated from sensory and motor information. These approaches have provided many impressive examples of intelligent behavior in artificial agents (for review see Vernon, Metta, & Sandini, 2007), and we do not deny that the sequence of decisions shown in our robotics experiments could be implemented by symbolic planning as well. However, it is now widely recognized by the robotics and cognitive science communities that the symbolic framework has notorious problems coping with real-time interactions in dynamic environments (Haazebroek, Van Dantzig, & Hommel, 2011; Kozma, 2008; Levesque & Lakemeyer, 2008). In human-robot joint tasks, the robot has to reason about a world that may change at any instance of time due to actions taken by the user. Even if we consider that the processing in the perceptual and decision modules would allow continuously updating the robot's plan in accordance with the user's intention and emotional state, the extra processing step needed to embody the abstract action plan in the autonomous robot would challenge the fluent and seemingly effortless coordination of decisions and actions that characterize human joint action in familiar tasks.

Bayesian models represent a popular alternative approach for modeling decision and integration processes in the face of uncertainty (Körding & Wolpert, 2006). It is important to note that the dynamic field framework is compatible with central aspects of probabilistic models. For instance, the pre-activation below threshold of several populations in the action execution layer due to prior task knowledge and contextual information may be interpreted in the sense of a probability density function for different complementary actions. This prior information has to be combined with evidence about the inferred goal and emotional state of the co-actor. In fact, it can be shown that in the input-driven regime the field dynamics may implement Bayes' rules (Cuijpers & Erlhagen, 2008). In our view, there are two major advantages of the dynamic neural field approach. First, stabilizing decision against noise, fluctuations and temporary absence of information in the input stream, is of particular importance. Second, as an example of the dynamical approach to cognition (Schöner, 2008), a DNF-based model allows us to address the important temporal dimension of coordination in joint action (Sebanz et al., 2006). The decision process linked to complementary actions unfolds over time under multiple influences which are themselves modeled as dynamic representations with proper time scales.

We have tested the DNF-architecture in real-time human-robot joint action experiments in the context of a construction task.

In Experiment 1, we have demonstrated how the emotional state of the human partner can affect the decisions made by the robot. Specifically, it was shown that in the same context, a different emotional state displayed by the human can trigger a different complementary behavior on the robot.

In Experiment 2, we have explored how the perceived emotions may play a role in the way the robot detects and handles different types of errors. When the human co-worker is in a positive emotional state, this is taken as a signal that the human is engaged in the task, and thus, it is not probable that he/she will commit errors. The load of the Error Monitoring processes can be decreased by deactivating the anticipation of errors in intention and errors in the action means. The result is that the robot can make decisions faster. In the case that the human co-worker makes an error, this is detected *a posteriori* as an execution error. Conversely, when the human is in a negative emotional state (e.g. Anger) this is used as a signal that the human user is not committed to the task, and thus it is probable that he/she is more prone to making errors. All Error Monitoring processes are activated and this enables the robot to prevent the occurrence of errors by anticipating errors at the goal/intention level.

In Experiment 3 we have demonstrated how the robot can deal with a human operator persisting in making an error. It was shown that by expressing emotional states and verbalization of more information, the robot can induce the (stubborn) human to change his attitude and accept the robot's corrective suggestions.

The above summarized experiments have shown that perceived emotions play an important role in an early stage, during decision making and action preparation of a complementary action (AEL layer). In Experiment 4 it was shown that perceived emotions also play a role later because they may affect the execution at the kinematics level (Motor control). In this experiment, three persons expressing different emotional states (Neutral, Fear, Happy) worked with the robot. When the human co-worker seemed to be in a fearful state, the robot adjusted the arm-hand movements to be slower and took more time verbalizing its reasoning in order to not startle the human. Conversely, when the human displayed a positive emotional state, the robot adjusted the arm-hand movements, and verbalization, to be faster. In a neutral state, the robot used a medium velocity for the arm-hand movements and verbalization. The overall result was that the time to complete the task decreases when the human partner is in a positive emotional state. However, to perform a more in depth study on this matter, a bigger study with more participants is required to make it possible to present statistically relevant results.

