Wojtak, W., Ferreira, F., Louro, L., Bicho, E., & Erlhagen, W. (2023).
Adaptive timing in a dynamic field architecture for natural human–robot interactions.
Cognitive Systems Research, 101148.
https://doi.org/10.1016/j.cogsys.2023.101148, available online 1 July 2023

Adaptive timing in a dynamic field architecture for natural human-robot interactions

Abstract

A close temporal coordination of actions and goals is crucial for natural and fluent human-robot interactions in collaborative tasks. How to endow an autonomous robot with a basic temporal cognition capacity is an open question. In this paper, we present a neurodynamics approach based on the theoretical framework of dynamic neural fields (DNF) which assumes that timing processes are closely integrated with other cognitive computations. The continuous evolution of neural population activity towards an attractor state provides an implicit sensation of the passage of time. Highly flexible sensorimotor timing can be achieved through manipulations of inputs or initial conditions that affect the speed with which the neural trajectory evolves. We test a DNF-based control architecture in an assembly paradigm in which an assistant hands over a series of pieces which the operator uses among others in the assembly process. By watching two experts, the robot first learns the serial order and relative timing of object transfers to subsequently substitute the assistant in the collaborative task. A dynamic adaptation rule exploiting a perceived temporal mismatch between the expected and the realized transfer timing allows the robot to quickly adapt its proactive motor timing to the pace of the operator even when an additional assembly step delays a handover. Moreover, the self-stabilizing properties of the population dynamics support the fast internal simulation of acquired task knowledge allowing the robot to anticipate serial order errors.

Keywords: temporal cognition, human-robot interactions, neurodynamics, adaptation, error monitoring

1. Introduction

Robots that are designed to assist humans in domestic or industrial workplaces must be capable of transferring objects to the human user for successful

cooperation. Examples include a robot serving a drink (Bohren et al., 2011; Gulletta et al., 2021), helping to unload a dishrack (Huang et al., 2015), or providing tools and parts in a joint assembly task (Glasauer et al., 2010; Bicho et al., 2011; Koene et al., 2014). The transfer may be achieved indirectly by placing the object at a reachable position for the user or by a direct handover. While placing only imposes loose constraints on the object handling, the cognitive and physical demands of object handovers are quite high since the receiver has a lifelong experience with smooth human-human handovers. Various aspects of robot-to-human handovers have been studied over the last couple of years with the ultimate goal to advance towards an efficient and seamless coordination of the joint action which meets the expectancies of the user (Strabala et al., 2013). This includes the role of verbal and non-verbal cues (e.g., commands, gaze, hand pose (Moon et al., 2014; Grigore et al., 2013; Bicho et al., 2010)), human-like motion trajectories (Glasauer et al., 2010; Gulletta et al., 2021), user-aware grasp configurations and modulation of grip forces (Medina et al., 2016; Cini et al., 2019) or different safety aspects (De Santis et al., 2008), just to mention a few (for a recent review see (Ortenzi et al., 2021)). However, to date much less attention has been paid to the important role of a close temporal action coordination for successful team performance (Sebanz and Knoblich, 2009; Maniadakis and Trahanias, 2011). Results of a recent user study in a task in which an assisting robot hands over a series of tools to a mechanics suggest that the temporal precision is more important for the user satisfaction than the spatial aspects of the object transfer (Koene et al., 2014). Some human-robot experiments have investigated computational mechanisms for action synchronization in quasi-rhythmic object transfer and manipulation tasks (Huber et al., 2010; Mörtl et al., 2014) which leaves open the question how the robot may flexibly adjust its action timing to variable temporal constraints of discrete object exchanges. A high level of adaptability must be built into any collaborative robot since the order and timing of object transfers may depend on the preferences of different users (Wilcox et al., 2013). Importantly, close temporal coordination does not necessarily mean that the robot should adopt a proactive strategy which minimizes the user's waiting time by holding out the object the human partner needs next (Huang et al., 2015). Users may want to interact with the robot at their own pace. This requires that the robot delays the handover until the user is ready to receive the object at the exchange position, still ensuring a fluent interaction. This is especially important in situations in which the human performs additional tasks between

consecutive object transfers.

In this paper, we test a robot control architecture for adaptive action timing in an assembly paradigm in which a robot has to hand over a series of objects in the right order and at the right time to an operator. The robot first takes a third-person perspective and learns by watching two human experts executing the task. Subsequently, it substitutes the giver and adapts its action timing to the pace of the receiver which may vary due to the introduction of new assembly steps. The control architecture builds on the neuro-computational principles of a recent Dynamic Neural Field (DNF) model of learning sequences with time constraints (Ferreira et al., 2021). It implements evidence from experimental findings that order and temporal information are bounded together in memory during learning (O'Reilly et al., 2008). The identity of a transferred object is represented through the persistent activity of a specific neural population which increases monotonically with elapsed time since the start of the object transfer at the exchange position. As a result, the neural dynamics of a "perceptual memory field" establish in response to a series of observed handovers a gradient of persistent activity over subpopulations which carries the information about the serial order and the relative timing of the transfer events. The stored information can be recalled in a "decision field" using a ramp-to-threshold dynamics. However, to guide the robot's actions in time and to meet the user's expectancy of a minimal waiting time, the memory gradient should be adapted to account for the duration of the robot's pre-handover phase. Theories of a minimal architecture for joint action postulate that a successful adaptation could be achieved in the simplest cases by direct environmental feedback without the need to explicitly represent the co-worker's task (Vesper et al., 2010). Following this idea, we implement an adaptation dynamic in a "motor memory field" which adjusts the strength of the persistent population activity as a function of an observed temporal mismatch between the arrival times of the two hands at the exchange position. In many cases, however, user-adaptive interactions require that the robot has an explicit model of what the co-worker is expected to do. For instance, to detect errors in the assembly work, the robot should have a memory of the temporal order of all parts that the user is going to manipulate including the ones which are not transferred by the robot. Moreover, the robot should be able to proactively inform the user about an error before it manifests in overt behavior. A fast internal simulation of the co-worker's action plan, which is often considered a fundamental building block of successful joint action (Sebanz et al., 2006), supports such proactive timing of a corrective response.

In the DNF architecture, the evolution of persistent population activity in an "error monitoring field" is driven by a mismatch between the predicted and the observed grasping of a specific object.
The remainder of the article is organized as follows. In Section 2, we describe the joint assembly task and the experimental setup for the HRI experiments. Section 3 presents the DNF model of adaptive timing. The implemented neuro-cognitive mechanisms, the model architecture, and the basic model equations are introduced, and the results of the HRI experiments are discussed. In Section 4, the extensions of the model architecture for prediction and action monitoring are explained and the results of an interaction with an anticipated order error are presented. The final Section 5 provides a critical discussion of the experimental results and some lines of future research.

2. Joint assembly task

2.1. Task description

The design of the experiment was inspired by a manufacturing scenario of a robotic assistant handing over a series of objects to a human operator. The object transfers occur in the context of an assembly task in which the operator puts together pieces of computer hardware inside a computer case. Besides receiving three different objects which are initially located in the workspace of the giver (Agent 1), the operator (Agent 2) also handles pieces which are located in his own reachable workspace (Figure 1).
The first experiment starts with a learning phase in which the robot observes two human experts with well synchronized joint actions executing the assembly task. The start and the end of the collaborative work are indicated by verbal commands given by Agent 2. The robot pays attention to the order and the timing of the three object transfers to subsequently substitute the human giver. The goal is to adapt its movement timing to the pace of the human operator thereby minimizing the waiting times for both agents (Hoffman, 2019). For safety reasons, the speed of the robotic movements is limited. A proactive coordination strategy thus implies that the robot should initiate the transfer of the next object when the operator is still working on the previous assembly step. Since there is no abrupt visual event which may serve as a go signal, the initiation of actions must be internally triggered (Maimon and Assad, 2006). The robot's adaptive timing capacity is further challenged when the operator changes the initial plan and introduces a new piece in the assembly work which delays the timing of subsequent object

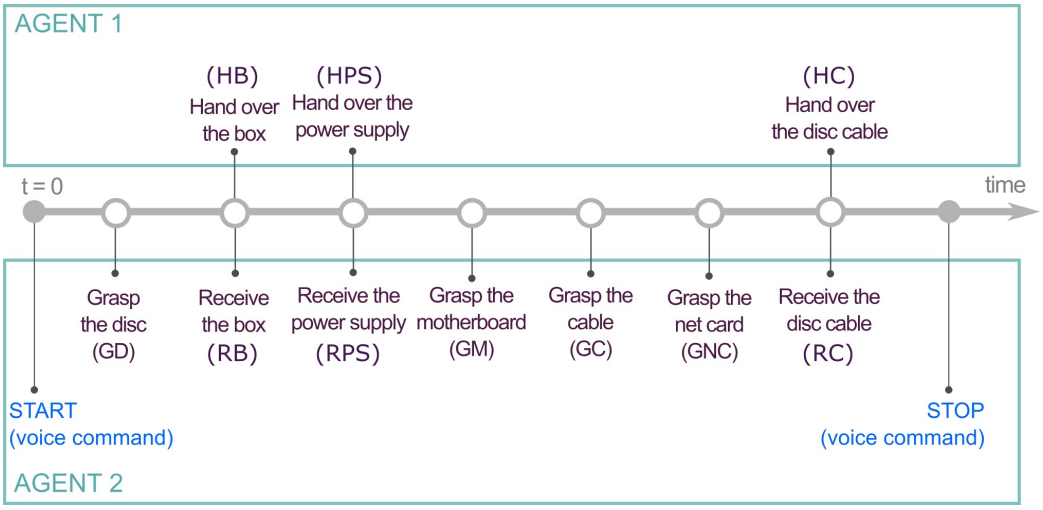

Figure 1: Joint assembly task. The assistant (Agent 1) has to handover in the correct order and at the right time a box, a power supply, and a disc cable to the operator (Agent 2) who also performs additional assembly steps which do not involve the assistant.

transfer.

