Modeling mental rotation in cognitive robots
Kristsana Seepanomwan, Daniele Caligiore, Gianluca Baldassarre and Angelo Cangelosi
Adaptive Behavior published online 11 July 2013
DOI: 10.1177/1059712313488782

The online version of this article can be found at:
http://adb.sagepub.com/content/early/2013/07/11/1059712313488782

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
ISAB
International Society of Adaptive Behavior

Additional services and information for Adaptive Behavior can be found at:

Email Alerts: http://adb.sagepub.com/cgi/alerts
Subscriptions: http://adb.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav

>> OnlineFirst Version of Record - Jul 11, 2013
What is This?
Modelling mental rotation in cognitive robots

Kristsana Seepanomwan¹, Daniele Caligiore², Gianluca Baldassarre² and Angelo Cangelosi¹

Abstract
Mental rotation concerns the cognitive processes that allow an agent mentally to rotate the image of an object in order to solve a given task, for example to say if two objects with different orientations are the same or different. Here we present a system-level bio-constrained model, developed within a neurorobotics framework, that provides an embodied account of mental rotation processes relying on neural mechanisms involving motor affordance encoding, motor simulation and the anticipation of the sensory consequences of actions (both visual and proprioceptive). This model and methodology are in agreement with the most recent theoretical and empirical research on mental rotation. The model was validated through experiments with a simulated humanoid robot (iCub) engaged in solving a classical mental rotation test. The results of the test show that the robot is able to solve the task and, in agreement with data from psychology experiments, exhibits response times linearly dependent on the angular disparity between the objects. This model represents a novel detailed operational account of the embodied brain mechanisms that may underlie mental rotation.

Keywords
Mental rotation, computational robotic model, neurorobotics, neural mechanisms, affordances and forward models, parietal/premotor/prefrontal cortex.

1 Introduction
Since it was first described by Shepard and Metzler (1971), mental rotation has attracted an enormous research interest in the field of cognitive psychology. This is in part due to the attempts to understand why object comparison using imagery seems to obey the same physical principles as overt rotation, considering that humans are capable of using imagery that is not limited by the laws of physics (Kosslyn, 1994). In a typical mental rotation task, human participants are asked to make a decision on whether two objects presented with different rotational orientations are the same or a mirror version of each other. The results show that the response times (RTs), as well as the errors, of the participants’ answers are highly dependent on the angular disparity between the two stimuli (Shepard & Metzler, 1971; Wexler, Kosslyn, & Berthoz, 1998). In particular, participants exhibit RTs that linearly increase with the disparity angle between the orientations of the objects. The number of errors also increases with the increment of disparity. The most accredited explanation of these results is that participants might rotate a ‘mental’ image of one object until its orientation matches the one of the other object (Kosslyn, 1994). Once mentally rotated, the participants can ascertain if the two objects are the same or not.

Early attempts to explain brain mechanisms underlying mental rotation processes relied upon a visuo-spatial perception hypothesis (Corballis & McLaren, 1982; Shepard & Metzler, 1971). According to this view, mental rotation is performed on the basis of processes mainly involving the internal manipulation of the visual and spatial features of objects. This view makes the prediction that these processes mainly implicate brain areas underlying visual and spatial perception. Contrary to this, recent behavioural and neuroscientific evidence indicates an important role of motor processes, beside the perceptual ones, in mental rotation. In this respect, several behavioural works show interferences between action planning/execution, and mental

¹Plymouth University, Plymouth, UK
²Laboratory of Computational Embodied Neuroscience, Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche (LOCEN-ISTC-CNR), Rome, Italy

Corresponding author:
Kristsana Seepanomwan, Plymouth University, Drake Circus, Plymouth PL4 8AA, UK.
Email: kristsana.seepanomwan@plymouth.ac.uk
rotation processes (Wexler et al., 1998; Wohlschläger, 2001; Wohlschläger & Wohlschläger, 1998). In a typical experiment, participants are asked to perform a classical mental rotation task (Shepard & Metzler, 1971) while performing a manual rotation on a custom joystick in both congruent and incongruent conditions with respect to the direction of rotation of the mental image. The results show that RTs (and error rates) are faster (lower) when the direction of the two rotations (manual and mental) is congruent, whereas they are slower (higher) when they are inconsistent (Wexler et al., 1998; Wohlschläger, 2001). This supports the idea that motor processes play a key role in mental rotation, as otherwise it would be difficult to explain why the production of overt motor actions interferes with mental rotation only when the two are incongruent.