Finally, Experiment 5 has shown a longer interaction scenario – the complete construction of the toy vehicle – with the human shifting his emotional state, and the robot adapting in real time its behavior to these changes.

As we have shown, the adopted dynamic perspective offers in general a high degree of flexibility in joint task execution. However, in the present implementation of the DNF-architecture the neural representations and their connectivity were tailored by the designer. It is highly desirable to endow the robot with a developmental program that would allow it to autonomously learn and represent new representations (Asada et al., 2009; Weng, 2004). Using correlation-based learning rules (Gerstner & Kistler, 2002) with a gating that signals the success of behavior, we have shown for instance how goal-directed mappings between action observation and action execution that support an action understanding capacity may develop during learning and practice (Erlhagen, Mukovskiy, & Bicho, 2006a; Erlhagen, Mukovskiy, Bicho, Panin, et al., 2006). Importantly, the developmental process, through Hebbian learning rules, may explain the emergence of new task-specific populations that have not been introduced to the architecture by the human designer. Recently, we have demonstrated how the robot may autonomously develop – through tutor demonstration and feedback during joint performance – the connections between the populations in the two layers of the CSGL that code the possible serial orders and the longer term dependencies between subgoals.

The work on learning and development in the DNF-architecture for joint action is consistent with the work of Keysers and Gazzola (2014) who have analysed how mirror neurons could develop and become a dynamic system that performs active inferences about the actions, sensations and emotions of others and allows joint actions despite sensory motor delays.

Various works have explored automatic facial expression recognition in human-computer interaction (see Pantic & Bartlett, 2007; Tian, Kanade, & Cohn, 2005). However, a human-robot scenario presents additional challenges: lack of control over lighting conditions, relative poses, the inherent mobility of the robot and separation between robot and human. These are limitations imposed on our robot that are also present in other works (e.g Wimmer, MacDonald, Jayamuni, & Yadav, 2008). The vision system limitations prevented us from performing experiments with a larger numbers of human subjects. The vision system relies on the acquisition of a neutral face of the subject to perform the Action Units coding, which might not be possible at all times. Also, the features extraction is not robust enough to detect subtle and micro expressions, which in more naturalistic scenarios would be the most common expressions. Tests conducted to the system by using the Cohn–Kanade face database (Kanade, Cohn, & Tian, 2000) reveal detection rates for some Action Units above 70% (4, 12 15), others have detection rates just above 50% (1, 2, 5, 26). This led us to instruct the participants in our studies to perform posed expressions to improve the system detection rate.

Regardless of the sensorial limitations, the DNF-architecture proved to be ready to cope with the demands of truly real world human-robot joint action scenarios. When dealing with multiple information sources, which in the real world might not be reliable or consistent, our DNF based cognitive architecture is able to cope with these situations, even when the information is not available all at the same time. Being able to synthesize, in an embodied artificial agent, the cognitive demands of real-time interactions with a human co-actor whose displayed emotional states modulate the robot's behavior shows that the dynamic neural field theory provides a promising research program for bridging the gap that still exists in natural and (socially) intelligent human-robot joint action.

In the future, further user studies need to be conducted to assess how the robot can be more expressive, and also how we can explore the subject of face recognition to allow the robot to customize the interaction based on the person that is interacting with it.

Supplemental material

In the supplemental material one can find the meaning and connection scheme for the neural pools in the layered DNF architecture, numerical values for the dynamic field parameters, and the numerical values for the inter-field synaptic weights.

Acknowledgements

This work benefited from the invaluable contribution of people in our Lab, namely, Angela Loureiro, António Ramiro, Emanuel Sousa and Paulo Vicente.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was possible in part by the funding of research grants from the Portuguese Foundation for Science and Technology (grant numbers SFRH/BD/48527/2008, SFRH/BPD/71874/2010, SFRH/BD/81334/2011), and with funding from FP6-IST2 EU-IP Project JAST (project number 003747) and FP7 Marie Curie ITN Neural Engineering Transformative Technologies NETT (project number 289146).