The second experiment aims at testing an extended joint action architecture which supports user-adaptive behavior based on shared task knowledge about the entire assembly plan (Vesper et al., 2010). Here, the robot pays attention to all object-directed actions performed by the operator including the ones which do not require a handover. The acquired serial order knowledge allows the robot to detect order errors based on a mismatch between the predicted and the observed object-directed action. Importantly, the predictive mechanism should be fast in order to timely communicate to the operator which object he should handle instead.

2.2. Experimental setup

The two agents operate on opposite sides of a table where the computer hardware parts are initially distributed in the two separate working areas (Figure 2). The humanoid robot ARoS (Anthropomorphic Robotic System) was used in the experiments. It consists of a stationary torus on which a 7 DOFs AMTEC arm (Schunk GmbH) with a 3-fingered dexterous hand (Barrett Technology Inc.) and a stereo camera system are mounted. A speech synthesizer/recognizer (Microsoft Speech API 6.2) supports basic verbal communication (Bicho et al., 2010).

To focus on the temporal cognition and action prediction challenges of the task, several aspects have been simplified for the robot's vision and motor systems. The hands and the objects are color coded so that the vision system can robustly detect when the operator's hand approaches a certain object or has reached the exchange position. The spatial dimension x of the neural fields thus represents color. The positions of localized activity patterns or bumps in the field encode the respective objects (using the abbreviations in Figure 1 as labels) while the relative timing information is stored in the bump amplitudes (for details see Section 3). In addition, proprioceptive information is used to detect when the robot hand has realized the planned goal posture at the handover location. Sensory information indicating any delay between the arrival times of the two hands (compare Figure 5) drives the adaptation of the robot's movement onset in order to achieve a more fluent handover in the next joint task execution.

To generate the object-directed robotic movements, we have applied the posture-based motion planner HUMP (Gulletta et al., 2021). It has been developed to include human-like motion features which are known to support more natural and intuitive human-robot interactions. The handovers take place at a pre-specified position roughly midway between the giver and the receiver. The grasping of the objects and the final hand-object pose are chosen to guarantee that the object transfer is safe and comfortable for the user. For the physical object transfer, we used force information from the robot's joints to determine whether or not the object can be released since it was grasped by the human operator.

The focus in our study on the temporal aspect of object transfers is motivated by user studies (Koene et al., 2014) showing that transfer fluency has a much higher weight for user satisfaction than the spatial aspects (e.g., users accept an optional final adjustment for picking up the object from the robot hand). Note that the situation might be different when task constraints (e.g., obstacles) limit the motion of the user (Ortenzi et al., 2021). In this case, models of the kinematics and the dynamics of the human receiver should be taken into account in the motion planning to devise an adequate handover location. In this paper, we address a user model only on the cognitive level of shared task knowledge in order to support an error monitoring capacity.

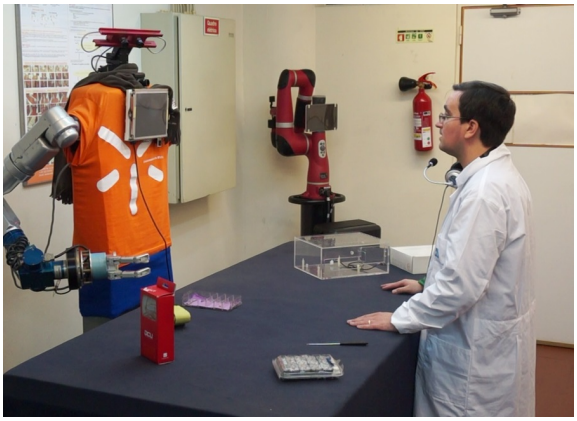

Figure 2: Experimental setup.

3. Dynamic neural field model I: Adaptive timing

3.1. Neuro-computational mechanisms

We start by giving a brief overview about the main neuro-computational processing principles implemented in the DNF model.
The capacity to hold and manipulate items mentally over behaviorally relevant time scales is fundamental for many cognitive processes such as decision making, learning, or planning action sequences. Neural population activity which persists after the removal of triggering input stimuli is commonly believed to represent a critical neural substrate for cognitive behavior (Miller and Cohen, 2001; Curtis and Lee, 2010). Dynamic Neural Field models explain neural activity patterns that remain stable in time by assuming balanced excitatory and inhibitory feedback loops between neurons in populations coding for particular inputs or their behaviorally relevant attributes (Amari, 1977; Schöner, 2016). For continuous-valued information such as the spatial location or the color of a stimulus, the attractor state takes the form of a spatially localized activity pattern or bump. Due to the assumed translation-invariant structure of the recurrent connections, the network can hold a continuous family of bumps, each representing the memory of a specific input value. However, the stationary bump attractor constitutes an incomplete description of stimulus-specific, persistent population activity as observed in the brain. Often, the persistent activity is not static but varies systematically over the course of maintenance. The observed monotonic increase (or decrease) has been interpreted as evidence that the self-sustained

population activity encodes in addition to input identity also elapsed time (Brody et al., 2003; Hass and Durstewitz, 2016). The integration of an additional threshold accommodation dynamics (Coombes and Owen, 2007) accounts for this pattern of persistent "ramping" activity (Ferreira et al., 2021) (for an alternative, spatial DNF representation of elapsed time based on a traveling wave mechanism see (Duran and Sandamirskaya, 2017; Lima et al., 2022)). In response to a series of transient inputs, the field dynamics establishes a stable multi-bump pattern over neural subpopulations with an activation gradient reflecting the relative timing or temporal order of stimulus events. Ramping activity has been also described in motor-related brain areas (for review see (Svoboda and Li, 2018)). Here, the population activity typically starts from a pre-activated state reflecting prior task knowledge (Bastian et al., 1998; Erlhagen and Schöner, 2002) (e.g., a specific object takes part in the sequential task). The preparatory activity evolves toward a threshold to trigger a motor response (e,g., a reach-to-grasp movement directed toward the object). Importantly, experimental evidence shows that the gradual firing rate elevation is not necessarily linked to the integration of any immediate external event but may also represent an internally triggered, proactive timing signal (Maimon and Assad, 2006). Finally, and crucial for our current purposes, the proactive action timing can be efficiently adapted on a fine temporal scale. Sensory feedback indicating a mismatch between the expected and the perceived timing of the desired action effect can be used to adjust the baseline level from which the rise-to-threshold dynamics starts (Erlhagen and Schöner, 2002; Bogacz et al., 2010).

3.2. Model architecture

An overview of the cognitive control architecture consisting of several coupled DNFs is presented in Figure 3. The architecture is divided into a learning part which comprises perceptual and motor memory and an execution part which controls the what and when of the robotic actions. A start command spoken by Agent 2 initiates the assembly work. Modeled as a transient Gaussian input, it triggers the evolution of a bump in the Start field u_{START}. During observation, the presence of this bump controls the linear threshold accommodation dynamics (blue arrows) responsible for bump growth in the Sequence Memory and Task Duration fields, u_{MEM_1} and u_D, respectively. During execution, the suprathreshold activity in u_{START} drives through excitatory connections the rise-to-threshold dynamics in the Action Onset field u_{ACT}. The voice command signaling the end of the joint action

(modeled as a brief inhibitory input) destabilizes the bump and the activity in u_{START} decays back to resting state. The start command also triggers a bump in u_D which at the end of the task represents in its amplitude the total task duration.
When the vision system detects that the receiver's hand and a certain object transported by the giver arrive at the exchange position (compare Figure 4), the color input of the object triggers the evolution of a bump in the Sequence Memory field u_{MEM_1}. A demonstrated sequence of object transfers creates a stable multi-bump pattern in u_{MEM_1} that stores all transfer events with an activation strength decreasing from item to item as a function of elapsed time since the start of the collaboration. The stored information about the order and relative timing of handovers can be sequentially recalled with the linear rise-to-threshold dynamics in the Action Onset field u_{ACT} driven by the bump in u_{START} (Ferreira et al., 2021). The initial subthreshold state of the field is defined by the homogenous inhibitory input from u_D given by the bump amplitude, $max(u_D)$, and the spatially modulated excitatory input from u_{MEM_1}. However, to guarantee a tight synchronization of the robotic actions with the operator, the robot should adapt its pace in accordance with the time it takes to reach, grasp and transport each individual object to the exchange location. This is done in the Action Onset Memory field u_{AMEM} which at the beginning of the first execution trial contains a copy of the activity gradient in u_{MEM_1} mediated by excitatory connections. The two feedback fields, u_{F_1} and u_{F_2}, appear to be preshaped by the activity pattern in u_{MEM_1} and receive input from the vision system about the detection of the receiver's hand (Agent 2) and the object (Agent 1) at the transfer location. Differences in the temporal evolution of the suprathreshold activities drive a local threshold adaptation dynamics (red arrows) resulting in a change of the respective bump amplitude in u_{AMEM} (Wojtak et al., 2017). If the bump evolution starts earlier in u_{F_2} than in u_{F_1}, meaning that the operator had to wait for the object, the amplitude appears to be increased whereas in the opposite case the amplitude decreases. Finally, bumps in the Past Events field u_{PE}, which are triggered by the bumps in u_{F_1} and u_{F_2}, represent the memory of already executed object transfers. Inhibitory connections (dashed line) to u_{F_1}, u_{F_2} and u_{ACT} guarantee that the suprathreshold activity patterns in these fields become suppressed and the next handover is processed.