Single cell recordings in the motor cortices of monkeys also supply direct neural evidence for the involvement of motor processes in mental rotation (Georgopoulos, Lurito, Petrides, Schwartz, & Massey, 1989). In humans, a number of neuroscientific studies using different research techniques, such as transcranial magnetic stimulation (TMS), event-related potentials (ERPs) and functional magnetic resonance imaging (fMRI), show an involvement of lateral and medial premotor areas (lateral premotor cortex/precentral gyrus and supplementary motor area) during mental rotation (Lamm, Windischberger, Moser, & Bauer, 2007; Richter et al., 2000). The fMRI study of Richter and colleagues (Richter et al., 2000), for example, shows a significant correlation between the hemodynamic response in lateral premotor areas with the RT of participants involved in the classical Shepard and Metzler mental rotation task (Shepard & Metzler, 1971). This result suggests that mental rotation is an imagined (covert) object rotation action rather than an image transformation relying exclusively upon visuo-spatial processing. This claim has been further confirmed by other studies (cf. Lamm et al., 2007; Lamm, Fischmeister, & Bauer, 2005; Wohlschläger, 2001).

Importantly, despite these consistent results about the involvement of motor processes during mental rotation, we still lack a comprehensive hypothesis of the specific brain mechanisms involving motor simulation that might underlie mental rotation processes. One proposal that might help to explain the role of premotor areas during mental rotation pivots on the concept of affordance (Gibson, 1979) and its behavioural manifestations (Tucker & Ellis, 2001), brain correlates (Rizzolatti & Craighero, 2004) and models (Caligiore et al., 2010; Fagg & Arbib, 1998). According to this perspective, affordances are the possible actions that objects and the environment offer to a certain agent. In particular, the visual presentation of objects triggers the activation of internal representations of the affordances needed for the online guidance of actions over them within the parietal–premotor circuits (Grafton, Fadiga, Arbib, & Rizzolatti, 1997; Grezès & Decety, 2001). In this respect, the activation of affordance representations might be involved in the mental rotation processes, as in brain they play a key role in the first stage of motor preparation.

Another hypothesis on how motor areas might participate in mental rotation comes from theories (Grush, 2004), neuroscientific evidence (Miall, 2003) and computational architectures (Baldassarre, 2002, 2003; Wolpert & Kawato, 1998) on motor control based on forward models. This perspective suggests that preparatory/planning covert motor processes play a key role in the mental simulation and understanding of the environment, and involve the same brain motor areas involved in overt action execution. This view would suggest that mental rotation involves the same motor areas and mechanisms used in the physical execution of active rotations of objects (e.g., manual rotations), and the imagined anticipation of their sensory consequences.

So, both views would give important indications on the possible involvement of motor areas in mental rotation phenomena. Wexler et al. (1998, p. 77) stated the hypotheses that ‘transformations of mental image are at least in part guided by motor processes.’ This view also supports the existence of a relation between affordance learning (motor processes) and forward models (mental imagery). In addition, the dual task paradigm (Kosslyn, 1994) is the best example that also supports the view of shared location between motor processes and mental rotations in motor cortex. Moreover, the generation of affordances and motor commands can be generated, from the initial configuration of body and environment, and selected, on the basis of goal-related information (Thill, Caligiore, Borghi, Ziemke, & Baldassarre, 2013). However, these views are partial as mental rotation is a complex process requiring the coordinated operation of several distinct elemental cognitive processes. These processes include (Lamm et al., 2007): (a) stimulus encoding and mental image generation, (b) planning and execution of the mental rotation, (c) comparison (matching) of the rotated stimulus with the target stimulus, and finally (d) execution of the same/different response.