References

Adolphs, R. (2006). How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. *Brain research*, *1079*(1), 25–35. doi: 10.1016/j.brainres.2005.12.127

Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, *27*(2), 77–87. doi: 10.1007/BF00337259

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., & . . . Yoshida, C. (2009). Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development*, *1*(1), 12–34.

Bekkering, H., de Bruijn, E. R. A., Cuijpers, R. H., Newman-Norlund, R. D., van Schie, H. T., & Meulenbroek, R. G. J. (2009). Joint action: Neurocognitive mechanisms supporting human interaction. *Topics in Cognitive Science*, *1*, 340–352. doi: 10.1111/j.1756-8765.2009.01023.x

Bicho, E., Erlhagen, W., Louro, L., & Costa e Silva, E. (2011). Neuro-cognitive mechanisms of decision making in joint action: A human-robot interaction study. *Human Movement Science*, *30*, 846–868. doi: 10.1016/j.humov.2010.08.012

Bicho, E., Erlhagen, W., Louro, L., Costa, e, Silva, E., Silva, R., & Hipólito, N. (2011). A dynamic field approach to goal inference, error detection and anticipatory action selection in human-robot collaboration. In K. Dautenhahn, & J. Saunders (Eds.), *New frontiers in human-robot interaction (advances in interaction studies)* (6th ed., pp. 135–164). Amsterdam: John Benjamins Publishing Company.

Bicho, E., Louro, L., & Erlhagen, W. (2010). Integrating verbal and nonverbal communication in a dynamic neural field architecture for human-robot interaction. *Frontiers in Neurorobotics*, *4*, 1–13. doi: 10.3389/fnbot.2010.00005

Bicho, E., Mallet, P., & Schöner, G. (2000). Target representation on an autonomous vehicle with lowlevel sensors. *The International Journal of Robotics Research*, *19*, 424–447. doi: 10.1177/02783640022066950

Blakemore, S., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*, 561–567.

Breazeal, C. (2003a). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*(1–2), 119–155. doi: 10.1016/S1071-5819(03)00018-1

Breazeal, C. (2003b). Toward sociable robots. *Robotics and Autonomous Systems*, *42*(3–4), 167–175.

Cañamero, L. (2005). Emotion understanding from the perspective of autonomous robots research. *Neural networks : the official journal of the International Neural Network Society*, *18*, 445–55. doi: 10.1016/j.neunet.2005.03.003

Cañamero, L., & Fredslund, J. (2000). How Does It Feel? Emotional Interaction with a Humanoid LEGO Robot. In K. Dautenhahn (Ed.), *Socially intelligent agents: The human in the loop. papers from the aaai 2000 fall symposium* (pp. 23–28). Cape Cod, MA: AAAI Press.

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(9), 5497–502. doi: 10.1073/pnas.0935845100

Costa e Silva, E., Costa, F., Bicho, E., & Erlhagen, W. (2011). Nonlinear optimization for humanlike movements of a high degree of freedom robotics arm-hand system. In *Computational science and its applications-iccsa 2011* (pp. 327–342). Berlin Heidelberg: Springer.

Cuijpers, R. H., & Erlhagen, W. (2008). Implementing bayesain rule with neural fields. In *Artificial neural networks-icann 2008* (pp. 228–237). Berlin Heidelberg: Springer.

Ekman, P., & Friesen, W. V. (1978). Facial action coding system: A technique for the measurement of facial movement. In P. C. Ellsworth, & C. A. Smith (Eds.), *From appraisal to emotion: Differences among unpleasant feelings. motivation and emotion* (*Vol. 12*, pp. 271–302). Palo Alto, CA: Consulting Psychologists Press (1988).

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system.* Salt Lake City, USA: Research Nexus division of Network Information Research Corporation.

Enticott, P. G., Johnston, P. J., Herring, S. E., Hoy, K. E., & Fitzgerald, P. B. (2008). Mirror neuron activation is associated with facial emotion processing. *Neuropsychologia*, *46*, 2851–2854.

Erlhagen, W., & Bicho, E. (2006). The dynamic neural field approach to cognitive robotics. *Journal of Neural Engineering*, *3*, R36–R54.