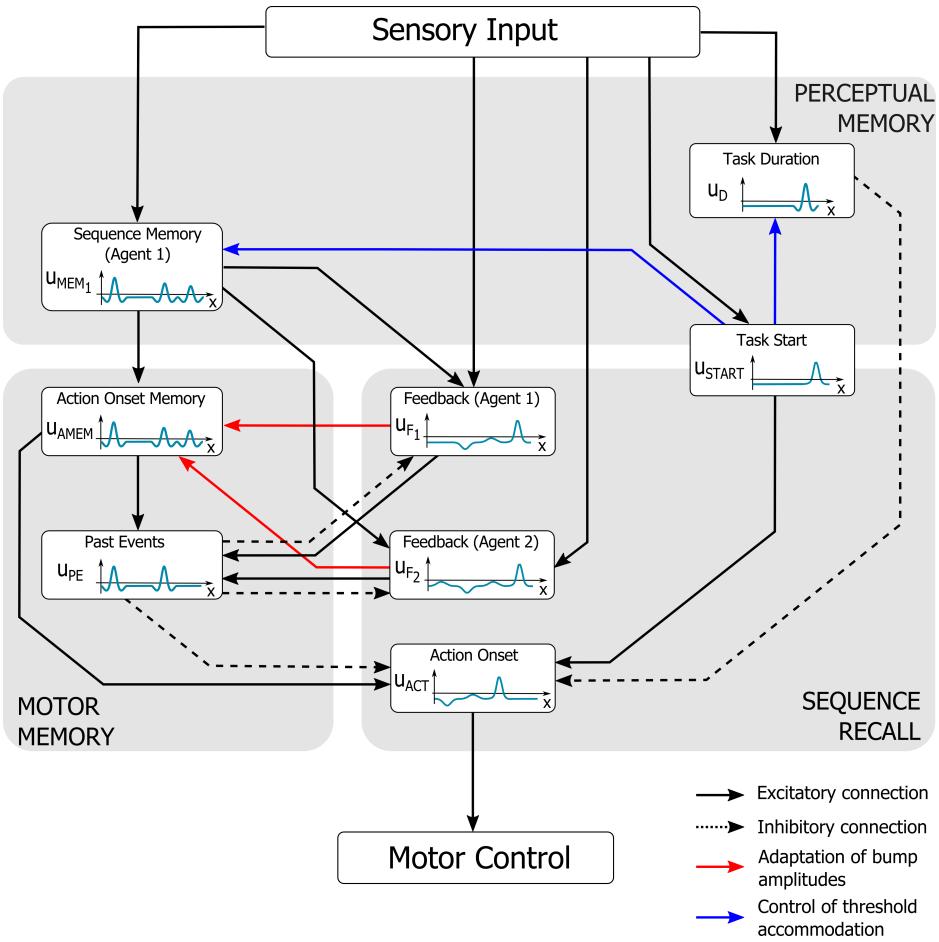

Figure 3: Schematic view of the DNF architecture with several interconnected neural fields implementing perceptual sequence memory, motor sequence memory, and sequence recall. The neural fields are spanned over the continuous dimension color which is used by the vision system to identify the different objects. The spatial positions of the activation bumps indicate which objects are currently processed. Sensory inputs include vision, proprioception and simple voice commands indicating the start and the end of the assembly work. For details see the text.

3.3. Model equations

We describe here the main equations of dynamic field theory that we have used to implement the neural processing mechanisms. A detailed description of each layer of the architecture together with the model parameters can be found in the Appendix.

To implement persistent neural population activity, we use the continuous

attractor network first proposed and analyzed by Amari (1977). In each field, the activity $u(x,t)$ at time t of a neuron at position x, representing a certain value of the continuous color dimension, is governed by the following integro-differential equation:

$$\tau \frac{\partial u(x,t)}{\partial t} = -u(x,t) + \int_{\Omega} w(x-y) f(u(y,t)) \mathrm{d}y - h(t) + S(x,t), \quad (1)$$

where τ defines the time scale of the field dynamics, $S(x,t)$ represents the time dependent localized input at site x from external sources (e.g., speech, vision) and/or connected fields, $h(t)$ defines the baseline level to which the field activity converges without external stimuli, the coupling function $w(x-y)$ determines the distance-dependent connection strength between neurons x and y, and $f(u)$ gives the firing rate which is chosen as the Heaviside function,

$$f(u) = \begin{cases} 1, & u > 0, \\ 0, & u \leq 0. \end{cases} \quad (2)$$

In model layers in which only a single bump may exist at a time (e.g., u_{ACT}) the connectivity function is of lateral inhibition type:

$$w(x) = w_{exc} e^{(x^2/2\sigma_{exc}^2)} - w_{inhib}, \quad (3)$$

where $w_{exc} > 0$ and $\sigma_{exc} > 0$ define, respectively, the amplitude and standard deviation, and $w_{inhib} > 0$ represents a constant lateral inhibition.

To enable in the fields with memory functionality a robust encoding of a series of inputs in a multi-bump pattern, we adopt a connectivity function with oscillatory rather than monotonic decay (Ferreira et al., 2016; Laing et al., 2002):

$$w(x) = A^{-b|x|}(b \sin|\alpha x| + \cos(\alpha x)), \quad (4)$$

where the parameter $b > 0$ controls the rate at which the oscillations in $w(x)$ decay with distance. The parameters A and α control the amplitude and the zero crossings of $w(x)$, respectively.

A multi-bump solution may also exists for a kernel of type (3) with moderate lateral inhibition level w_{inhib} (Erlhagen and Bicho, 2006). However, the creation of this pattern with sequentially presented inputs is much more vulnerable to noise and variations in input characteristics such as strength, duration and width.

The strength of individual event representations in u_{MEM_1} (in an analogous

way also the bump amplitude in u_D) is controlled by a state-dependent dynamics for the baseline activity $h_{MEM_1}(x,t)$ (Coombes and Owen, 2007):

$$\frac{\partial h_{MEM_1}(x,t)}{\partial t} = (1-f(u_{MEM_1}(x,t)))(-h_{MEM_1}(x,t)+h_{MEM_{0_1}})+\frac{1}{\tau_{h_{MEM_1}}}f(u_{MEM_1}(x,t)), \tag{5}$$

where $f(u)$ is again the Heaviside function, $h_{MEM_0} < 0$ defines the level to which h_{MEM} converges without suprathreshold activity at position x and $\tau_{h_{MEM_1}} > 0$ measures the growth rate when it is present.
For the recall of the memorized task information in u_{ACT}, we apply a linear ramping dynamics for the baseline activity driven by the integral of the population activity in u_{START}:

$$\frac{\mathrm{d}h_{ACT}(t)}{\mathrm{d}t} = \beta_{ACT}\int_{\Omega} f(u_{START}(x))\mathrm{d}x, \qquad h_{ACT}(t_0) \leq 0. \tag{6}$$

Changing the value of $\beta_{ACT} > 0$ affects the speed of task execution, while preserving the memorized relative timing of object-directed robotic actions.

The adaptation of the bump amplitudes in u_{AMEM} is based on a comparison between the bump evolution in u_{F_1} and u_{F_2} which should occur simultaneously in case of synchronized team behavior. To compensate for a detected temporal mismatch, we apply a learning rule proposed in (Almeida and Ledberg, 2010) to adapt the baseline level in u_{AMEM} in the k-th trial:

$$h_{AMEM_{k+1}} = \begin{cases} \dfrac{h_{AMEM_k}}{1+\beta}, & \text{if } u_{F_2} > 0 \text{ and } u_{F_1} < 0, \text{ i.e., robot is too late,} \\ \dfrac{h_{AMEM_k}}{1-\beta}, & \text{if } u_{F_2} < 0 \text{ and } u_{F_1} > 0, \text{ i.e., robot is too early,} \\ h_{AMEM_k}, & \text{otherwise.} \end{cases} \tag{7}$$

The change in the baseline activity of u_{AMEM} is formalized by the following h-dynamics:

$$\frac{\partial h_{AMEM}(x,t)}{\partial t} = \beta(1-f(u_{F_2}(x,t))f(u_{F_1}(x,t)))(f(u_{F_1}(x,t))-f(u_{F_2}(x,t))), \tag{8}$$

where $f(u)$ is again the Heaviside function and β is the adaptation rate parameter.
As the required temporal adjustments will most likely differ for each of the

object transfers, we apply the adaptation dynamics locally at field sites with suprathreshold activity representing the currently prepared object-directed action. A larger bump amplitude in u_{AMEM} reduces the gap between the level of pre-activation and the response threshold in u_{ACT}, leading to a faster initiation of the object transfer movements, whereas a reduced bump amplitude has the opposite effect.

3.4. Results

Figure 4 shows snapshots of the learning phase of the first human-robot experiment in which the robot ARoS observes the human team performing a fluent transfer of the three pieces of computer hardware as part of the assembly task (top row). The giver adapted her movement timing to the pace of the operator by taking into consideration his availability to receive the object. The robot's camera view (middle row) reveals that at the moment of each handover, ARoS has detected the respective color-coded object and hands in the exchange area. The object color drives the evolution of a bump in the memory field u_{MEM_1}. Panel (H) depicts the time course of activity of the three subpopulations tuned to the different color inputs where time $t = 0$ indicates the start of the collaboration. Due to the threshold accommodation dynamics, the firing rate consistently increases until the end of the assembly task. Panel (G) shows the established activation gradient encoding the temporal order of the handovers (magenta(HB)-red(HPS)-yellow(HC)).

ARoS is able to autonomously recall the stored order and the relative timing information with the rise-to-threshold dynamics in the action onset field u_{ACT} which appears to be preshaped by the inputs from u_D and u_{AMEM}. However, a tight action synchronization with the receiver is not guaranteed since the recall speed can be too fast or too slow resulting in significant waiting times for the robot or the operator at the exchange position. Figure 5 shows examples from the camera view in which the robotic hand with the object arrives too late (panel A) or too early (panel B).

