TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 14 (2022) 327–343

ISSN: 1756-8765 online
DOI: 10.1111/tops.12587

This article is part of the topic "Everyday Activities," Holger Schultheis and Richard P. Cooper (Topic Editors).

A Robotic Cognitive Control Framework for Collaborative Task Execution and Learning

Riccardo Caccavale, Alberto Finzi

Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione (DIETI), Università degli Studi di Napoli "Federico II"

Received 7 September 2021; received in revised form 20 October 2021; accepted 21 October 2021

Abstract

In social and service robotics, complex collaborative tasks are expected to be executed while interacting with humans in a natural and fluent manner. In this scenario, the robotic system is typically provided with structured tasks to be accomplished, but must also continuously adapt to human activities, commands, and interventions. We propose to tackle these issues by exploiting the concept of cognitive control, introduced in cognitive psychology and neuroscience to describe the executive mechanisms needed to support adaptive responses and complex goal-directed behaviors. Specifically, we rely on a supervisory attentional system to orchestrate the execution of hierarchically organized robotic behaviors. This paradigm seems particularly effective not only for flexible plan execution but also for human–robot interaction, because it directly provides attention mechanisms considered as pivotal for implicit, non-verbal human–human communication. Following this approach, we are currently developing a robotic cognitive control framework enabling collaborative task execution and incremental task learning. In this paper, we provide a uniform overview of the framework illustrating its main features and discussing the potential of the supervisory attentional system paradigm in different scenarios where humans and robots have to collaborate for learning and executing everyday activities.

Keywords: Cognitive robotics; Cognitive control; Cognitive architecture; Attention; Human–robot collaboration

Correspondence should be sent to Alberto Finzi, Dipartimento di Ingegneria Elettrica e Tecnologie dell'Informazione (DIETI), Università degli Studi di Napoli "Federico II," Via Claudio 21, Napoli, Italy, 80125. Email: alberto.finzi@unina.it

1. Introduction

Social and service robots are expected to effectively assist humans during their daily activities providing a natural and fluent interaction. In these settings, robotic systems are often required to collaborate with humans during the execution of multiple tasks, while continuously adapting their behavior to the human activities and intentions. From the human side, it is also desirable that the robot behavior is perceived as safe, compliant, and predictable. In order to support this natural and effective interaction, human monitoring and communication processes are to be situated in the operative context and strictly integrated with the robot control processes (activity planning, behavior orchestration, action execution, etc.).

In the robotics literature, several frameworks have been proposed to conciliate natural human–robot interaction and the coordinated execution of goal-oriented activities. The dominant approach relies on architectures for plan-based autonomy, which leverage several planning systems and replanning processes to continuously align the robot planned actions with respect to the interpreted human activities and commands. This paradigm has been successfully deployed in the field and industrial robotics to enable mixed initiative interaction and adjustable autonomy (Carbone, Finzi, Orlandini, & Pirri, 2008; Karpas, Levine, Yu, & Williams, 2015); on the other hand, when humans and robots interact in close proximity during daily tasks this continuous planning/replanning process is usually computationally expensive and not reactive and smooth enough to provide a natural and fluent interaction. Alternative approaches are provided by the behavior-based robotics paradigm (Breazeal, Edsinger, Fitzpatrick, & Scassellati, 2001; Scheutz, Harris, & Schermerhorn, 2013) and cognitive robotics (Baxter, de Greeff, & Belpaeme, 2013; Trafton et al., 2013). In particular, cognitive models and architectures have been exploited to capture several processes involved in human–robot interaction (motivations, emotions, drives, attention, communication, social interaction, etc.). However, the integration of human–robot interaction/communication processes with the processes of generation and orchestration of collaborative task-oriented activities remains a challenging issue. In cognitive psychology and neuroscience, the executive mechanisms needed to support flexible, adaptive responses and complex goal-directed cognitive processes and behaviors are associated with the concept of cognitive control (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Norman & Shallice, 1986; Posner & Snyder, 1975; Rubinstein, Meyer, & Evan, 2001). Notwithstanding their relevance in cognitive science, cognitive control models have been rarely exploited in robotic systems (Caccavale & Finzi, 2017; Caccavale, Saveriano, Finzi, & Lee, 2019; Garforth, McHale, & Meehan, 2006; Kasderidis & Taylor, 2004; Kawamura, Gordon, Ratanaswasd, Erdemir, & Hall, 2008). In particular, executive attention processes, which are considered as key mechanisms for flexible action execution and coordination in humans (Cooper & Shallice, 2000, 2006; Norman & Shallice, 1986), have been largely neglected in the robot planning and execution literature. In contrast, we believe that similar attention-based control processes are expected to play a crucial role in robotic systems as well, since the associated activation/regulation mechanisms can support human monitoring, flexible orchestration of multiple tasks, human–robot activity coordination, learning by demonstration, etc. Specifically, we propose to deploy a supervisory attentional system (SAS) executive model (Cooper & Shallice, 2000, 2006; Norman & Shallice,

1986) to monitor and orchestrate the execution of hierarchically organized robot (and human) behaviors. This paradigm seems particularly effective not only for flexible execution of multiple competing tasks but also for human–robot interaction, since it naturally provides attention mechanisms (attention manipulation, joint attention, etc.) considered as pivotal for implicit, non-verbal human–human communication (Tomasello, 2010). In a SAS-based cognitive control framework, activities can be performed exploiting action schemata, which specify well-learned patterns of behaviors or cognitive processes enabling task/subtask accomplishment. Schemata are hierarchically organized, each endowed with an activation value. Each schema can be activated, aroused, or inhibited by perceptual stimuli or other active schemata. Multiple conflicting schemata can be active at the same time; therefore, orchestration mechanisms are needed. In this respect, SAS provides two main processes: contention scheduling and supervisory attention system. Contention scheduling is a low-level mechanism that exploits schemata activation values to solve conflicts among competing schemata in a reactive fashion. Instead, the supervisory attentional system is a higher level mechanism that affects contention scheduling in the case of non-routine situations to enable flexible/adaptive behavior orchestration and goal-oriented responses.