Erlhagen, W., & Bicho, E. (2014). A Dynamic Neural Field Approach to Natural and Efficient Human-Robot Collaboration. In *Neural fields* (pp. 341–365). Berlin Heidelberg: Springer.

Erlhagen, W., Mukovskiy, A., & Bicho, E. (2006a). A dynamic model for action understanding and goaldirected imitation. *Brain Research*, *1083*, 174–188.

Erlhagen, W., Mukovskiy, A., Bicho, E., Panin, G., Kiss, C., Knoll, A., & . . . Bekkering, H. (2006). Goaldirected imitation for robots: A bio-inspired approach to action understanding and skill learning. *Robotics and autonomous systems*, *54*, 353–360.

Erlhagen, W., Mukovskiy, A., Chersi, F., & Bicho, E. (2007b, jul). On the development of intention understanding for joint action tasks. In *2007 ieee 6th international conference on development and learning* (pp. 140–145). London: Imperial College London. doi: 10.1109/DEVLRN.2007.4354022

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience*, *17*, 1703–1714. doi: 10.1046/j.1460-9568.2003.02601.x

Ferri, F., Campione, G. C., Dalla Volta, R., Gianelli, C., & Gentilucci, M. (2010). To me or to you? When the self is advantaged. *Experimental brain research*, *203*, 637–646.

Ferri, F., Stoianov, I. P., Gianelli, C., D'Amico, L., Borghi, A. M., & Gallese, V. (2010). When action meets emotions: how facial displays of emotion influence goal-related behavior. *PloS One*, *5*, e13126.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science (New York, N.Y.)*, *308*(5722), 662–667. doi: 10.1126/science.1106138

Fogassi, L., & Rizzolatti, G. (2013). The Mirror Mechanism as Neurophysiological Basis for Action and Intention Understanding. In A. Suarez, & P. Adams (Eds.), *Is science compatible with free will?* (pp. 117–134). New York, NY: Springer New York. doi: 10.1007/978-1-4614-5212-6_9

Fong, T., Nourbakhsh, I. R., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, *42*(3–4), 143–166. doi: 10.1016/S0921-8890(02)00372-X

Fuster, J. M. (2006). The cognit: a network model of cortical representation. *International Journal of Psychophysiology*, *60*, 125–132.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609. doi: 10.1093/brain/119.2.593

Genovesio, A., Brasted, P. J., & Wise, S. P. (2006). Representation of future and previous spatial goals by separate neural populations in prefrontal cortex. *The Journal of Neuroscience*, *26*, 7305–7316.

Gerstner, W., & Kistler, W. M. (2002). Mathematical formulations of hebbian learning. *Biological Cybernetics*, *87*(5–6), 404–415.

Grecucci, A., Cooper, R. P., & Rumiati, R. I. (2007). A computational model of action resonance and its modulation by emotional stimulation. *Cognitive Systems Research*, *8*(3), 143–160.

Haazebroek, P., Van Dantzig, S., & Hommel, B. (2011). A computational model of perception and action for cognitive robotics. *Cognitive Processing*, *12*, 355–365.

Hegel, F., Spexard, T., Wrede, B., Horstmann, G., & Vogt, T. (2006). Playing a different imitation game: Interaction with an Empathic Android Robot. In G. Sandini, & A. Billard (Eds.), *2006 6th ieee-ras international conference on humanoid robots* (pp. 56–61). Genova, Italy: IEEE. doi: 10.1109/ICHR.2006.321363

Hu, & Ming-Kuei. (1962). Visual pattern recognition by moment invariants. *IEEE Transactions on Information Theory*, *8*, 179–187. doi: 10.1109/TIT.1962.1057692

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*, e79. doi: 10.1371/journal.pbio.0030079

Kanade, T., Cohn, J. F., & Tian, Y. (2000). Comprehensive database for facial expression analysis. In F. M. Titsworth (Ed.), *Proceedings of the fourth IEEE international conference on automatic face and gesture recognition (fg'00)* (pp. 46–53). Grenoble, France: IEEE Comput. Soc. doi: 10.1109/AFGR.2000.840611

Keysers, C., & Gazzola, V. (2006). Towards a unifying neural theory of social cognition. *Progress in Brain Research*, *156*, 379–401. doi: 10.1016/S0079-6123(06)56021-2

Keysers, C., & Gazzola, V. (2014). Hebbian learning and predictive mirror neurons for actions, sensations and emotions. *Philosophical Transactions of the Royal Society B*, *369*(1644), 20130175, 1–11.