In this article, we propose a system-level computational model suggesting a specific operational hypothesis on how the information processes taking place in brain sensorimotor areas might interplay to perform mental rotation. This hypothesis first draws ideas from the affordance and forward model view introduced above, and integrates and specifies them to make them applicable to the explanation of mental rotation. Second, it introduces some additional elements to allow the implementation of not only the processes (a) and (b) indicated above (mental rotation proper), but also (c) and (d) (control and exploitation of mental rotation processes).
To this purpose, the model builds on some ideas from the computational model ‘TRoPICALS’ (Caligiore et al., 2010, 2013) developed to study affordance compatibility effects (Tucker & Ellis, 2001). TRoPICALS is a good starting point to design a model on mental rotation as it reproduces some key functions of the parietal–premotor circuits, crucial for stimulus encoding and extraction of object affordances (process a). TRoPICALS also includes important features of the prefrontal–premotor circuit pivotal for managing other key aspects of mental rotation (processes c and d). However, it cannot perform mental simulations as it lacks the needed feedback circuits. In this context, to address the core mental rotation process (process b) the model proposed here enhances the functions of TRoPICALS by developing some new key features. First, it is endowed with premotor–parietal feedback loops that allow it to implement mental simulation and sensory prediction based on forward models. Second, it is endowed with enhanced parietal functions for encoding somatosensory information, important to elaborate anticipated proprioceptive signals. Third, it is endowed with an improved visual and motor system allowing it to scale up to more realistic 3D environments and robotic setups.

The rest of the paper is organized as follows. Section 2 discusses the main features of the model, the learning algorithms used to train it, and the robotic setup used to validate it. Section 3 presents and discusses the results. Finally, Section 4 drives the conclusions.

2 Methods

2.1 The simulated mental rotation experiment: task, participant and stimuli

A mental rotation task used in this work follows the typical mental rotation tasks that have been used in the field of experimental psychology, for example by Shepard and Metzler (1971), and Chu and Kita (2008, 2011). The goal is to let a simulated participant make a judgement on whether a pair of stimuli is the same or a mirror version of each other. The stimuli are stylized geometrical shapes. At each trial, the stimuli can change in terms of object type and orientation. After the simulated participant produces an answer, a new trial will be started by changing the current pair of stimuli and/or their rotation.

Figure 1 shows the simulated humanoid robot iCub (Tikhanoff et al., 2008) used as a participant to model the targeted psychological experiments. The simulation replicates the same body and control scheme of the real iCub robot (Sandini, Metta, & Vernon, 2007), which is an open source robotic platform built for studying cognitive development in humans. iCub looks like a 3–5-year-old child, in great part designed on the basis of human body structure and movements. Thanks to these features, the iCub platform is widely used as a cognitive robotics tool in many laboratories (Cangelosi & Schlesinger, in press).

The iCub simulator provides visual perception via simulated cameras and can perform actions corresponding to specific motor commands. Each arm of the iCub has 16 joints. Here we use the joint number 5 of the right arm affecting the robot’s wrist angle. If the robot holds an object with the right hand, rotating the wrist will change the orientation of the object on a plane. During the mental rotation task, the model has to compare two visual stimuli having different orientation as in the target experiments.

Within the perspective of embodied cognition, the robot platform used in this work furnishes to the model perception and action capabilities through simulated cameras and motor outputs. Here a small subset of the sensorimotor possibilities of the iCub simulator was used to demonstrate the possibility of performing the mental rotation task within a robotic embodied setup. However, in future work we will consider the implementation of more complex mental rotation tasks and the role of gestures. In these cases, the rich perception and multiple degrees of freedom of the iCub platform will allow the investigation of sophisticated cognitive skills related to object recognition, management of mental images (creation/rotation) and problem solving.

The stimuli are coloured in red to ease their detection using the iCub camera. A red colour filter is first used to this purpose. An edge detection method is then applied to mimic an early visual processing stage. The edges of the object are extracted with the Canny edge detection technique (Canny, 1986) based on the OpenCV library. The output from the edge detection process is used as input to the neural network system. The right object is considered to be the target stimulus, whereas the left object represents the current ‘comparison’ stimulus, which has to be mentally rotated (the left stimulus image is fed to PP, whereas the right stimulus image to PFC_1 (Figure 2); Section 2.2.1 will present a...
The eye position of the iCub is fixed, with the left object centred on the fovea throughout the experiment. Regarding the motor response, the iCub’s wrist angle can rotate in the range of $[−90°; 90°]$. Counter-clockwise rotations and orientations are indicated by positive values, while clockwise ones are indicated by negative values.