In this paper, we discuss the effectiveness of similar action orchestration mechanisms for robot control and human–robot collaboration. In particular, we claim that the SAS paradigm provides key attention-based regulation mechanisms, which not only are crucial for natural human–robot communication and collaborative task execution but are also practical and effective for the accomplishment of real-world everyday robotic tasks. We discuss these issues presenting a uniform overview of a SAS-based cognitive control framework we are currently developing (Caccavale & Finzi, 2017, 2019; Caccavale et al., 2019; Caccavale et al., 2017), illustrating its main features along with application scenarios where humans and robots have to collaborate for learning and executing incrementally complex everyday activities. In these case studies, we highlight the relevance of attention-based regulation mechanisms for flexible and collaborative execution of multiple competing tasks/subtasks, implicit human–robot communication, and task learning by demonstration.

2. An attention-based robotic executive framework

In order to enable natural and fluent human–robot interaction during collaborative execution of everyday activities, the robotic system needs to be endowed with executive functions to orchestrate and modulate multiple processes, both reactive and goal directed, while rapidly and smoothly adapting task execution to environmental changes and human interventions. For instance, a robotic assistant that collaborates with a human coworker for beverage preparation in a kitchen workspace is expected to be capable of monitoring and executing different concurrent, competing, and interleaved tasks (e.g., prepare tea, coffee, cocktails) composed of various subtasks (e.g., take bottles, glasses, pour liquids), in continuous interaction with human coworkers, which may either directly execute subtasks (e.g., open bottles, pick/place containers) or delegate/change tasks and subtasks to the robotic coworker (e.g., prepare a coffee or take a teapot) through rapid verbal/non-verbal signals (e.g., pointing, verbal cues).

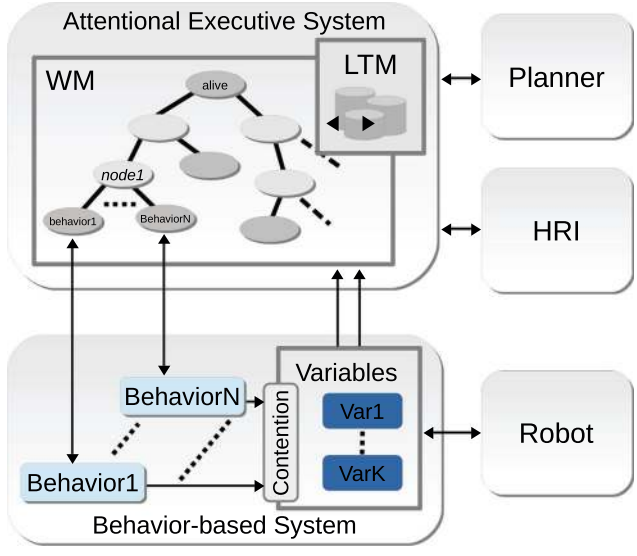

Fig. 1. System architecture. The *attentional executive system* (AS) orchestrates the sensorimotor processes collected in the behavior-based system (BS). AS continuously allocates/deallocates hierarchical tasks in WM retrieved from the LTM, while behaviors in BS compete for shared resources (state variables and robot devices). AS also interacts with external modules like task/motion planners (Planner) and human–robot interaction (HRI) systems.

Here, not only the order of operations to accomplish the tasks can dynamically change, but also the division of work between the human and the robot is not predetermined. Moreover, the interpretation of human activities and intentions is task and context dependent. In these scenarios, executive attention mechanisms play a primary role in activity monitoring, task regulation and switching, attention manipulation, and implicit communication. To endow a robotic system with such mechanisms, we propose to deploy a SAS-inspired cognitive control paradigm, where attention to action (Norman & Shallice, 1986) is exploited to regulate the execution of robotic activities at different levels of abstraction. The underlying assumption is that supervisory attention and contention scheduling can be exploited at the core of a robotic executive control process to reduce continuous task replanning, support flexible execution of multiple concurrent tasks with smooth and human-legible task switching, and enable attention-based human–robot interaction/communication. The design principle is to provide a practical attention-based executive framework, suitable for real-world collaborative robotic systems, which is also compatible with AI methods for planning, execution, learning, and human–robot interaction/communication.

2.1. System architecture

An abstract representation of the proposed system is illustrated in Fig. 1. The architecture is based on a working memory (Baddeley, 1993, 1996), with executive control processes

regulated by a SAS-inspired framework. The executive system exploits a long-term memory (LTM), a procedural memory where all the available/learned tasks are stored (e.g., procedures to prepare beverages), a working memory (WM) containing context-relevant tasks recruited for execution or monitoring, and a behavior-based system (BS) collecting sets of routinized behaviors, representing instantiated sensorimotor processes competing for the execution (e.g., picking a glass). Behaviors can be designed or trained to control the robot actuators, acquire data from the sensors, and update the WM. Competitions among behaviors for shared resources are regulated by attentional processes.

2.2. Task representation and long-term memory

We assume that the repository of procedural knowledge available to the system is stored in a LTM, which collects the description of the activities the robotic system can monitor and execute. Following a typical approach in cognitive science (Cooper & Shallice, 2000, 2006; Kleijn, Kachergis, & Hommel, 2014; Lashley, 1951; Norman & Shallice, 1986) and AI (Nau et al., 2003; Nicolescu & Mataric, 2003), we assume that activities are organized as hierarchical (and goal-directed) tasks to be accomplished (e.g., a coffee preparation can be decomposed into different steps). In our computational framework, each activity is an action schema symbolically represented by a predicate in the form $t(x_1, \ldots, x_n)$, where t is the name of the task and $x_1, \ldots, x_n$ are parameters to be online instantiated. Tasks can be either *concrete* or *abstract*, where the concrete ones represent primitive sensorimotor processes, while the abstract ones represent complex activities to be further hierarchically decomposed into simpler subtasks. As in hierarchical task network representations (HTNs) (Nau et al., 2003), multiple decompositions can be available for an abstract task, representing alternative methods for task accomplishment. Analogously to Cooper and Shallice (2006), tasks are also associated with *preconditions* and *postconditions*. In our framework, *preconditions* are propositional formulas to be satisfied to enable the execution of the task, while *postconditions* are to be satisfied when a task is completed. For example, during a beverage preparation, the task $take(glass)$ can be associated with the precondition $on(glass, table)$ and the $holding(glass)$ postcondition. In the case of *concrete* tasks, preconditions and postconditions are represented by a Stanford Research Institute Problem Solver (STRIPS)-like representation (Fikes & Nilsson, 1971), this way our task representation can also be exploited as a planning domain for both classical and HTN planners, which can be invoked by the executive system.