Assuming equal time constant for the h-dynamics in u_{MEM_1} and u_{ACT} ($\tau_{h_{MEM_1}} = 1/\beta_{ACT}$), in the first joint execution trial, the handover of the first object appears to be delayed by the time the robots takes to transport the object to the exchange position. Since the duration of the robot's pre-handover phases are object (and hardware) specific, a local adaptation of the bump amplitude in u_{AMEM} is required to compensate for the timing errors. Figure 6 shows snapshots with time stamps (in seconds) of the three handovers. In the first joint task execution (top row), ARoS starts the transfers of the box ($t = 33s$)

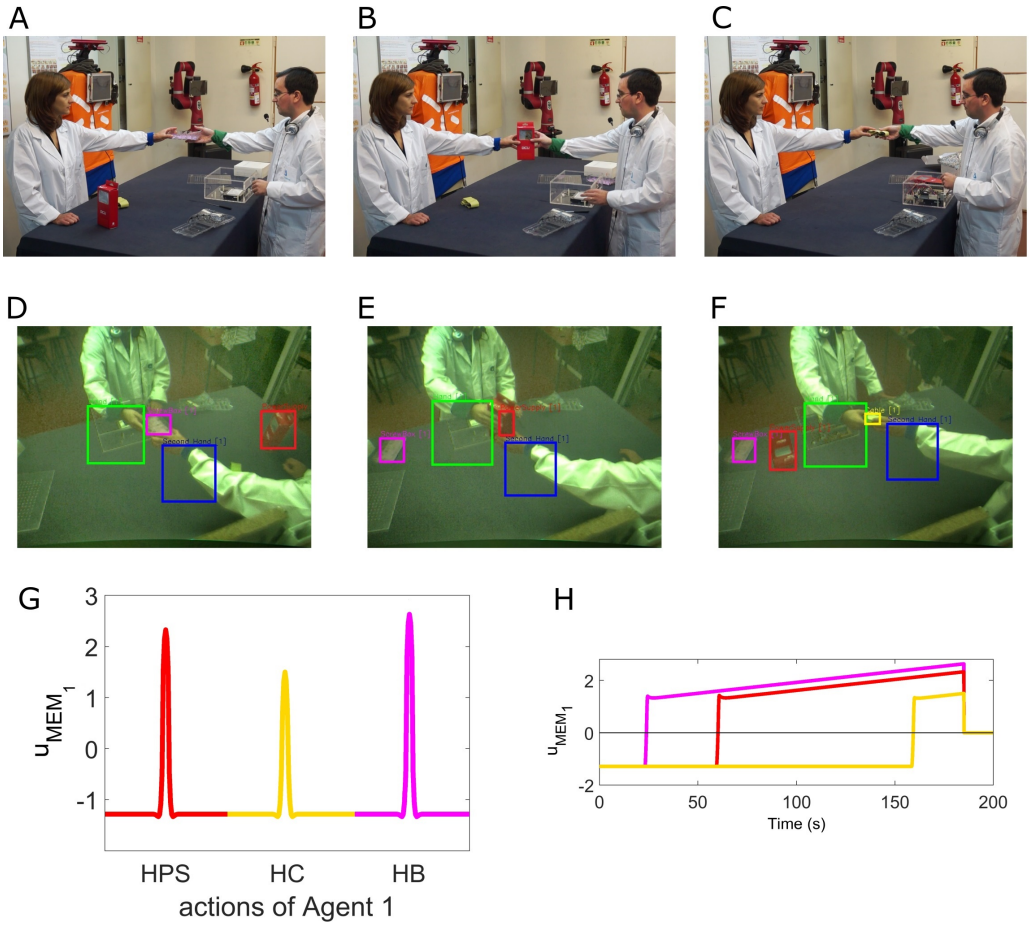

Figure 4: (**A**-**C**) Video snapshots of the sequence learning.(**D**-**F**) Camera view. (**G**) Sequence memory of Agent 1. (**H**) Time course of activation in u_{MEM_1}.

and the disc cable ($t = 175s$) too late whereas it holds out the power supply for the operator when he is still mounting the disc ($t = 84s$). The bottom row shows that in the third joint execution trial, the adapted movement onset timings result in object transfer with minimal waiting times for both agents at the exchange location.

Figure 7 presents a closer look at the preshaped field u_{ACT} at time $t = 0$ (left) and the time course of neural activity in u_{ACT} and in the two feedback fields, u_{F_1} and u_{F_2} (right), for the three execution trials. In the first trial, the preshape of u_{ACT} reflects the memory pattern in u_{MEM_1} (A). A significant

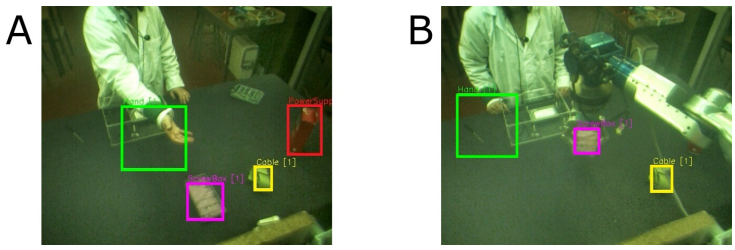

Figure 5: (**A**) Vision input to u_{F_2} during recall while the operator's hand is in a waiting position. (**B**) Robot arm in a waiting position. The input to u_{F_1} is provided by the proprioception of the robotic arm joint angles.

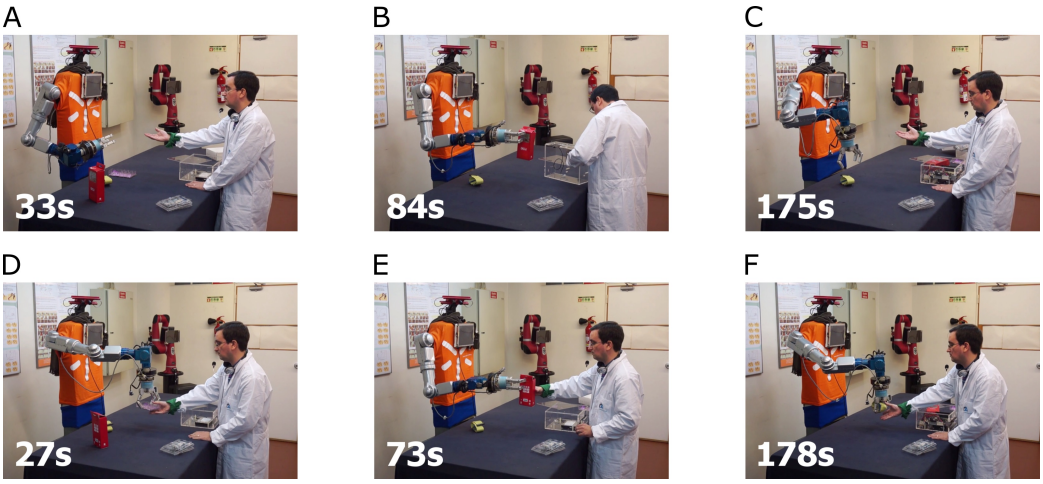

Figure 6: (**A**-**C**) Video snapshots of the first execution trial. The first and the third handovers are delayed by ARoS while it starts the second object transfer too early. (**D**-**F**) Video snapshots of the third execution trial. The timing of all three handovers appears to be synchronized.

waiting time for the operator during the exchange of the first and the third object and a shorter waiting time for the robot during the transfer of the second object can be seen by comparing the time courses in u_{F_1} and u_{F_2} (B). In the second trial, the pre-activation at time $t = 0$ following the adaptation in u_{AMEM} is now higher for objects one and three and lower for object two (C, solid line), resulting in an earlier respectively later movement onset (compare the top panels in B and D). The operator's hand still arrives earlier than the object at the transfer location as shown by the earlier evolution of the bump in u_{F_2}. This temporal offset appears to be compensated in the