Keysers, C., Wicker, B., Gazzola, V., Anton, J.-L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron*, *42*, 335–346. doi: 10.1016/S0896-6273(04)00156-4

Kirby, R., Forlizzi, J., & Simmons, R. (2010). Affective social robots. *Robotics and Autonomous Systems*, *58*, 322–332. doi: 10.1016/j.robot.2009.09.015

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10*, 319–326.

Kozma, R. (2008). Intentional systems: Review of neurodynamics, modeling, and robotics implementation. *Physics of Life Reviews*, *5*, 1–21.

Kedzierski, J., Muszynski, R., Zoll, C., Oleksy, A., & Frontkiewicz, M. (2013). EMYS - Emotive Head of a Social Robot. *International Journal of Social Robotics*, *5*, 237–249. doi: 10.1007/s12369-013-0183-1

Leslie, K. R., Johnson-Frey, S. H., & Grafton, S. T. (2004). Functional imaging of face and hand imitation: Towards a motor theory of empathy. *NeuroImage*, *21*, 601–607. doi: 10.1016/j.neuroimage.2003.09.038

Levesque, H., & Lakemeyer, G. (2008). Cognitive robotics. *Foundations of Artificial Intelligence*, *3*, 869–886.

Locke, K. D., & Horowitz, L. M. (1990). Satisfaction in interpersonal interactions as a function of similarity in level of dysphoria. *Journal of Personality and Social Psychology*, *58*, 823–31.

Lowe, R., Herrera, C., Morse, A., & Ziemke, T. (2007). The Embodied Dynamics of Emotion, Appraisal and Attention. In L. Paletta, & E. Rome (Eds.), *Attention in cognitive systems. theories and systems from an interdisciplinary viewpoint* (*Vol. 4840*, pp. 1–20). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-540-77343-6_1

Michael, J. (2011). Shared emotions and joint action. *Review of Philosophy and Psychology*, *2*, 355–373. doi: 10.1007/s13164-011-0055-2

Newman-Norlund, R. D., van Schie, H. T., van Zuijlen, A. M., & Bekkering, H. (2007). The mirror neuron system is more active during complementary compared with imitative action. *Nature Neuroscience*, *10*, 817–818.

Novikova, J., & Watts, L. (2015). Towards artificial emotions to assist social coordination in HRI. *International Journal of Social Robotics*, *7*, 77–88. doi: 10.1007/s12369-014-0254-y

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. *Cognition and Emotion*, *1*, 29–50.

Oatley, K., & Johnson-Laird, P. (2014). Cognitive approaches to emotions. *Trends in Cognitive Sciences*, *18*, 134–140.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, *9*, 242–9. doi: 10.1016/j.tics.2005.03.010

Pantic, M., & Bartlett, M. S. (2007). Machine Analysis of Facial Expressions. In K. Kurihara (Ed.), *Face recognition* (pp. 237–366). Vienna, Austria: I-Tech Education and Publishing.

Poljac, E., van Schie, H. T., & Bekkering, H. (2009). Understanding the flexibility of action-perception coupling. *Psychological Research*, *73*, 578–86. doi: 10.1007/s00426-009-0238-y

Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., & Rozzi, S. (2014). Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. *Physiological Reviews*, *94*, 655–706. doi: 10.1152/physrev.00009 .2013

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–92. doi: 10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*, 661–70. doi: 10.1038/35090060

Rizzolatti, G., & Sinigaglia, C. (2008). Mirrors in the Brain: *How Our Minds Share Actions and Emotions* (Translated ed.). New York: Oxford University Press.