During the experiment, pairs of target and comparison object images having different orientations are used. The objects are displayed in the space in front of the iCub (Figure 1). For the training, the rotation of the comparison object is varied by $30°$, so that each stimulus can assume seven orientations ($−90°$, $−60°$, $−30°$, $0°$, $30°$, $60°$, $90°$). During the process of affordance training, only one comparison stimulus is shown in the left position, with the experimenter varying the orientation of the object and assigning a corresponding position for the robot’s wrist angle. In the testing session, two stimuli are displayed, the comparison stimulus at the left and the target stimulus at the right positions.

After training, the generalization ability of the model is tested using 196 pairs of stimuli supplied in sequence. The experiment has been repeated 10 times to test the consistency of the model. Each time the pair of stimuli is changed, the model internally rotates the left stimulus to match it with the right one and produce an answer. Three types of information are recorded during the experiment: the RTs, which are the result of a neural dynamical competition (see Section 2.2.2 and cf. Caligiore et al., 2010; Erlhagen & Schöner, 2002); the answer for the current mental rotation task (Section 2.2.2); the successful degree of rotation (Section 2.2.2). When the number of rotation cycles reaches 10, this indicates that the model cannot correctly perform the mental image rotation of the left stimulus and so it is forced to do the matching process by using the last rotated image.

### 2.2 Neural architecture, simulated mental rotation, learning phase

#### 2.2.1 Neural architecture

The neural network model (Figure 2) proposed in this article suggests an operational hypothesis about the interplay of the visual and motor neural processes during mental rotation. To this
purpose, the model extends some features of the TRoPICALS model. TRoPICALS (Caligiore et al., 2010, 2013) is a computational model of affordance control designed to account for action-language and stimulus–response compatibility effects studied experimentally in cognitive psychology (Tucker & Ellis, 2001, 2004). It does this based on an architecture that considers prefrontal cortex as a key source of the top-down control of the areas that participate to the selection of affordances and execution of actions.

The account of compatibility effects given by TRoPICALS is based on four general brain organization principles incorporated in its architecture (Caligiore et al., 2010): (a) the two-route organization of the sensorimotor brain into the ventral and a dorsal neural pathways; (b) the guidance of action selection based on prefrontal cortex ‘instructions’; (c) the selection of actions within premotor cortex based on a neural competition between different affordances with bias from prefrontal cortex; (d) the capability of language to trigger internal simulations of the referents of words (Barsalou, Santos, Simmons, & Wilson, 2008). The acronym ‘TRoPICALS’ summarizes these principles: Two Route, Prefrontal Instruction, Competition of Affordances, Language Simulation. The model reproduces compatibility effects on the basis of the agreement or disagreement (compatibility or incompatibility) of the top-down bias from prefrontal cortex with the available affordances of objects as this produces respectively fast or slow reaction times. TRoPICALS provides a broad framework to account for several types of affordance related compatibility effects involving grasping, reaching and language, and is capable of generating novel testable predictions, including some predictions on the possible outcomes of compatibility experiments with Parkinson patients (Caligiore et al., 2013; the latter predictions are relevant as Parkinson patients have damaged excitatory and inhibitory neural circuits linking prefrontal cortex to premotor cortex via supplementary motor cortex).

The architecture of the model presented here is shown in Figure 2. It consists of four parts corresponding to main areas of the brain involved in mental rotation tasks (Lamm et al., 2007; Richter et al., 2000): the parietal cortex (PC), the premotor cortex (PMC), the prefrontal cortex (PFC) and the primary motor cortex (M1). The dorsal pathway through the circuit PC–PMC is responsible for the ‘how’ sub-task in this case, i.e., for the pre-activation and selection of affordances of the seen objects. The ventral pathway via PC–PFC is instead the circuit that recognizes objects (‘what’ sub-task). The matching and answer triggering processes are the result of the integration of the maps PC, PMC and PFC. The M1 is responsible for overt control of the robot’s wrist movement. Repeating processes within PC and PMC drive mental image rotation, which is supported by the interaction between affordance processing and forward model actions. The proprioceptive input from the robot’s wrist posture (PC) plays a key role in the forward model used during mental rotation.