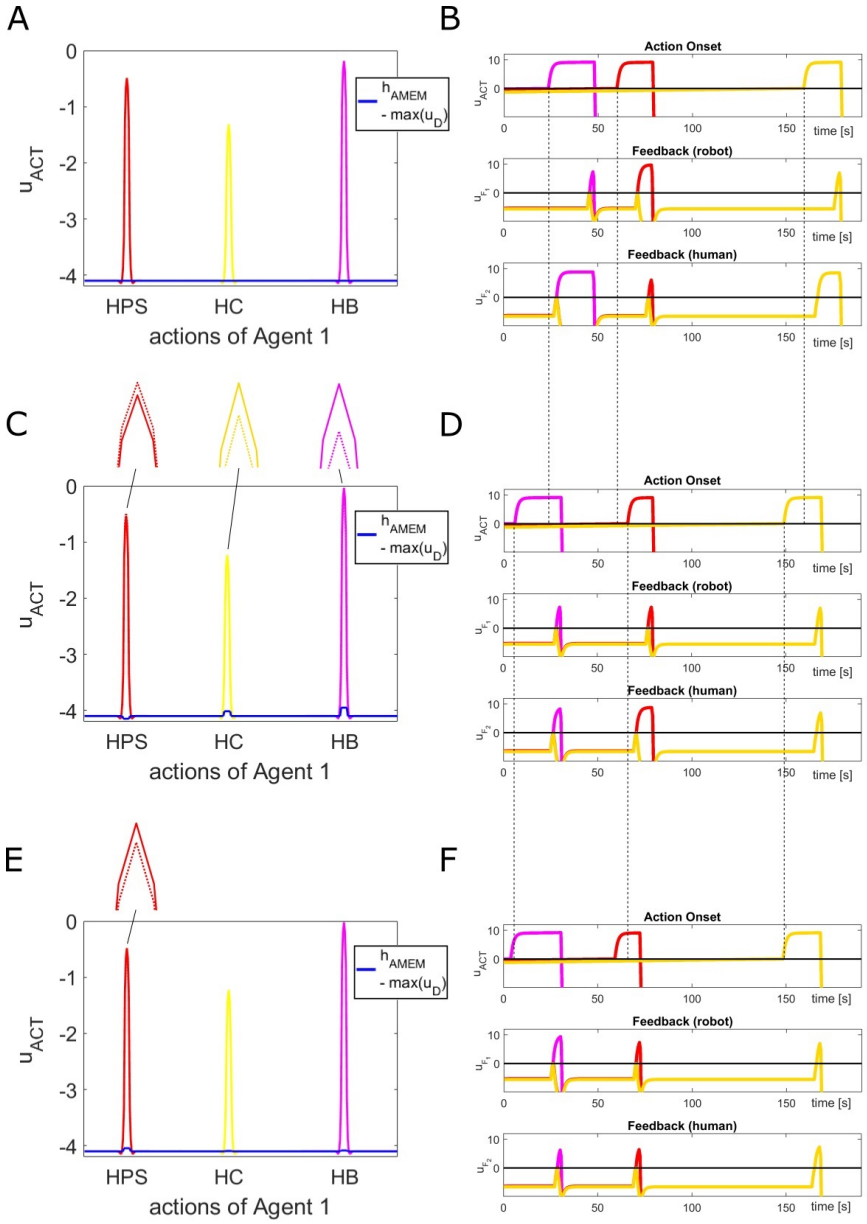

Figure 7: (**A**, **C** and **E**) Action onset field u_{ACT} at time $t = 0$ during the first (**A**), second (**C**) and third (**E**) execution trial. (**B**, **D** and **F**) Time course of activation in u_{ACT}, u_{F_1} and u_{F_2} during the first (**B**), second (**D**) and third (**F**) execution trial.

third trial (F). The robot starts the second handover earlier due to the higher pre-activation at the start of the joint task execution (E). Important for user acceptance, any effort of the robot to increase the interaction fluency signals to the human partner the willingness to cooperate (Koene et al., 2014).
Table 1 summarizes the waiting times for human and robot for all three object transfers in three consecutive joint execution trials. The waiting times are defined as $\Delta t = t_H - t_R$ where t_H and t_R indicate the arrival times of the human hand and the robot hand at the exchange position, respectively. As can be clearly seen, with increasing team experience a drastic reduction of the waiting times for all handovers is achieved, realizing a high level of synchronization. Note that the expected precision of the temporal coordination based on the previous experience is constrained by the fact that the operator's execution time for the different assembly steps varies to some extent from trial to trial.

	1st object	2nd object	3rd object
1st trial	-17	6	-9
2nd trial	-2	-6	-1
3rd trial	2	0	-1

Table 1: Waiting time $\Delta t = t_H - t_R$ (seconds). Negative values indicate that the human hand arrived first.

Figure 8 shows the results of an additional experiment in which the operator introduced a memory card as a new piece between the second and third handover in the assembly process (A). ARoS is not aware of this change in the initial assembly plan but is still able to adapt its action timing in order to transfer the third object when the operator is ready to receive it. The time course of population activity in the fields u_{ACT}, u_{F_1} and u_{F_2} before (B) and after (C) adaptation shows that the robot starts the third transfer with an additional delay, Δt, causing a synchronous bump evolution in the two feedback fields.
It is important to emphasize that the feedback mechanism works irrespectively of the reason for an observed temporal mismatch. For instance, after the first interactions, the operator may have reduced his overall movement speed since he perceives the robot in general as less agile compared to a human co-worker. The robot in turn will automatically adjust the initiation of the object transfer movements to the new pace in order to achieve

more synchronous handovers. Such mutual adaptation between co-workers is a common observation in human-human collaborative tasks (Sebanz and Knoblich, 2009).

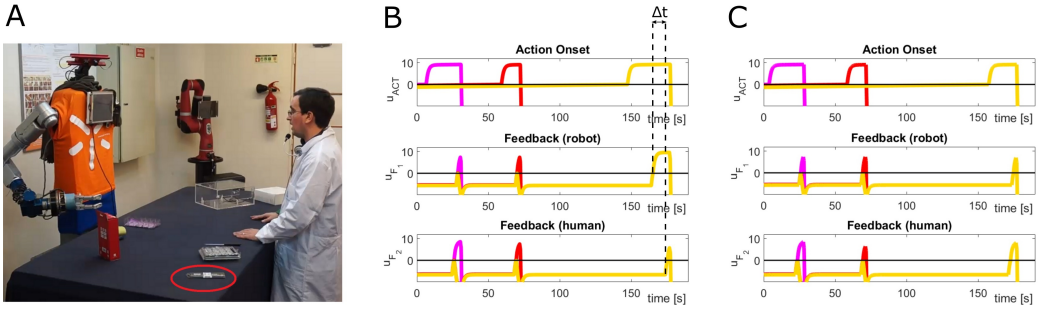

Figure 8: (**A**) The operator assembles an additional memory card which is located in his workspace (red circle). (**B-C**) Temporal evolution of activity in u_{ACT}, u_{F_1} and u_{F_2} before (**B**) and after (**C**) adaptation. The waiting time Δt changes from +10s for the robot to -1s for the human operator.

4. Dynamic Field Model II: Monitoring and prediction

4.1. Extended DNF architecture

In the previous experiments, the precise temporal coordination of transfer actions has been achieved based on a short-term adaptation to the operator's behavior without knowing what the partner should do next. In many cases, successful and fluent team performance will benefit from shared task knowledge. A concrete example is the ability to detect errors based on an online comparison between the actual and the predicted goal of the partner's action. Ideally, errors should be communicated to the partner as fast as possible before they manifest in overt behavior. Figure 9 depicts an extended model architecture with three new fields (green box) which support this ability. A stable activation gradient in the field u_{MEM_2} encodes the information about the temporal order in which Agent 2 manipulates all assembly parts. The color cue driving the evolution of the activity pattern is the moment of grasping the respective part at the exchange position during handover or at the location in the operator's workspace. Driven by the bump in u_{START}, the stored information is sequentially recalled in u_{SIM}. This can be seen as an internal simulation of the operator's task. Since the objective is to anticipate the operator's next goal, the bump evolutions should happen on a much

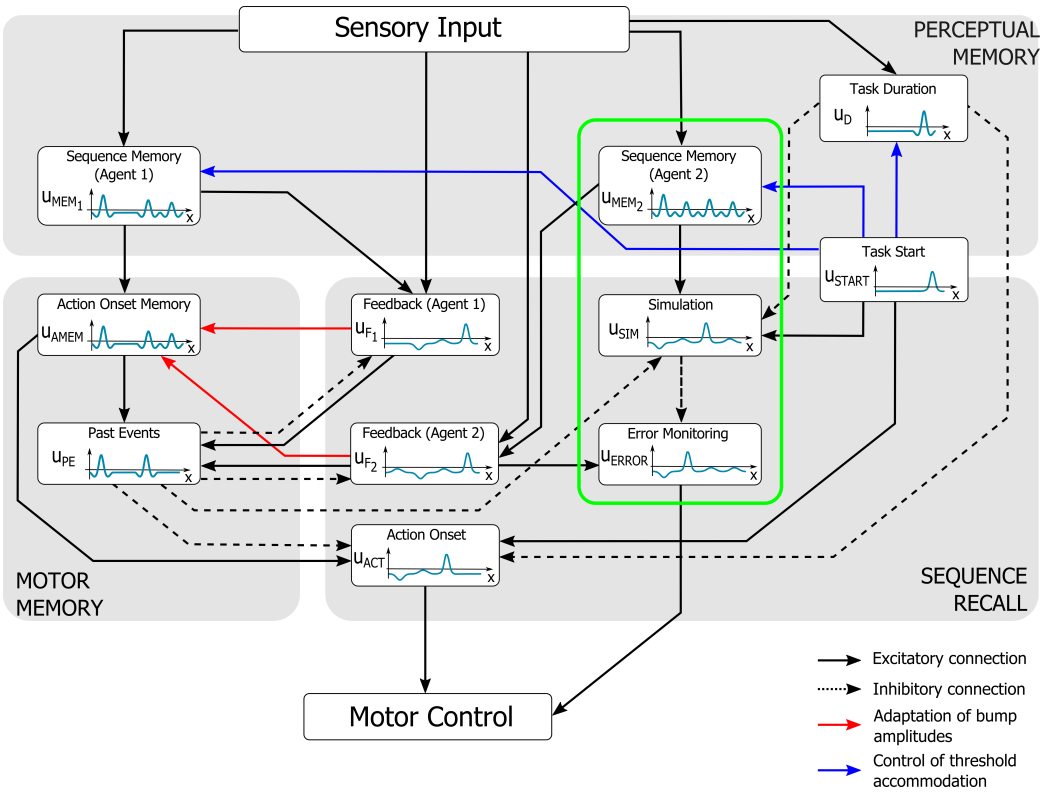

Figure 9: Schematic view of the extended DNF architecture supporting prediction and error monitoring based on shared task knowledge. For details see the text.

faster time scale compared to the encoding process. This can be achieved by choosing a sufficiently large strength parameter $\beta_{SIM} > 0$ for the input integration which increases the slope of the linear ramping dynamics. Mediated by inhibitory connections, a bump in u_{SIM} representing a prediction suppresses the neurons at the corresponding site of the Error Monitoring field, u_{ERROR}, below the resting level. Such a reduction in neural activity associated with an expected sensory event is broadly referred to as "expectation suppression" in the neuroscience literature (De Lange et al., 2018). The net effect is that excitatory input from the feedback field u_{F_2} representing the object Agent 2 is actually going to manipulate may drive a bump in u_{ERROR} in case of a mismatch only. Moreover, the bump positions in u_{SIM} and u_{ERROR} can be in principle read out by the system to communicate the serial order error to the operator (e.g., "not object B, but object A").

4.2. Results

Figure 10 shows an example of a successful collaboration in which the robot has established during observation the activation gradients for the serial order of task execution for the assistant (A) and the operator (B). The temporal evolution of activity in the two feedback fields in the first joint execution trial (C) reveals that the collaboration occurred without error but the handovers of the first object (box) and of the last object (disk cable) were not executed in synchrony. As shown in panel (D), the timing errors have been corrected in the fourth joint execution trial.