Schaal, S. (1999). Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences*, *3*, 233–242. doi: 10.1016/S1364-6613(99)01327-3

Scheutz, M. (2011). The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robotics* (pp. 205–221). Cambridge, Massachusetts: MIT Press.

Scheutz, M., Schermerhorn, P., & Kramer, J. (2006). The utility of affect expression in natural language interactions in joint human-robot tasks. In M. A. Goodrich, A. C. Schultz, & D. J. Bruemmer (Eds.), *Proceeding of the 1st acm sigchi/sigart conference on humanrobot interaction - hri '06* (p. 226). New York, USA: ACM Press. doi: 10.1145/1121241. 1121281

Schöner, G. (2008). Dynamical systems approaches to cognition. In *Cambridge handbook of computational cognitive modeling* (pp. 101–126). Cambridge, UK: Cambridge University Press.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*, 70–76.

Silva, R., Bicho, E., & Erlhagen, W. (2008). ARoS: An anthropomorphic robot for human-robot interaction and coordination studies. In *Proceedings of the controlo'2008 conference - 8th portuguese conference on automatic control* (pp. 819–826). UTAD – Vila Real, Portugal: UTAD.

Sousa, E., Erlhagen, W., Ferreira, F., & Bicho, E. (2015). Off-line simulation inspires insight: A neurodynamics approach to efficient robot task learning. *Neural Networks*, *72*, 123–139. doi: 10.1016/j.neunet.2015.09.002

Talanov, M., Vallverdu, J., Distefano, S., Mazzara, M., & Delhibabu, R. (2015). Neuromodulating Cognitive Architecture: Towards Biomimetic Emotional AI. In L. Barolli, M. Takizawa, F. Xhafa, T. Enokido, & J. H. Park (Eds.), *2015 ieee 29th international conference on advanced information networking and applications* (pp. 587–592). Los Alamitos, California: IEEE. doi: 10.1109/AINA.2015.240

Tian, Y.-L., Kanade, T., & Cohn, J. F. (2005). Facial Expression Analysis. In *Handbook of face recognition* (pp. 247–275). New York: Springer-Verlag. doi: 10.1007/0-387-27257-7_12

van der Gaag, C., Minderaa, R. B., & Keysers, C. (2007). Facial expressions: What the mirror neuron system can and cannot tell us. *Social neuroscience*, *2*(3–4), 179–222. doi: 10.1080/17470910701376878

van Schie, H. T., van Waterschoot, B. M., & Bekkering, H. (2008). Understanding action beyond imitation: reversed compatibility effects of action observation in imitation and joint action. *Journal of Experimental Psychology. Human Perception and Performance*, *34*, 1493–500. doi: 10.1037/a0011750

Vernon, D., Metta, G., & Sandini, G. (2007). A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. *IEEE Transactions on Evolutionary Computation*, *11*, 151.

Weng, J. (2004). Developmental robotics: Theory and experiments. *International Journal of Humanoid Robotics*, *1*, 199–236.

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*, 655–664. doi: 10.1016/S0896-6273(03)00679-2

Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, *131*, 460.

Wimmer, M., MacDonald, B. A., Jayamuni, D., & Yadav, A. (2008). Facial expression recognition for human-robot interaction - A prototype. In G. Sommer, & R. Klette (Eds.), *Robot vision* (*Vol. 4931*, pp. 139–152). Berlin Heidelberg: Springer. doi: 10.1007/978-3-540-78157-8_11

Appendix

Table 1. Description of Action Units (AUs) and appearance changes caused by each AU in the face, according to Ekman et al. (2002).

AU	Name	Description
AU1	Inner Brow Raiser	Inner portion of the eyebrows up.
AU2	Outer Brow Raiser	Outer portion of the eyebrows up.
AU4	Brow Lowerer	Lowers the entire eyebrow.
AU5	Upper Lid Raiser	Eyes wide open.
AU7	Lid Tightener	Eyes half closed.
AU9	Nose Wrinkler	Causes wrinkles to appear in nose.
AU12	Lip Corner Puller	Lip corners up (Smile).
AU15	Lip Corner Depressor	Lip corners down.
AU20	Lip Stretcher	Pulls the lips back laterally.
AU25	Lips Part	Open mouth slightly.
AU26	Jaw Drop	Open the mouth.