Each cortical area is formed by two neural maps encoding information using population code methods (Pouget, Dayan, & Zemel, 2003). Population code methods claim that information (e.g., on stimuli and actions) is encoded in the brain on the basis of the activation of populations of neurons having a broad response field and topologically organized in neural maps. In particular, each neuron of a map responds maximally to a certain value of the variables to encode, and then progressively less intensely to less similar values (based on a Gaussian-like function).

PC is formed by two distinct areas: the posterior–parietal cortex (PP) and the somatosensory cortex (SS). The neurons of the PP map (32×32 neurons) encode the shape and the orientation of the object that has to be mentally rotated (Rizzolatti & Craighero, 2004).

The neurons of the SS map (31×100 neurons) elaborate the proprioceptive signal related to the robot wrist orientation (Caligiore et al., 2010). The PMC region is formed by two neural maps PMC_1 (31×100 neurons) and PMC_2 (10×20 neurons). The two maps encode motor programs related to different arm parts (Caligiore et al., 2008a; Rizzolatti & Craighero, 2004): PMC_1 neurons encode the wrist posture of the robot corresponding to the object orientation encoded in PP; PMC_2 neurons encode the hand posture that the robot produces to accomplish the mental rotation results (i.e., to indicate if two objects are same or different). The PFC (Fuster, 2001) also consists of two maps implementing a working memory encoding the target stimulus (PFC_1, 32×32 neurons) and performing the matching process (PFC_2, 64×64 neurons; cf. Baldassarre, 2002, 2003, for an embodied neural-network model of planning based on visual imagery and using a goal-matching mechanism).

The visual input for the model is the captured image from one ‘eye’ (camera) of the simulated iCub robot. The edge information for the object on the left is passed to the PP, while the one for the target object on the right is sent to the PFC_1. The target object is used as a reference for the rotational process. The robot has mentally to rotate the object encoded by PP and check if it is the same or it is different with respect to the target object encoded in PFC_1. For each image, PP pre-activates all possible wrist postures in PMC_1. This pre-activation is equal to 0.2 and represents the possible actions afforded by the current image in PP. At the same time, PFC_1 supplies a bias signal to PMC_1 to lead to the full activation, equal to 1.0, of one desired final wrist posture among the ones afforded by PP. This posture corresponds to the desired final orientation of the object that the robot has (mentally) to accomplish to overlap the image within PP with the
target image within PFC_1. In parallel with these processes, the PFC_2 performs the matching process. PFC_2 is formed by a Kohonen self-organizing map (SOM; Kohonen, 1997), which takes inputs from PP and PFC_1 and represents each possible combination of their activation as a whole cluster. This represents the current situation used by PFC to decide what to do (cf. Caligiore et al., 2010). The winning clusters of PFC_2 cause the PMC_2 activation, in turn encoding the answer of the system.

M1 consists of two areas M1_1 and M1_2. M1_1 is a SOM map (64×64 neurons) responsible for encoding a combination of the current posture from SS and the desired posture from PMC_1. The neural activation of M1_1 feedbacks to SS as a reference copy of the motor program during the mental rotation process (see below). M1_1 also triggers a wrist rotation movement through M1_2. M1_2 is a neural array formed by 10×30 neurons grouped in three separated clusters (N1, N2, N3). The activation of N1 causes a 30° clockwise rotation of the wrist; the activation of N3 causes a 30° counter-clockwise rotation of the wrist; the activation of N2 does not lead to any rotation of the wrist.

2.2.2 The simulated mental rotation process. This section briefly summarizes how the model reproduces the mental rotation processes. The following points refer to the model functioning after the learning processes, illustrated in Section 2.2.3, have terminated. Affordance-based action pre-activation (C1). The left object image encoded by PP neurons pre-activates all the possible write postures within PMC_1 at the same time. Since one object could assume seven different orientations, we have seven different clusters of neurons pre-activated within PMC_1. This affordance-based pre-activation of possible actions mimics the preparatory processes for actions present when people see an object.

Action selection (C6): PFC_1 supplies a bias signal to PMC_1 to lead the full activation (with a level of neural activation of 1.0) of one affordance/action among the elicited ones so transforming it into the representation of a specific desired final wrist posture. This cluster represents the desired posture that the robot has (mentally) to reach mentally to rotate and overlap the image within PP with the target image within PFC_1.