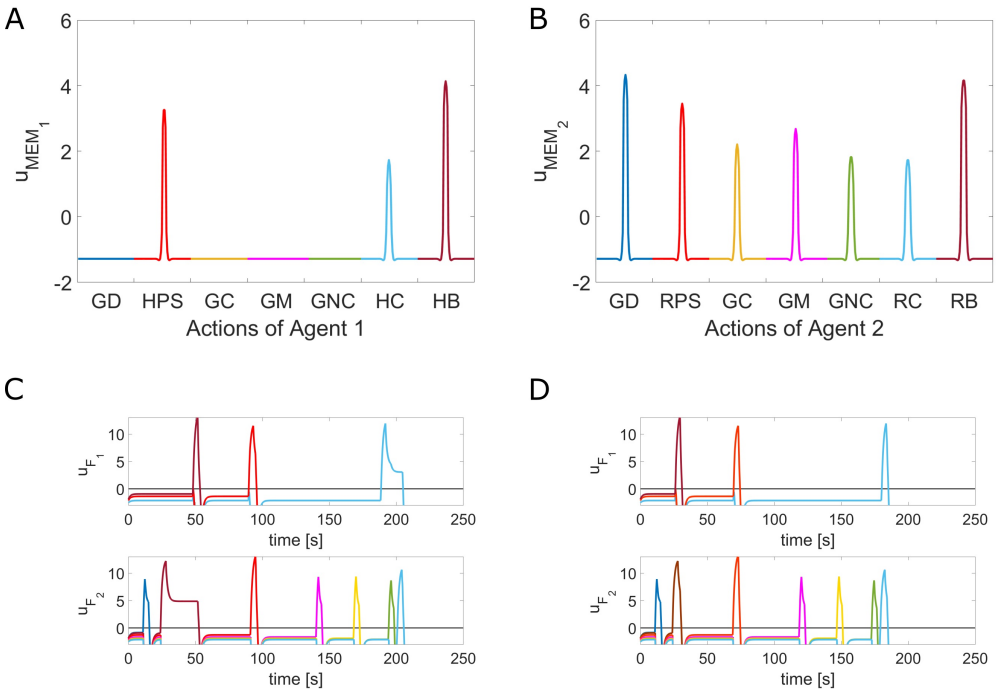

Figure 10: Top: Sequence memory of Agent 1 (**A**) and Agent 2 (**B**). For meaning of labels representing action goals see Figure 1. Bottom: Time course of activation in u_{F_1} and u_{F_2} during the first (**C**) and fourth (**D**) joint execution trial.

An example of an execution trial in which a serial error occurred is demonstrated in Figure 11 where the activation patterns in u_{SIM}, u_{F_2} and u_{ERROR} are overlaid. At time $t = 65s$, due to the internal task simulation, the robot expects that the operator is preparing for receiving the power supply. The activity of the neural population encoding this object-directed action

in u_{ERROR} appears to be suppressed. However, since the operator's hand approaches the cable, a bump representing this erroneous goal has evolved in u_{F_2} which in turn drives the bump evolution in u_{ERROR}. The snapshots at time $t = 70s$ show the stable activation patterns representing the serial order error. The robot is able to explain the correct object manipulation according to the assembly plan to the operator. For the present implementation, we have assumed that this verbal feedback destabilized the bumps in u_{F_2} and u_{ERROR} so that the collaborative work may proceed as planned when the operator finally grasps the power supply at the exchange position.

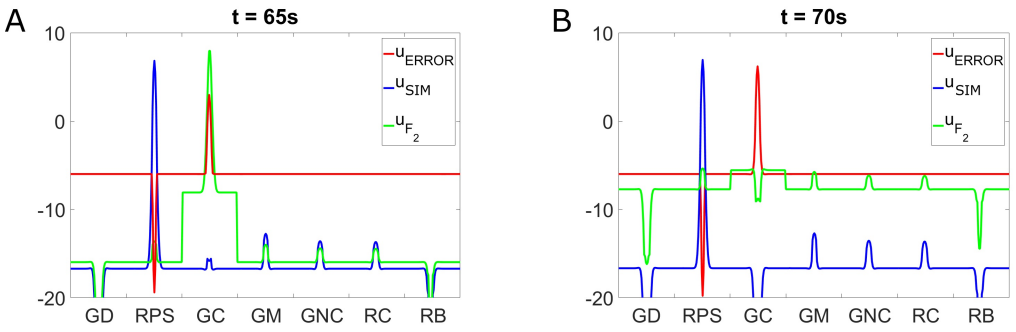

Figure 11: Example of an error trial. Snapshots at times $t = 65s$ (**A**) and $t = 70s$ (**B**) of the activity distribution in the fields u_{ERROR} (red), u_{SIM} (blue) and u_{F_2} (green) are shown. The bumps indicate that the predicted action goal RPS (receive power supply) does not match the observed action goal GS (grasp cable) of Agent 2.

5. Discussion

A close temporal coordination of actions and goals between partners in a collaborative task is crucial for successful and fluent joint action. As new generations of robotic assistants move into human-populated environments, the question of how to endow them with a human-like temporal prediction and adaptation capacity becomes increasingly important (Maniadakis and Trahanias, 2011; Wojtak et al., 2017; Basgol et al., 2021). We have tested in a sequential object transfer task a neurodynamics approach which implements principles of cortical population dynamics as observed in time perception tasks. The neural activity evolves continuously in time towards an attractor state representing the memory of a transfer event and elapsed time. Highly flexible sensorimotor timing can be achieved through manipulations of inputs or initial conditions affecting the speed of the neural trajectory towards

a threshold (Remington et al., 2018). Our study of a temporal cognition capacity is conceptually in line with DNF-based architectures that achieve dynamically adapting behaviors to the spatial aspects of perception and action (Sandamirskaya et al., 2013; Zibner et al., 2015). The hypothesis that timing processes in the brain are closely integrated with other cognitive computations has been exploited by neurodynamics models before. For instance, Maniadakis and colleagues (Maniadakis et al., 2009) used Continuous Time Recurrent Neural Networks (CTRNN) to learn the temporal constraints of a rule-switching task. However, different to the fast, activation-based learning with an existing bump attractor, the shaping of the neural trajectories in a CTRNN is achieved by weight-based neural plasticity which is slow and typically requires extensive training (see also (Remington et al., 2018)). The dynamical systems perspective differs fundamentally from more classical, AI-based implementations with dedicated time processing modules, possibly controlled by a computer clock. In (Huang et al., 2015), a rule-based waiting strategy for handovers with a fixed delay threshold was implemented which takes into account the states of the giver and the receiver. An optimization framework has been used in (Wilcox et al., 2013) to compute a flexible scheduling policy for a robotics assistant working together with different user groups in an industrial assembly task. The objective function is derived from group-specific preference values based on a time assignment to each assembly event. In (Huber et al., 2010), a Kalman filter was applied for online prediction of the time to pass the next part. This was possible due to the statistically observed linear dependency between the complexity and the duration of repetitive assembly steps in the particular task. As shown in Figure 8, ARoS is able without any explicit user model to adjust in a single trial its action timing to sudden temporal changes in the task execution. This is in agreement with experimental evidence showing that humans can anticipate events and deliberately control movement initiation based on single observations (Remington et al., 2018).

The self-stabilizing properties of the field dynamics support the autonomous recall of the stored information in u_{AMEM} with an unspecific start signal given by the operator. For predicting the next user goal and detecting errors, the transition between stable states of the fast internal simulation in u_{SIM} is controlled by bumps encoding sensory feedback.

In future work, we plan to address several limitations of the present HRI study. The objects and the transfer events were identified by simple color cues. The DNF model could benefit from more naturalistic sensory signals in

several ways. First, our focus on an internally generated, cognitive signal for action timing does not exclude of course that also kinematic features of the operator's movements might play a role. Indeed, many studies have revealed experimental evidence that humans use discriminative features in observed movements for anticipating others' motor intentions (Sebanz and Knoblich, 2009). In previous DNF architectures for natural human-robot interactions, we have used the perception of certain kinematic features to endow robots with an action understanding capacity (Bicho et al., 2011; Erlhagen and Bicho, 2006). Using an arm-hand tracker would allow us to address the potential role of kinematic cues for temporal action coordination. However, finding in the continuous stream of bodily motion a salient feature which is predictive for the synchronous arrival of both hands at the exchange location will be in many cases quite difficult and probably requires many execution trials (Cavallo et al., 2016). As a related topic, we also plan to extend the DNF approach to a fluent physical handover phase which requires an interplay of predictive and feedback mechanisms to guarantee a tight temporal coordination of the object release (Ortenzi et al., 2021). Second, additional distinctive object features such as for instance position, size, or orientation should support the identification of a larger set of objects. Higher-dimensional fields in which the bump position represents simultaneously several features or architectures with coupled one-dimensional fields have been used for this purpose in the past (e.g., (Wojtak et al., 2021; Faubel and Schöner, 2008)). In the current robotics experiments, the attractor dynamics supported one-shot learning of the camera input during observation. To increase the robustness of the encoding process in the face of weak and noisy inputs, we have proposed in (Ferreira et al., 2021) an additional "perceptual field" in the DNF architecture in which a gradual buildup of pre-activation during successive task demonstrations represents expectations about likely objects. As detailed in (Ferreira et al., 2021), the existence of a perceptual field also allows us to address the problem of repeated items which naturally occur in many assembly paradigms and other sequential tasks. Typical connectionist models of serial order implementing the idea of an activation gradient assume that an item is represented by a single neuron (Bradski et al., 1994). Since a second instance of the same item will increase the activation level of the already excited neuron, additional mechanisms have to be introduced into the network architecture (e.g., specific "rank-order" neurons (Silver et al., 2012)) to cope with item repetitions. The field concept assumes that large populations of recurrently connected neurons encode a specific object but only

a small portion becomes active in response to a single input presentation. Excitatory-inhibitory connections between the memory field and the perceptual field then ensure that for each object repetition, new subpopulations are automatically recruited to stabilize an activation bump.