2.3. Working memory

In order to be monitored and executed, the activities specified in LTM are to be instantiated and allocated into WM. Specifically, an activity is enabled in WM once its definition is retrieved from LTM, instantiated with concrete parameters, and linked to the WM structure. This process is recursively repeated with the associated subtasks, until the primitive actions. Notice that multiple competing activities can be allocated and expanded in the WM (e.g., concurrent tasks, alternative conflicting schemata, or instances of the same task/subtask, etc.). In our computational model, this *task set* is represented by an annotated rooted directed graph

(r, S, E), whose nodes $s \in S$ represent tasks to be accomplished, while the edges E represent parental relations among tasks/subtasks. Each node $s \in S$ is denoted by a tuple (b, t, q, e, p), where b is the name of a task, t represents the set of the associated subtasks of b, q represents the enabling condition (precondition), e the activation value, while p is the postcondition. Precondition, postconditions, and activation values are continuously monitored and updated during task execution. An activity/task is enabled when its precondition is satisfied along with the preconditions of all the ancestor nodes in WM; conversely, if a task is accomplished or dismissed (e.g., due to a failure or an external request), this is removed from the WM along with its hierarchical decomposition. Activation values are updated by top-down and bottom-up attention mechanisms as explained below. Following the beverage preparation example, in Eq. 1, we describe a node that implements the abstract subtask *take(glass)*.

$$\begin{aligned} b &= take(glass), \\ t &= \langle goto(glass), pickup(glass) \rangle, \\ q &= on(glass, table), \\ e &= 1, \\ p &= holding(glass). \end{aligned} \tag{1}$$

This subtask is further decomposed into two subtasks: *goto(glass)* (abstract) and *pickup(glass)* (concrete). The node is enabled if the glass is on the table ($on(glass, table)$ precondition) with an activation value of 1 and it is accomplished when the robot holds the glass.

2.4. Activity allocation in WM

The WM structure is managed by the control cycle, which continuously monitors, updates, and allocates/deallocates hierarchical activities. Task allocation may depend on external requests (e.g., a *make(coffee)* command issued by a human or by a planning system), subtask expansion (e.g., a *take(coffee)* method retrieved from LTM to execute *make(coffee)*), or environmental affordance elicitation (e.g., *take(glass)* if *glass* is reachable). If multiple schemata in LTM are eligible to expand the current WM structure, different selection policies can be deployed. For instance, we can allocate a n-best set of instances whose assessed activation values exceed a suitable threshold. More complex activity recruiting mechanisms as in Anderson, Matessa, and Lebiere (1997) and Franklin, Madl, and D'Mello (2014) may be introduced as well. Since the LTM schemata can be mapped into HTN planning domains (Caccavale, Cacace, Fiore, Alami, & Finzi, 2016), a task hierarchy instance can also be externally generated by an HTN planner (de Silva, Lallement, & Alami, 2015; Nau et al., 2003) and then directly allocated in WM. Multiple plans may also be allocated in WM, while concrete activity execution depends on the attention regulations and contention-scheduling mechanisms. Therefore, in contrast to typical AI plan-monitoring systems, the allocated plans do not fully constraint the execution, instead they provide attentional guidance used to bias the executive system toward the accomplishment of planned activities. In Caccavale et al. (2016), we also show how multiple alternative methods competing in WM for the same task can support fast opportunistic plan repairs when the human behavior diverges from the planned one.

2.5. Behavior-based system

The concrete tasks allocated in WM represent real sensorimotor processes, that is, primitive behaviors that can be reactivity executed by the robotic system. Each behavior is endowed with a *perceptual schema* that monitors the operative and environmental state by reading sensors or variables, and a *motor schema* providing commands for the robot actuators or internal updates of the robot state. Only enabled behaviors (along with all the ancestors in WM) can run their motor schemata, while disabled behaviors can only monitor sensors and variables. For each concrete behavior, the activation value is exploited for contention scheduling (see below) but also to regulate a monitoring mechanism: the higher the activation, the higher the resolution/frequency at which the behavior is monitored and controlled. Therefore, multiple behaviors can be active and enabled, with different sampling rates depending on their contextual relevance (Broquère et al., 2014).

2.6. Contentions

Multiple tasks can be concurrently executed, hence several behaviors can compete to acquire shared resources generating conflicts. Contentions are solved by exploiting activations and attention mechanisms: The most activated behaviors are selected following a winner-takes-all approach. The activation values of abstract and concrete nodes in WM are online regulated by top-down and bottom-up stimuli. Intuitively, bottom-up stimuli emphasize concrete nodes of the hierarchy that are more accessible or attractive for the robot. For example, nodes associated with the objects *glass* can be stimulated by the glass proximity. On the other hand, top-down regulations affect nodes at all levels of abstraction and are propagated through the WM hierarchy (e.g., $take(coffee)$ arouses $goto(coffee)$ and $pickup(coffee)$). To allow such propagation of activations, each edge $(i, j) \in E$ of the WM structure is associated with a weight $w_{i,j}$ that regulates the intensity of the attentional influence from the upper node i to the subnode j ($i = j$ is used to weight the bottom-up influence for i). The overall *emphasis* value e_j associated with each node j in the WM is obtained from the weighted sum of the contributions c_i combining top-down and bottom-up influences, that is, $e_j = \sum_i w_{i,j} c_i$ (see Fig. 2, left). This hierarchical propagation can be complemented by additional attention regulation mechanisms. In particular, we introduce a top-down mechanism (*teleology*) to induce the system toward task accomplishment: for each successfully accomplished subtask (postcondition satisfied), the parent activation value is increased by a k value (i.e., n accomplished subtasks provide a $k\,n$ increment), which is suitably weighted and propagated to the successor nodes.