Mental rotation by the inverse model (C7, C8) and the forward model (C10, C2): The desired wrist posture encoded by PMC_1 and the current wrist posture encoded by the SS are combined within M1_1 (C7, C8). Together with C9 connections, this forms an inverse model (inverse models map the current state and the desired state into the action needed to move from the former to the latter one). M1_1 and SS form a forward model (forward models map the current state and planned action into the future state). In particular, the winning cluster within M1_1 evokes a cluster within SS corresponding to the next anticipated wrist posture (C10). In turn, this cluster within SS activates the new rotated image within PP (C2), so causing a mental rotation step. In particular, the connection C2 from SS to PP underlies the process of mental image generation based on the anticipated proprioception. After a specific proprioceptive cluster in SS has been formed, this causes the corresponding image back to PP so that a progressive sequence of clusters in SS will cause a corresponding progressive rotation of the image in PP.

In line with empirical evidence (Chu & Kita, 2008, 2011) the current proprioceptive signal that affects the mental rotation processes based on the activation of SS depends on both the signal from M1_1 (C10) related to the planned action. This process might be disturbed by the current actual posture that is still (Figure 2). In this respect, we assume that attention mechanisms not explicitly simulated here (Logan, 1996; Roelfsema, Lamme, & Spekreijse, 1998) might drive the system to be more focused on the mental rotation task rather than on the wrist condition. This assumption is supported by recent evidence showing the presence of reciprocal interference between mechanisms of mental rotation and the deployment of visual–spatial attention (Pannebakker et al., 2011). The effect of the attention focus assumed here is simulated by setting (within SS) a weaker signal from current proprioception than from the forward model.

The mental rotation in this work is achieved through a training strategy that considers the angular difference between the two stimuli of the task. When the orientation of the left object is greater than that of the right target object, the model generates a mental image of the left object rotated one step (30°) clockwise. In contrast, the model performs a one-step counter-clockwise rotation when the left object’s orientation is smaller than the right one. The RTs expressed by the model, proportional to the discrepancy of orientation between the target and the rotated object, are so strongly dependent on the specific mechanisms assumed here to perform mental rotation. These mechanisms are consistent with what might happen in human working memory of subjects engaged in mental rotation tasks. The model always uses the last image in PP to perform the matching process. The maximum number of rotation cycles is set to 10, more than needed by a maximum rotation, as in some cases the model cannot rotate the image of one position at the first cycle and so requires extra rotations.

2.2.3 Learning process. Connections between maps are trained using Hebbian learning and SOM competitive learning (summarized in Table 1), which are widely accepted as biologically plausible learning mechanisms involving cortical areas (Doya, 2000). The specific Hebbian learning method used in this model is the Oja
Table 1. The parameters used in the network.

<table>
<thead>
<tr>
<th>Connection</th>
<th>Type</th>
<th>Number of patterns</th>
<th>Training cycles</th>
<th>Type of output</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Hebb</td>
<td>14</td>
<td>84</td>
<td>Cluster of activity</td>
</tr>
<tr>
<td>C2</td>
<td>Hebb</td>
<td>14</td>
<td>84</td>
<td>Image</td>
</tr>
<tr>
<td>C3 and C4</td>
<td>Kohonen</td>
<td>98</td>
<td>10,000</td>
<td>Cluster of activity</td>
</tr>
<tr>
<td>C5</td>
<td>Hebb</td>
<td>196</td>
<td>1176</td>
<td>Cluster of activity</td>
</tr>
<tr>
<td>C6</td>
<td>Hebb</td>
<td>14</td>
<td>84</td>
<td>Cluster of activity</td>
</tr>
<tr>
<td>C7 and C8</td>
<td>Kohonen</td>
<td>196</td>
<td>10,000</td>
<td>Cluster of activity</td>
</tr>
<tr>
<td>C9</td>
<td>Hebb</td>
<td>98</td>
<td>1960</td>
<td>Cluster of activity</td>
</tr>
<tr>
<td>C10</td>
<td>Hebb</td>
<td>98</td>
<td>1960</td>
<td>Cluster of activity</td>
</tr>
</tbody>
</table>