In conclusion, the current study contributes to the largely unexplored research area of autonomous robots that are able to use temporal information in the service of more natural and pleasant interactions with human users. The DNF model implements our thinking that taking inspiration from neurocognitive mechanisms supporting human joint action represents a promising way to advance toward an efficient learning and flexible control of temporal coordination in artificial agents. The results of the robotics experiments give support for the hypothesis that the brain achieves this flexibility by modifying the initial condition and inputs of fixed recurrent networks (Remington et al., 2018).

Appendix A. Model equations and parameters

Perceptual Memory

Sequence Memory fields:

$$\begin{aligned}\tau_{u_{MEM}} \frac{\partial u_{MEM_i}(x,t)}{\partial t} =& - u_{MEM_i}(x,t) + \int w_{MEM}(x-y) f(u_{MEM_i}(y,t)) \mathrm{d}y \\ & - h_{MEM_i}(x,t) + S_{MEM_i}(x,t), \quad \text{for } i = 1,2,\end{aligned} \tag{A.1}$$

where

$$\begin{aligned}\frac{\partial h_{MEM_i}(x,t)}{\partial t} =& (1 - f(u_{MEM_i}(x,t)))(-h_{MEM_i}(x,t) + h_{MEM}(t_0)) \\ & + \frac{1}{\tau_{h_{MEM}}} f(u_{MEM_i}(x,t)).\end{aligned} \tag{A.2}$$

Task start field:

$$\begin{aligned}\tau_{u_{START}} \frac{\partial u_{START}(x,t)}{\partial t} =& - u_{START}(x,t) + \int w_{START}(x-y) f(u_{START}(y,t)) \mathrm{d}y \\ & + S_{ON}(x,t) - h_{START}(x,t),\end{aligned} \tag{A.3}$$

where

$$\begin{aligned}\frac{\partial h_{START}(x,t)}{\partial t} =&(1-f(u_{START}(x,t)))(-h_{START}(x,t)+h_{START}(t_0)) \\ &+\frac{1}{\tau_{h_{START}}}f(u_{START}(x,t)).\end{aligned} \tag{A.4}$$

Task duration field:

$$\begin{aligned}\tau_{u_D}\frac{\partial u_D(x,t)}{\partial t} =&-u_D(x,t)+\int w_{MEM}(x-y)f(u_{START}(y,t))\mathrm{d}y \\ &+S_{ON}(x,t)-h_D(x,t),\end{aligned} \tag{A.5}$$

where

$$\frac{\partial h_D(x,t)}{\partial t} = (1-f(u_D(x,t)))(-h_D(x,t)+h_D(t_0))+\frac{1}{\tau_{h_D}}f(u_D(x,t)). \tag{A.6}$$

Motor Memory

Action Onset Memory field:

$$u_{AMEM}(x,t) = u_{MEM_1}(x) - h_{AMEM}(x,t), \tag{A.7}$$

where

$$\frac{\partial h_{AMEM}(x,t)}{\partial t} = \beta(1-f(u_{F_2}(x,t))f(u_{F_1}(x,t)))(f(u_{F_1}(x,t))-f(u_{F_2}(x,t))). \tag{A.8}$$

Past Events field:

$$\begin{aligned}\tau_{u_{PE}}\frac{\partial u_{PE}(x,t)}{\partial t} =&-u_{PE}(x,t)+\int w_{PE}(x-y)f(u_{PE}(y,t))\mathrm{d}y \\ &-h_{PE}+k_{PE}*u_{MEM_2}(x,t) \\ &+S_{PE}f(u_{F_1}(x,t-d_{PE}))f(u_{F_2}(x,t-d_{PE})).\end{aligned} \tag{A.9}$$

Sequence Recall

Action Onset field:

$$\begin{aligned}\tau_{u_{ACT}}\frac{\partial u_{ACT_1}(x,t)}{\partial t} =&-u_{ACT_1}(x,t)+\int w_{ACT}(x-y)f(u_{ACT_1}(y,t))\mathrm{d}y \\ &-h_{ACT}(t)+(u_{MEM_1}(x)-\max(u_D)) \\ &-k_{ACT}\int w_{PE}(x-y)f(u_{PE}(y,t))\mathrm{d}y.\end{aligned} \tag{A.10}$$

The dynamics of the resting level h_{ACT} is controlled by

$$\frac{\mathrm{d}h_{ACT}(t)}{\mathrm{d}t} = \beta_{ACT} \int_{\Omega} f(u_{START}(x))\mathrm{d}x, \qquad h_{ACT}(t_0) \leq 0. \tag{A.11}$$

Simulation field:

$$\begin{aligned} \tau_{u_{SIM}} \frac{\partial u_{SIM}(x,t)}{\partial t} = &- u_{SIM}(x,t) + \int w_{SIM}(x-y) f(u_{SIM}(y,t))\mathrm{d}y \\ &- h_{SIM}(t) + (u_{MEM_2}(x) - \max(u_D)) \\ &- k_{ACT} \int w_{PE}(x-y) f(u_{PE}(y,t))\mathrm{d}y. \end{aligned} \tag{A.12}$$

The dynamics of h_{SIM} is controlled by

$$\frac{\mathrm{d}h_{SIM}(t)}{\mathrm{d}t} = \beta_{SIM} \int_{\Omega} f(u_{START}(x))\mathrm{d}x, \qquad h_{SIM}(t_0) \leq 0. \tag{A.13}$$

Feedback fields:

$$\begin{aligned} \tau_{u_F} \frac{\partial u_{F_i}(x,t)}{\partial t} = &- u_{F_i}(x,t) + \int w_F(x-y) f(u_{F_i}(y,t))\mathrm{d}y + k_F u_{MEM_i} \\ &- h_F + S_{F_i}(t) - \int w_{PE}(x-y) f(u_{PE}(y,t))\mathrm{d}y, \qquad \text{for} \quad i = 1, 2. \end{aligned} \tag{A.14}$$

Error Monitoring field:

$$\begin{aligned} \tau_{u_{ERROR}} \frac{\partial u_{ERROR}(x,t)}{\partial t} = &- u_{ERROR}(x,t) + \int w_E(x-y) f(u_{ERROR}(y,t))\mathrm{d}y \\ &- h_E - k_E u_{SIM}(x,t) f(u_{SIM}(x,t)) \\ &+ u_{F_2}(x,t) f(u_{F_2}(x,t)). \end{aligned} \tag{A.15}$$

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Sequence Memory fields $u_{MEM_{1,2}}$	
$\tau_{u_{MEM}}$	14
$w_{MEM} = w_{osc}$	$A = 1$, $b = 0.8$, $\alpha = 0.6$
$\tau_{h_{MEM_{1,2}}}$	1200 (Model I); 600 (Model II)
$S_{MEM_{1,2}}$	$w_S = 2$, $\sigma_S = 1.5$, $g_S = 0.25$
$h_{MEM}(t_0)$	-4.3
Task Start field u_{START}	
$\tau_{u_{START}}$	6
$w_{START} = w_{lat}$	$w_{exc} = 4$, $\sigma_{exc} = 4.125$, $w_{inhib} = 2.5$
$h_{START}(t_0)$	-4.3
S_{ON}	$w_S = 2$, $\sigma_S = 1.5$, $g_S = 0.25$
Task Duration field u_D	
τ_{u_D}	14
$w_D = w_{osc}$	$A = 1$, $b = 0.8$, $\alpha = 0.6$
τ_{h_D}	1200 (Model I); 600 (Model II)
$h_D(t_0)$	-4.3
S_{ON}	$w_S = 2$, $\sigma_S = 1.5$, $g_S = 0.25$

Table A.2: Parameter values of the field equations used in the *Perceptual memory* part.

Author Contributions

Acknowledgments

Data Availability Statement

Not applicable.

Action Onset Memory field u_{AMEM}	
β	0.00085 (Model I); 0.0017 (Model II)
Past Events field u_{PE}	
$\tau_{u_{PE}}$	14
$w_{PE} = w_{osc}$	$A = 4$, $b = 0.8$, $\alpha = 0.4$
k_{PE}	0.5
S_{PE}	10
d_{PE}	3/dt
h_{PE}	5.6

Table A.3: Parameter values of the field equations used in the *Motor memory* part.

Action Onset field u_{ACT}	
$\tau_{u_{ACT}}$	14
$w_{ACT} = w_{lat}$	$w_{exc} = 4$, $\sigma_{exc} = 2.3$, $w_{inhib} = 2.5$
k_{ACT}	2
β_{ACT}	1/1200 (Model I); 1/600 (Model II)
$h_{ACT}(t_0)$	0
Simulation field u_{SIM}	
$\tau_{u_{ACT}}$	14
$w_{SIM} = w_{lat}$	$w_{exc} = 4$, $\sigma_{exc} = 2.3$, $w_{inhib} = 0$
k_{ACT}	2
β_{SIM}	1/2000
$h_{SIM}(t_0)$	0
Feedback fields $u_{F_{1,2}}$	
τ_{u_F}	14
$w_F = w_{lat}$	$w_{exc} = 4$, $\sigma_{exc} = 2.3$, $w_{inhib} = 2.5$
$S_{F_{1,2}}$	10
k_F	0.5
h_F	3
Error Monitoring field u_{ERROR}	
$\tau_{u_{ERROR}}$	14
$w_E = w_{osc}$	$A = 4$, $b = 0.8$, $\alpha = 0.4$
h_E	6
k_E	2

Table A.4: Parameter values of the field equations used in the *Sequence recall* part.