For instance, in Fig. 2, two abstract nodes $take(glass)$ and $take(bottle)$ propagate their values towards two conflicting concrete nodes $goto(glass)$ and $goto(bottle)$. In this case, $goto(glass)$ receives a bottom-up regulation

given by the proximity of the glass along with a top-down regulation from $take(glass)$ (due to the accomplished subtasks), instead $goto(bottle)$ receives the bottom-up proximity-based regulation only. As a result of the additional top-down contribution, $goto(glass)$ can win the competition even if the bottle is closer than the glass. *Lateral inhibition* mechanisms after contention (Cooper & Shallice, 2000) can be introduced as well, we usually neglect them in

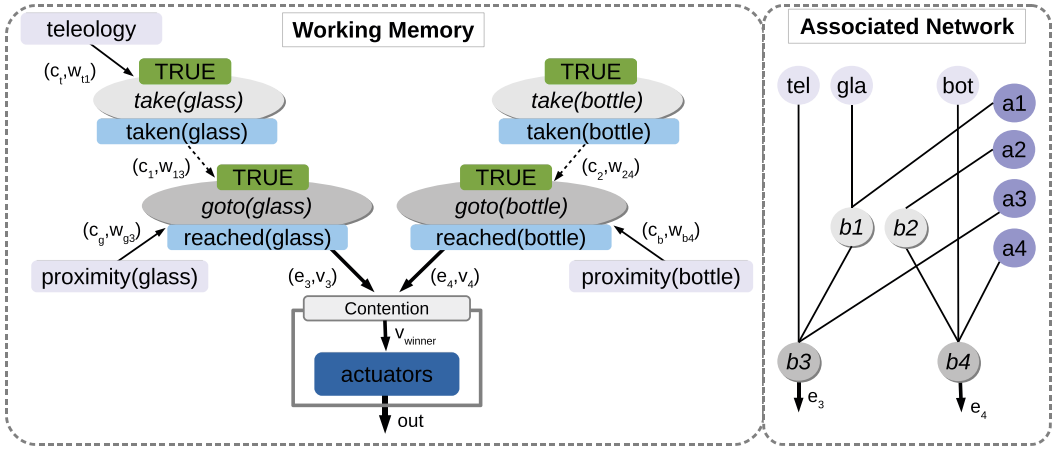

Fig. 2. Conflicting tasks in WM (left) and the associated network representation (right). Concrete behaviors (dark-gray ovals) belonging to conflicting tasks (light-gray ovals) compete for shared resources (robot actuators). Their *emphasis* values (e_3 and e_4) are affected by object proximity (bottom-up) and (top-down) aroused by task continuation drives (*teleology*). In the associated network (right), *b* nodes are for behaviors enabled by *a* provided as inputs along with regulations tel, gla, bot, emphasis values e_i are the outputs.

human–robot interaction applications to keep active and enabled alternatives for rapid task switching (Caccavale & Finzi, 2017) or plan repair (Caccavale et al., 2016).

2.7. Adaptive regulations

Given a task set allocated in WM, the weights are to be suitably regulated to trade-off flexible task switching and effective task accomplishment. For this purpose, the WM structure can be implicitly associated with a multilayered feed-forward neural network (see Fig. 2, right), whose nodes and edges represent, respectively, activities and hierarchical relations between them (Caccavale & Finzi, 2019). Such implicit mapping enables us to combine neural-based learning with symbolic activity representations (needed for task planning and flexible task execution). The network receives as input two vectors representing the enabled activities $\vec{a}$ and the associated attentional influences $\vec{r}$ (respectively, $a_1, \ldots, a_4$ and tel, gla, bot in Fig. 2), while it generates in output the activation values $\vec{e}$ used to regulate the competitions on contended variables. The weights $\vec{w}$ are then updated exploiting a backpropagation method (see Caccavale and Finzi (2019) for details). This way, the system can be online trained by a user that supervises task execution and takes the robot control to adjust its behavior. In this setting, the difference between the system behavior and the human correction is interpreted as an error to be backpropagated through the task hierarchy to adapt the associated weights. Following this approach, each task set allocated in WM can be associated with a trained vector $\vec{w}$ of weights to be suitably stored and retrieved when an analogous task set is obtained again in WM.

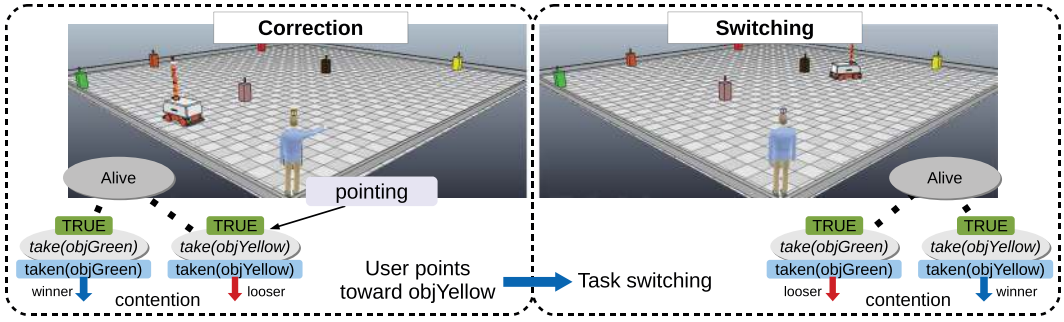

Fig. 3. Example of task switching from $take(objGreen)$ to $take(objYellow)$. Initially (left), the robot is approaching the most emphasized green object, but human pointing towards the yellow object arouses $take(objYellow)$ and induces the robot to switch task and move towards the yellow-object target (right).

3. Case studies: Collaborative task execution and incremental learning

The proposed cognitive control framework has been deployed in several real-world robotic scenarios, where attentional regulations and flexible task execution are used to rapidly adapt the robotic activities to environmental changes and human interventions or to incrementally learn/refine novel tasks from human guidance and demonstrations. In the following, we discuss some case studies to highlight the main features of the system.

3.1. Flexible and collaborative execution of multiple tasks

An attention-based executive system is particularly suited for human–robot collaboration since it naturally provides regulation mechanisms enabling smooth task switching in the presence of multiple concurrent structured tasks to be accomplished. Moreover, such mechanisms can also be manipulated and monitored by the operator in so enabling a natural implicit interaction between the human and the robotic system.