Table 2. Combinations of AUs and human movements capable of activating an emotional state in ESL (the detection of an AU implies that the face is detected).

Inferred emotion	AUs and human movements
Disgust	4 + 12 + 25 / 4 + 25 / 9 + 15 / 9 / 4 + 9 / 4 + 5 + 9
Anger	4 + 5 / 4 + 7 / 4 + Head Movement High
Fear	1 + 2 + 5 + 20 + 26 / 4 + 20 / 1 + 20 5 + Hand Movement Low
Sadness	1 + 4 + 15 / 1 + 4 / 1 + 15 / 4 + 15 1 + Hand Movement Low 15 + Hand Movement Low
Neutral	Face detected
Surprise	1 + 2 / 1 + 2 + 5 / 1 + 2 + 26 / 1 + 2 + 5 + 26
Happiness	12 / 1 + 2 + 12 / 12 + 26

Table 3. Experiment 1: Scenario 1-1.

Label	Time (s)
T1	5
T2	43
T3	46

Table 4. Experiment 1: Scenario 1-2.

Label	Time (s)
T1	5
T2	14
T3	21
T4	30

Table 5. Experiment 2: Scenario 2-1.

Label	Time (s)
T1	2
T2	8
T3	15
T4	23

Table 6. Experiment 2: Scenario 2-2.

Label	Time (s)
T1	3
T2	17
T3	20
T4	29

Table 7. Experiment 3.

Label	Time (s)
T1	3
T2	13
T3	17
T4	23
T5	38
T6	42
T7	58
T8	64

About the Authors

Rui Silva received his MSc in Industrial Electronics and Computers Engineering, with specialization on "Automation, Control and Robotics" at the University of Minho, Portugal, in 2008. He is finishing his doctoral studies on "Electronics and Computers Engineering". His research interests are focused on Computer Vision, Non-linear Dynamical Systems, Simulation, Robotics and Facial Expression Recognition. Currently he is working at Displax - Multitouch Technologies.

Luís Louro received a PhD in the area of Automation and Robotics (PhD program on Electronics and Computers Engineering) at University of Minho, in 2010. He was a research assistant at the European projects "Artesimit–Artefact Structural Learning through Imitation" and "JAST - Joint Action Science and Technology". His research interests are Autonomous and Anthropomorphic Robotics, Human-Robot Interaction. Currently he has a post-doc position at the University of Minho, Portugal, and he is an assistant professor at Lusíada University, Portugal.

Tiago Malheiro received his MSc in Industrial Electronics and Computers Engineering, with specialization on "Automation, Control and Robotics" and "Embedded Systems" from University of Minho, Portugal, in 2011. He is currently working toward a PhD in robotics focusing on the development of pro-active robots to assist dependent persons. His research interests are focused on Robotics, Non-linear Dynamical Systems, and Embedded Systems.

Wolfram Erlhagen is Associate Professor at the Department of Mathematics at the University of Minho, Portugal. He has been PI in several European and national projects in the ICT topic. His multidisciplinary research covers the multi-scale analysis of neuronal activity, the functional modeling of brain circuits, and the implementation of neuro-based models in autonomous robots. In close cooperation with experimental groups he applies his theoretical investigations to problems of motor planning, visual perception and reasoning with the ultimate goal to bridge Cognitive Sciences to Robotics.

Estela Bicho is Associate Professor at the Department of Industrial Electronics at University of Minho, Portugal, where she is responsible for courses in Non-linear Dynamical Systems, Control and Robotics and heads the research lab on Autonomous (mobile and anthropomorphic) Robotics & Dynamical systems. She obtained the PhD degree in Robotics, Automation and Control, in 1999, from the University of Minho. Her PhD work received the honor price from Portuguese-IBM (1999). Her research concentrates on the use of dynamical systems for the design and implementation of neuro-cognitive control architectures for flexible control of high-DOF robotics systems, including human-robot interaction and joint action. She has been PI in several national and EU funded research projects in robotics.