References

Almeida, R., Ledberg, A., 2010. A biologically plausible model of time-scale invariant interval timing. Journal of Computational Neuroscience 28, 155–175.

Amari, S., 1977. Dynamics of pattern formation in lateral-inhibition type neural fields. Biological Cybernetics 27, 77–87.

Basgol, H., Ayhan, I., Ugur, E., 2021. Time perception: A review on psychological, computational and robotic models. IEEE Transactions on Cognitive and Developmental Systems .

Bastian, A., Riehle, A., Erlhagen, W., Schöner, G., 1998. Prior information preshapes the population representation of movement direction in motor cortex. NeuroReport 9, 315–319.

Bicho, E., Erlhagen, W., Louro, L., e Silva, E.C., 2011. Neuro-cognitive mechanisms of decision making in joint action: A human–robot interaction study. Human Movement Science 30, 846–868.

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, 5.

Bogacz, R., Wagenmakers, E.J., Forstmann, B.U., Nieuwenhuis, S., 2010. The neural basis of the speed–accuracy tradeoff. Trends in Neurosciences 33, 10–16.

Bohren, J., Rusu, R.B., Jones, E.G., Marder-Eppstein, E., Pantofaru, C., Wise, M., Mösenlechner, L., Meeussen, W., Holzer, S., 2011. Towards autonomous robotic butlers: Lessons learned with the PR2, in: 2011 IEEE International Conference on Robotics and Automation, IEEE. pp. 5568–5575.

Bradski, G., Carpenter, G.A., Grossberg, S., 1994. STORE working memory networks for storage and recall of arbitrary temporal sequences. Biological Cybernetics 71, 469–480.

Brody, C.D., Romo, R., Kepecs, A., 2003. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. Current Opinion in Neurobiology 13, 204–211.

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., Becchio, C., 2016. Decoding intentions from movement kinematics. Scientific Reports 6, 1–8.

Cini, F., Ortenzi, V., Corke, P., Controzzi, M., 2019. On the choice of grasp type and location when handing over an object. Science Robotics 4, eaau9757.

Coombes, S., Owen, M., 2007. Exotic dynamics in a firing rate model of neural tissue, in: Fluids and Waves: Recent Trends in Applied Analysis: Research Conference, May 11-13, 2006, the Universtiy of Memphis, Memphis, TN, American Mathematical Soc.. p. 123.

Curtis, C.E., Lee, D., 2010. Beyond working memory: the role of persistent activity in decision making. Trends in Cognitive Sciences 14, 216–222.

De Lange, F.P., Heilbron, M., Kok, P., 2018. How do expectations shape perception? Trends in Cognitive Sciences 22, 764–779.

De Santis, A., Siciliano, B., De Luca, A., Bicchi, A., 2008. An atlas of physical human–robot interaction. Mechanism and Machine Theory 43, 253–270.

Duran, B., Sandamirskaya, Y., 2017. Learning temporal intervals in neural dynamics. IEEE Transactions on Cognitive and Developmental Systems 10, 359–372.

Erlhagen, W., Bicho, E., 2006. The dynamic neural field approach to cognitive robotics. Journal of Neural Engineering 3, 36–54. doi:10.1088/1741-2560/3/3/R02.

Erlhagen, W., Schöner, G., 2002. Dynamic field theory of movement preparation. Psychological Review 109, 545.

Faubel, C., Schöner, G., 2008. Learning to recognize objects on the fly: a neurally based dynamic field approach. Neural Networks 21, 562–576.

Ferreira, F., Erlhagen, W., Bicho, E., 2016. Multi-bump solutions in a neural field model with external inputs. Physica D: Nonlinear Phenomena 326, 32–51.

Ferreira, F., Wojtak, W., Sousa, E., Louro, L., Bicho, E., Erlhagen, W., 2021. Rapid learning of complex sequences with time constraints: A dynamic neural field model. IEEE Transactions on Cognitive and Developmental Systems 13, 853–864.

Glasauer, S., Huber, M., Basili, P., Knoll, A., Brandt, T., 2010. Interacting in time and space: Investigating human-human and human-robot joint action, in: 19th International Symposium in Robot and Human Interactive Communication, IEEE. pp. 252–257.

Grigore, E.C., Eder, K., Pipe, A.G., Melhuish, C., Leonards, U., 2013. Joint action understanding improves robot-to-human object handover, in: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE. pp. 4622–4629.

Gulletta, G., Silva, E.C.e., Erlhagen, W., Meulenbroek, R., Costa, M.F.P., Bicho, E., 2021. A human-like upper-limb motion planner: Generating naturalistic movements for humanoid robots. International Journal of Advanced Robotic Systems 18, 1–31.

Hass, J., Durstewitz, D., 2016. Time at the center, or time at the side? Assessing current models of time perception. Current Opinion in Behavioral Sciences 8, 238–244.

Hoffman, G., 2019. Evaluating fluency in human–robot collaboration. IEEE Transactions on Human-Machine Systems 49, 209–218.

Huang, C.M., Cakmak, M., Mutlu, B., 2015. Adaptive coordination strategies for human-robot handovers., in: Robotics: Science and Systems, pp. 1–10.

Huber, M., Knoll, A., Brandt, T., Glasauer, S., 2010. When to assist?-Modelling human behaviour for hybrid assembly systems, in: ISR 2010 (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics), VDE. pp. 1–6.

Koene, A., Endo, S., Remazeilles, A., Prada, M., Wing, A.M., 2014. Experimental testing of the coglaboration prototype system for fluent human-robot object handover interactions, in: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, IEEE. pp. 249–254.

Laing, C.R., Troy, W.C., Gutkin, B., Ermentrout, G.B., 2002. Multiple bumps in a neuronal model of working memory. SIAM Journal on Applied Mathematics 63, 62–97.

Lima, P., Erlhagen, W., Kulikova, M., Kulikov, G.Y., 2022. Numerical solution of the stochastic neural field equation with applications to working memory. Physica A: Statistical Mechanics and its Applications 596, 127166.

Maimon, G., Assad, J.A., 2006. A cognitive signal for the proactive timing of action in macaque LIP. Nature Neuroscience 9, 948–955.

Maniadakis, M., Trahanias, P., 2011. Temporal cognition: a key ingredient of intelligent systems. Frontiers in Neurorobotics 5, 2.

Maniadakis, M., Trahanias, P., Tani, J., 2009. Explorations on artificial time perception. Neural Networks 22, 509–517.

Medina, J.R., Duvallet, F., Karnam, M., Billard, A., 2016. A human-inspired controller for fluid human-robot handovers, in: 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), IEEE. pp. 324–331.

Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience 24, 167–202.

Moon, A., Troniak, D.M., Gleeson, B., Pan, M.K., Zheng, M., Blumer, B.A., MacLean, K., Croft, E.A., 2014. Meet me where I'm gazing: how shared attention gaze affects human-robot handover timing, in: Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, pp. 334–341.

Mörtl, A., Lorenz, T., Hirche, S., 2014. Rhythm patterns interaction-synchronization behavior for human-robot joint action. PloS One 9, e95195.

O'Reilly, J.X., McCarthy, K.J., Capizzi, M., Nobre, A.C., 2008. Acquisition of the temporal and ordinal structure of movement sequences in incidental learning. Journal of Neurophysiology 99, 2731–2735.

Ortenzi, V., Cosgun, A., Pardi, T., Chan, W.P., Croft, E., Kulić, D., 2021. Object handovers: a review for robotics. IEEE Transactions on Robotics 37, 1855–1873.

Remington, E.D., Egger, S.W., Narain, D., Wang, J., Jazayeri, M., 2018. A dynamical systems perspective on flexible motor timing. Trends in Cognitive Sciences 22, 938–952.

Sandamirskaya, Y., Zibner, S.K., Schneegans, S., Schöner, G., 2013. Using dynamic field theory to extend the embodiment stance toward higher cognition. New Ideas in Psychology 31, 322–339.

Schöner, G., 2016. Dynamic thinking: A primer on dynamic field theory. Oxford University Press.

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

Sebanz, N., Knoblich, G., 2009. Prediction in joint action: What, when, and where. Topics in Cognitive Science 1, 353–367.

Silver, M.R., Grossberg, S., Bullock, D., Histed, M.H., Miller, E.K., 2012. A neural model of sequential movement planning and control of eye movements: Item-order-rank working memory and saccade selection by the supplementary eye fields. Neural Networks 26, 29–58.

Strabala, K., Lee, M.K., Dragan, A., Forlizzi, J., Srinivasa, S.S., Cakmak, M., Micelli, V., 2013. Toward seamless human-robot handovers. Journal of Human-Robot Interaction 2, 112–132.

Svoboda, K., Li, N., 2018. Neural mechanisms of movement planning: motor cortex and beyond. Current Opinion in Neurobiology 49, 33–41.

Vesper, C., Butterfill, S., Knoblich, G., Sebanz, N., 2010. A minimal architecture for joint action. Neural Networks 23, 998–1003.

Wilcox, R., Nikolaidis, S., Shah, J., 2013. Optimization of temporal dynamics for adaptive human-robot interaction in assembly manufacturing. Robotics 8, 10–15.

Wojtak, W., Ferreira, F., Louro, L., Bicho, E., Erlhagen, W., 2017. Towards temporal cognition for robots: a neurodynamics approach, in: 2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), IEEE. pp. 407–412.

Wojtak, W., Ferreira, F., Vicente, P., Louro, L., Bicho, E., Erlhagen, W., 2021. A neural integrator model for planning and value-based decision making of a robotics assistant. Neural Computing and Applications 33, 3737–3756.

Zibner, S.K., Tekülve, J., Schöner, G., 2015. The neural dynamics of goal-directed arm movements: a developmental perspective, in: 2015 Joint IEEE international conference on development and learning and epigenetic robotics (ICDL-EpiRob), IEEE. pp. 154–161.