We deployed and demonstrated such features in several human–robot collaboration scenarios (Cacace, Caccavale, Finzi, & Lippiello, 2018; Caccavale et al., 2016; Caccavale & Finzi, 2017). For instance, Fig. 3 shows a mobile robot involved in pick carry and place tasks (e.g., collecting ingredients for beverage preparation) which can be influenced by human guidance. In this case, activations of concrete behaviors are bottom-up affected by target proximity (e.g., activations for $take(objGreen)$ inversely proportional to the object distance), while the task structure provides top-down influences (e.g., teleology). Moreover, the robot activities are also affected by human pointing gestures, which are monitored and detected by an activity recognition module (provided by the HRI system in Fig. 1) along with the referred objects. Specifically, pointed targets enhance the activations of the associated behaviors in WM (e.g., $take(Yellow)$ aroused by human pointing in Fig. 3). This way, when multiple tasks are allocated in WM (e.g., collect and deliver different sets of objects with different ordering constraints), a human supervisor can smoothly induce the robot to switch from one task to another (see Fig. 3) with intuitive and controllable guidance based on simple cueing

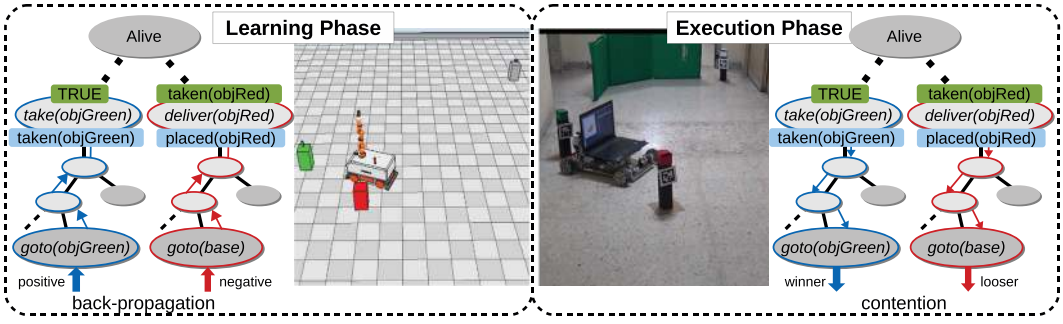

Fig. 4. Training multiple pick-and-place tasks. The system is first trained in different simulated environments with a variable number of objects (left) then tested on a real robot (right).

(see Caccavale & Finzi, 2017, for an empirical assessment). We believe that such manipulation of the robot executive attention, which is directly enabled by a SAS-based system, is a primitive and crucial mechanism for human–robot communication/collaboration in natural everyday scenarios. Similar mechanisms can be deployed to support a smooth interaction during human–robot collaborative manipulation with physical interaction. For instance, in Cacace et al. (2018), the estimated human intentions from the human–robot physical contacts are exploited to online select and adjust the robotic tasks/subtasks or motions, while also regulating the robot compliance with respect to human physical guidance.

3.2. *Learning attention regulations from human guidance*

In the proposed framework, for each task set in WM, the integration of top-down and bottom-up attention regulations depends on weights to be suitably adapted to trade-off flexible task switching and effective task accomplishment.

These regulations can be interactively trained by exploiting human demonstrations. An exemplification of this process is provided in Fig. 4, where a mobile manipulator is tasked to take two colored objects (red and green) and deliver them to a target location (in the right-upper side). Since the two *carrying* tasks are not forced to be sequential, the system can decide how to schedule them given bottom-up (objects *proximity*) and top-down (*teleology*) regulations. During the execution, a human supervisor can always take the robot control to correct the tasks execution through teleoperation. For instance, the human corrections can induce the system to first collect the two objects together and then deliver them to the target location. In Caccavale and Finzi (2019), we show how incrementally structured mobile manipulation activities can be trained in this manner. The trained system is then assessed by checking correct conflict resolutions and by evaluating the overall task performance (e.g., delivered objects, time to deliver, etc.). However, the more complex and diverse the activities in WM are, the more difficult it is to find an adaptive regulation suitable for several operational contexts; therefore, the approach is effective if we can keep limited task sets in WM, each associated with its specific weight regulation. Notice that such limitation does not

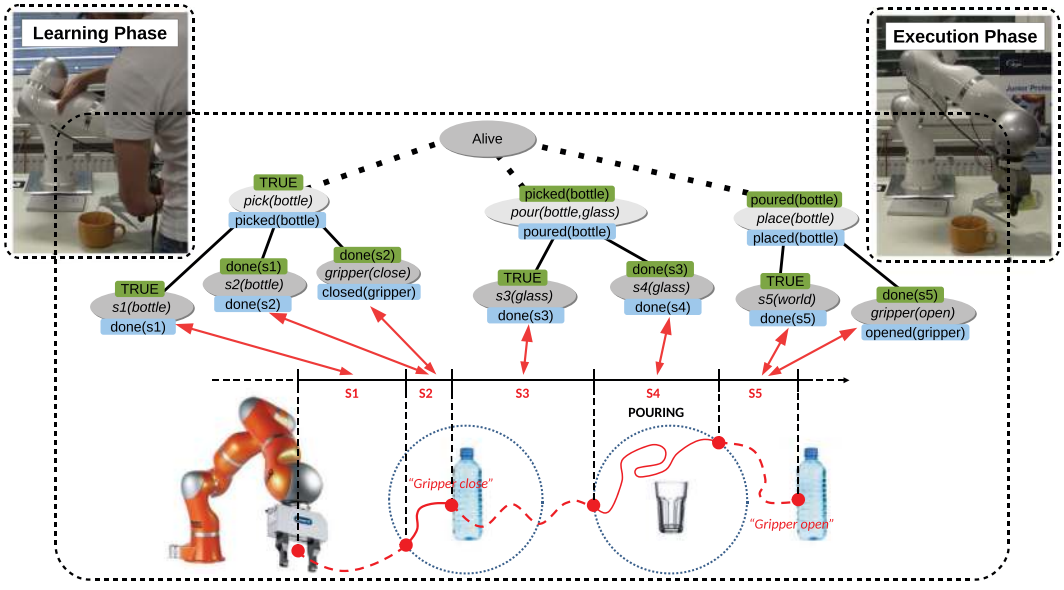

Fig. 5. Kinesthetic teaching of the water-pouring task. The human demonstration is simultaneously segmented ($S1, \ldots, S5$) and monitored by abstract tasks in WM (e.g., $pick(bottle)$). For each segment, a control model is generated (e.g., $s1(bottle)$) and linked to the most emphasized lower active node in WM (e.g., $pick(bottle)$).

directly impair complex task execution and long-term autonomy, since task sets in WM are not static, but continuously retrieved from LTM and allocated in WM or deallocated (e.g., once accomplished or in the presence of failures). In this context, the learned weight regulations could be suitably stored in LTM to be retrieved (when similar sets of activities are recognized in WM) and, in case, readjusted through additional user corrections.

3.3. *Incremental teaching/learning of novel activities*

Lifelong learning of incrementally complex activities is a key capability to enable long-term robot autonomy in everyday scenarios. In this respect, attention-based task supervision and execution provide natural and effective support to task teaching and learning from demonstrations. This feature is illustrated in Caccavale et al. (2019), where the proposed framework is exploited to enable kinesthetic teaching of hierarchically structured tasks. In this scenario, the operator can physically guide a robot manipulator to demonstrate how to execute complex operations (e.g., how to make a soluble coffee or a tea). The human demonstration is supervised by the attentional executive system, which tracks and monitors both the human and the robot activities at different levels of abstraction exploiting abstract/incomplete descriptions of the tasks/subtasks to be demonstrated (e.g., water pouring in Fig. 5). The human demonstration is simultaneously segmented ($S1, \ldots, S5$ in Fig. 5) to generate low-level control models (sensorimotor processes), which are linked to the abstract task structure, providing them with concrete/executable primitives. During the demonstration, attention manipulation

(object or verbal cueing) can be exploited by the human to facilitate hierarchical task monitoring and the match between (top-down) proposed tasks/subtasks and (bottom-up) generated segments/models.

This process is exemplified in Fig. 5. As explained above, during the water-pouring demonstration, the system is already provided with an incomplete and abstract description of the tasks/subtasks, that is, $pick(bottle)$, $pour(bottle, glass)$, $place(bottle)$, which are used to monitor the operator demonstration. On the other hand, the abstract task is not executable since the concrete control models associated with these three abstract nodes are to be online learned from the human guidance. During the kinesthetic demonstration, a low-level monitoring process segments the trajectory illustrated by the human and generates a dynamic model (Ijspeert, Nakanishi, Pastor, Hoffmann, & Schaal, 2013) for each segment ($s1(bottle), \ldots, gripper(open)$ in Fig. 5) that is linked to the abstract nodes in WM (e.g., $s1(bottle)$ linked to $pick(bottle)$). New segments are here online generated when the robot arm enters/leaves an object proximity region (e.g., $S2$ starts in the bottle proximity) or when the human commands something (e.g., $S2$ ends with "gripper open"), while contention scheduling is exploited to associate such generated segments/models to the most emphasized abstract node among the lower nodes enabled in WM. During the teaching process, the novel concrete activities along with the associated preconditions, effects, and hierarchical relations are also stored into the LTM, ready to be retrieved and allocated in WM. This way, the demonstrated task can be autonomously and flexibly executed by the robotic system (e.g., during coffee or tea making). Notice that the same abstract tasks allocated in WM to monitor the human demonstration are also exploited for autonomous and flexible task execution. This is a simple mirroring mechanism, which is naturally provided by a SAS-based cognitive control framework. We are currently investigating methods to simultaneously learn abstract and concrete tasks from multiple demonstrations.

4. Related works

In the AI and robotics literature, flexible execution of complex human–robot collaborative activities is usually managed exploiting integrated planning and execution frameworks, where the human interventions are continuously aligned to the planned activities exploiting replanning (or plan repair) cycles (Carbone et al., 2008; Karpas et al., 2015). This is a typical approach also in human-aware planning systems (de Silva et al., 2015), where structured plans are generated for both the human and the robot agents involved in cooperative activities, and then generated again when the human behavior diverges from the expected one. Such continuous replanning process is computationally expensive and not well legible from the human side, since the robot motions are fragmented, while tasks are frequently interrupted and restarted.

In contrast to these approaches, in our framework multiple competing collaborative plans can be concurrently allocated and enabled in WM. This way, a complex task set can be monitored and executed, while task switching can be smoothly affected by attention regulation mechanisms and contention scheduling, which are also comprehensible and controllable from

the human side. On the other hand, AI methods for task supervision, failure detection, plan repair, etc. could also be available to enable safe and effective task execution (Caccavale et al., 2016); for this purpose, we rely on symbolic task and action representations (e.g., HTNs; Nau et al., 2003), which are compliant with these techniques.

As far as human–robot collaboration is concerned, attention models are usually deployed for visual perception and exploited for implicit nonverbal communication (Breazeal, Kidd, Thomaz, Hoffman, & Berlin, 2005; Muller & Knoll, 2009), joint attention (Scassellati, 1999), anticipation (Hoffman & Breazeal, 2007), perspective taking (Trafton et al., 2005), active perception (Breazeal et al., 2001; Demiris & Khadhouri, 2006), etc. In contrast, we propose attention mechanisms for executive control, which are rarely considered in the robot literature (Broquère et al., 2014; Garforth et al., 2006; Kawamura et al., 2008) and usually not exploited for the orchestration of concurrent structured tasks in real-world robotics systems.

Attention-based contention scheduling has been investigated for behavior selection in behavior-based systems (Scheutz & Andronache, 2004), more complex, hierarchical, representations of goals are considered in Kasderidis and Taylor (2004), where several attentional mechanisms (sensory, motor, boundary attention, etc.) are deployed in combination with heterogeneous attributes (commitment, engagement, emotion, etc.) for goal prioritization and selection. Conversely, we propose to deploy a supervisory attentional system as a uniform executive paradigm for activity orchestration.

Cognitive control mechanisms have also been proposed within the the Intelligent Soft Arm Control (ISAC) architecture (Kawamura et al., 2008), where attention is mainly deployed to assign priority values to orient the focus of perception. More related to our approach, a neural SAS-based executive system for robot control in a simulated environment is proposed by Garforth et al. (2006), where only simple foraging tasks are considered as a proof of concept. In contrast, we are interested in a practical framework that can scale the complexity of real-world robotic collaborative tasks in everyday scenarios. In this direction, we pursue a hybrid neurosymbolic (d'Avila Garcez, & Lamb, 2020) approach to keep the system not only compatible with AI planning and execution methods but also modular, extensible, and explainable for users/developers.

Attentional processes are also crucial in the global workspace theory (GWT) paradigm, (Baars, 1997; Franklin et al., 2014) to recruit content in the workspace. In this setting, the integration of deep-learning attention-model (e.g., transformer networks; Chen et al., 2021) for memory access and content retrieval is an interesting line of active research (Bengio, 2019; VanRullen & Kanai, 2021). In our framework, we focus on attention mechanisms for task orchestration, but analogous complementary attention-based methods could be investigated and integrated to enable context-dependent task set recruitment (from LTM to WM).

5. Conclusion

Executive attention mechanisms are considered crucial regulators for activity orchestration in cognitive neuroscience; however, they are usually neglected in robotic systems. In

contrast, we believe that attention-based control provides an effective paradigm that supports flexible orchestration of multiple concurrent structured tasks, while enabling natural human–robot collaboration. In support of this claim, we provided an overview of a practical SAS-inspired robotic cognitive control framework suitable for collaborative task learning and execution in realistic everyday scenarios. The system was designed to be also compatible with external AI and robotics-based methods (i.e., task and motion planning, multimodal communication, dialogue management, etc.). We illustrated the main features of the framework and discussed case studies to highlight the relevance of attentional supervision/regulation for orchestration of multiple tasks and smooth task switching, activity monitoring and implicit human–robot communication, and incremental learning from demonstration. In contrast with typical cognitive architecture approaches (Anderson et al., 1997; Franklin et al., 2014; Kawamura et al., 2008; Trafton et al., 2005), where several interacting components are involved in activity execution, we illustrated an executive framework based on restricted executive mechanisms inspired by a SAS paradigm (i.e., structured tasks execution, activation values, attention mechanisms, contention scheduling, etc.) and suitable for robot cognitive control. Such mechanisms are rarely deployed for robot task orchestration, typically in combination with multiple heterogeneous mechanisms (Kasderidis & Taylor, 2004; Kawamura et al., 2008) or not intended for realistic robotic scenarios (Garforth et al., 2006). Moreover, the effectiveness of the proposed paradigm and the associated processes to human–robot collaborative task execution (Cacace et al., 2018; Caccavale & Finzi, 2017; Caccavale et al., 2016) and learning (Caccavale & Finzi, 2019; Caccavale et al., 2019) also adds weight to claims in the cognitive psychological literature that action schemata modulated by high level and attention-based control mechanisms (Cooper, 2021; Cooper & Shallice, 2006) play a relevant role in the performance of everyday activities. The proposed case studies also suggest that the SAS paradigm not only supports flexible execution of multiple tasks but also implicit communication and incremental learning.

To further support natural human–robot collaboration during both task teaching and flexible task execution, additional and complementary attentional mechanisms could be integrated into our framework (visual attention, joint attention, active perception, affordances, etc.). We are particularly interested in interaction scenarios where executive attention is affected by multimodal sources (e.g., utterance, gaze direction, gestures, physical interaction, body postures, etc.). Preliminary examples of integrated frameworks for multimodal attention-based interaction can be found in Caccavale et al. (2016). We are also investigating the effectiveness of the described attention-based executive framework for long-term autonomy in complex everyday scenarios. In this context, the aim is to further develop incremental task teaching and adaptation, from primitive to complex tasks. In Caccavale et al. (2019), we assumed already available abstract task descriptions in LTM and considered the problem of grounding them to concrete sensorimotor processes through human demonstrations, the approach can be extended to enable hierarchical tasks learning. In this direction, AI methods for symbolic task learning (Zhuo, Muñoz-Avila, & Yang, 2014) could be integrated in a SAS-based framework to simultaneously learn hierarchical tasks, sensorimotor processes, and attention regulations from human demonstrations. Notice also that we mainly considered learning from human demonstration methods, which are particularly suited for

collaborative robotic settings, but unsupervised learning techniques (e.g., attention-based deep reinforcement learning; Mott, Zoran, Chrzanowski, Wierstra, & Rezende, 2019) could also be integrated to enable task refinement by robot self-practice. Incremental learning processes in long-term autonomy scenarios may continuously generate novel, refined, and incrementally structured/specialized task descriptions to be stored in LTM along with the associated regulations. In this setting, effective mechanisms are needed to retrieve and reuse learned tasks depending on the operational and the environmental context. For this purpose, we are currently investigating how to extend the framework with attention-based task recruiting approaches inspired by the GWT framework (Franklin et al., 2014; VanRullen & Kanai, 2021).

Acknowledgments

The research leading to these results has been partially supported by the projects ICOSAF (PON R&I 2014-2020) and HARMONY by EU H2020 R&I under agreement No. 101017008.

References

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). Act-r: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, *12*(4), 439–462.

Baars, B. (1997). *In the theated of consciousness: The workspace of the mind*. Oxford, England: Oxford University Press.

Baddeley, A. (1993). *Working memory or working attention*, (pp. 152–170). Oxford, England: Oxford University Press.

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology: Section A*, *49*(1), 5–28.

Baxter, P., de Greeff, J., & Belpaeme, T. (2013). Cognitive architecture for human-robot interaction: Towards behavioural alignment. *Journal of Biologically Inspired Cognitive Architectures*, *6*, 30–39.

Bengio, Y. (2019). The consciousness prior. arXiv:1709.08568.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652.

Breazeal, C., Edsinger, A., Fitzpatrick, P., & Scassellati, B. (2001). Active vision for sociable robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, *31*(5), 443–453.

Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS-05; pp. 708–713). Piscataway, NJ: IEEE.

Broquère, X., Finzi, A., Mainprice, J., Rossi, S., Sidobre, D., & Staffa, M. (2014). An attentional approach to human-robot interactive manipulation. *International Journal of Social Robotics*, *6*(4), 533–553.

Cacace, J., Caccavale, R., Finzi, A., & Lippiello, V. (2018). Interactive plan execution during human-robot cooperative manipulation. *IFAC-PapersOnLine*, *51*(22), 500–505.

Caccavale, R., Cacace, J., Fiore, M., Alami, R., & Finzi, A. (2016). Attentional supervision of human-robot collaborative plans. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication* (RO-MAN-16; pp. 867–873). Piscataway, NJ: IEEE.

Caccavale, R., & Finzi, A. (2017). Flexible task execution and attentional regulations in human-robot interaction. *IEEE Transactions on Cognitive and Developmental Systems*, *9*(1), 68–79.

Caccavale, R., & Finzi, A. (2019). Learning attentional regulations for structured tasks execution in robotic cognitive control. *Autonomous Robots*, *43*(8), 2229–2243.

Caccavale, R., Saveriano, M., Finzi, A., & Lee, D. (2019). Kinesthetic teaching and attentional supervision of structured tasks in human-robot interaction. *Autonomous Robots*, *43*(6), 1291–1307.

Caccavale, R., Saveriano, M., Fontanelli, G. A., Ficuciello, F., Lee, D., & Finzi, A. (2017). Imitation learning and attentional supervision of dual-arm structured tasks. In *2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob-17*, pp. 66–71). Piscataway, NJ: IEEE.

Carbone, A., Finzi, A., Orlandini, A., & Pirri, F. (2008). Model-based control architecture for attentive robots in rescue scenarios. *Autonomous Robots*, *24*(1), 87–120.

Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., & Liu, Y. (2021). Deep learning for sensor-based human activity recognition: Overview, challenges and opportunities. arXiv:2001.07416.

Cooper, R. (2021). Action production and event perception as routine sequential behaviors. *Topics in Cognitive Science*, *13*(1), 63–78.

Cooper, R., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*, 297–338.

Cooper, R., & Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review*, *113*(4), 887–916.

d'Avila Garcez, A., & Lamb, L. C. (2020). Neurosymbolic AI: The 3rd wave. arXiv:2012.05876.

de Silva, L., Lallement, R., & Alami, R. (2015). The HATP hierarchical planner: Formalisation and an initial study of its usability and practicality. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 6465–6472). Piscataway, NJ: IEEE.

Demiris, Y., & Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition of actions. *Robotics and Autonomous Systems*, *54*(5), 361–369.

Fikes, R. E., & Nilsson, N. J. (1971). Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, *2*, 189.

Franklin, S., Madl, T., & D'Mello, S. (2014). Lida: A systems-level architecture for cognition, emotion, and learning. *IEEE Transactions on Autonomous Mental Development*, *6*(1), 19–41.

Garforth, J., McHale, S. L., & Meehan, A. (2006). Executive attention, task selection and attention-based learning in a neurally controlled simulated robot. *Neurocomputing*, *69*(16–18), 1923–1945.

Hoffman, G., & Breazeal, C. (2007). Cost-based anticipatory action selection for human–robot fluency. *IEEE Transactions. on Robotics*, *23*(5), 952–961.

Ijspeert, A., Nakanishi, J., Pastor, P., Hoffmann, H., & Schaal, S. (2013). Dynamical movement primitives: Learning attractor models for motor behaviors. *Neural Computation*, *25*(2), 328–373.

Karpas, E., Levine, S. J., Yu, P., & Williams, B. C. (2015). Robust execution of plans for human-robot teams. In *Proceedings of 25th International Conference on Automated Planning and Scheduling* (ICAPS-15; pp. 342–346). Menlo Park, CA: Association for the Advancement of Artificial Intelligence.

Kasderidis, S., & Taylor, J. (2004). Attentional agents and robot control. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, *8*(2), 69–89.

Kawamura, K., Gordon, S. M., Ratanaswasd, P., Erdemir, E., & Hall, J. F. (2008). Implementation of cognitive control for a humanoid robot. *International Journal of Humanoid Robotics*, *5*(04), 547–586.

Kleijn, R. D., Kachergis, G., & Hommel, B. (2014). Everyday robotic action: Lessons from human action control. *Frontiers Neurorobotics*, *8*, 13.

Lashley. (1951). The problem of serial order in behavior. In L. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–146). New York: Wiley.

Mott, A., Zoran, D., Chrzanowski, M., Wierstra, D., & Rezende, D. J. (2019). Towards interpretable reinforcement learning using attention augmented agents. arXiv:1906.02500.

Muller, T., & Knoll, A. (2009). Attention driven visual processing for an interactive dialog robot. In *Proceedings of 2009 ACM Symposium on Applied Computing (SAC-09*; pp. 1151–1155). New York: ACM Press.

Nau, D., Au, T., Ilghami, O., Kuter, U., Murdock, J. W., Wu, D., & Yaman, F. (2003). SHOP2: An HTN planning system. *Journal of Artificial Intelligence Research*, *20*, 379–404.

Nicolescu, M. N. & Mataric, M. J. (2003). Natural methods for robot task learning: Instructive demonstrations, generalization and practice. In Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-03; pp. 241–248). New York: ACM.

Norman, D. A. & Shallice, T. (1986). Attention to Action. In R. J. Davidson, G. E. Schwartz, D. Shapiro (Eds) *Consciousness and Self-Regulation* (pp. 1–18). Springer, Boston, MA.

Posner, M. I. & Snyder, C. R. R. (1975). Attention and cognitive control. *In information Processing and Cognition: The Loyola Symposium*, (pp. 55–85), R. L. Solso (Ed.), Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ.

Rubinstein, J., Meyer, E., & Evan, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 763–797.

Scassellati, B. (1999). Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot. In C. Nehaniv (Ed.), *Computation for metaphors, analogy, and agents* (pp. 176–195). LNCS, Vol. 1562. Berlin: Springer.

Scheutz, M., & Andronache, V. (2004). Architectural mechanisms for dynamic changes of behavior selection strategies in behavior-based systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *34*(6), 2377–2395.

Scheutz, M., Harris, J., & Schermerhorn, P. (2013). Systematic integration of cognitive and robotic architectures. *Advances in Cognitive Systems*, *2*, 277–296.

Tomasello, M. (2010). *Origins of human communication*. Cambridge, MA: MIT Press.

Trafton, G., Hiatt, L., Harrison, A., Tamborello, F., Khemlani, S., & Schultz, A. (2013). Act-r/e: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction*, *2*(1), 30–55.

Trafton, J. G., Cassimatis, N. L., Bugajska, M. D., Brock, D. P., Mintz, F. E., & Schultz, A. C. (2005). Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, *35*(4), 460–470.

VanRullen, R., & Kanai, R. (2021). Deep learning and the global workspace theory. arXiv:2012.10390.

Zhuo, H. H., Muñoz-Avila, H., & Yang, Q. (2014). Learning hierarchical task network domains from partially observed plan traces. *Artificial Intelligence*, *212*, 134–157.