rule (Oja, 1982; this is a Hebbian rule that solves the problem of the basic Hebb rule causing weights to grow without bound). The equation used to implement the Hebbian learning process is as follows:

$$\Delta w_{ij} = \eta a_i (a_j - w_{ij}); w(t)_{ij} = w(t-1)_{ij} + \Delta w_{ij}$$  \hspace{1cm} (1)

where \(\Delta w_{ij}\) denotes the weight’s change from neuron \(i\) to neuron \(j\), \(a_i\) and \(a_j\) denote activation potential of neuron \(i\) and \(j\) respectively, \(\eta\) denotes the learning rate, which is set to 0.15, and \(w(t)_{ij}\) is a weight value at a particular time step. The SOM learning rule has been implemented using the following equation:

$$w(t)_i = w(t-1)_i + \Theta(t-1)_i \eta(t-1)_i (v(t-1)_i - w(t-1)_i)$$  \hspace{1cm} (2)

where \(w(t)_i\) denotes current weight value of neuron \(i\) at time \(t\), \(w(t-1)_i\) denotes an old weight value of the neuron \(i\), \(\Theta\) denotes the amount of influence on distance between neuron \(i\) and the best matching neuron in a map, \(\eta\) denotes the learning rate, which is set to 0.15. Note that, \(\Theta\) and \(\eta\) decrease over time. Table 1 shows the parameters used for learning of the various connections.

Now we describe the training phases leading the model to perform the mental rotation task.

Learning of the inverse model (C7, C8, C9) and of the forward model (C10, C2): The aim of the inverse model learning phase is to obtain the values of the connection weights between SS-M1_1, between PMC_1-M1_1 and between M1_1-M1_2, needed to perform a wrist rotation (encoded by M1_2) driving the current wrist posture (encoded by SS) towards the desired wrist posture (encoded by PMC_1). The learning phase pivots on the following ‘motor babbling procedure’ done with the object rotated by the robot: (a) the robot assumes a random wrist posture within \([-90^\circ, 90^\circ]\), which is encoded by a Gaussian cluster within SS; (b) the random generator randomly decides the direction of rotation (DR) and the number of rotations (NR). For example, if \(DR = 1\) and \(NR = 3\), the robot rotates its wrist clockwise through \(90^\circ (3 \times 30^\circ)\). \(DR = 1\) causes the activation of the neuron N1 of M1_2. \(NR = 3\) implies that N1 is activated for three sequential steps. We assume that ‘one time step’ is the time the robot needs to rotate its wrist through \(30^\circ\); (c) the value of the wrist rotation is used to compute the total rotation (in this case \(3 \times 30^\circ = 90^\circ\)) and, based on the current posture, this is used to activate the PMC_1 map as a possible desired wrist posture; (d) PP neurons encode the current object orientation; (e) at the end of each step the Kohonen rule (Equation 2) is used to update the connection values (C7, C8) in order to obtain different cluster within the M1_1 representing all the combinations of the desired final wrist posture (PMC_1) and the current wrist posture (SS); (e) aside the SOM M1_1, at the end of each step we also train the forward model (C10, C2). Each SOM cluster (M1_1) is associated by the Hebbian rule (Equation 1), with the following wrist posture cluster (SS), which is in turn associated (Equation 1), with the corresponding object orientation (PP) (this corresponds to perform a rotation with an object in the hand and associating the felt proprioception with the seen object image); (f) at the end of each step the clusters activated within the SOM M1_1 are associated to M1_2 activated neuron (C9) using Equation 1. The use of the SOM M1_1 is necessary to learn all the possible combinations between current posture (SS), the desired posture (PMC_1) and control signal (M1_2) needed to accomplish the desired posture. Overall, there are seven possible desired postures encoded in PMC_1 and \(7 \times 14\) possible combinations to be encoded in M1_1.

Learning the affordance-based action pre-activation (C1): The training pattern is formed by two series of rotating images, which differ of \(30^\circ\) of orientation per step. Each image is loaded in PP as the activity level of a set of neurons in the map. The aim of the training process is to create a mapping between the input image (PP) and the corresponding wrist postures of the robot encoded by a cluster of active neurons (Gaussian tuning curve) within PMC_1. The signal from PP pre-activates the clusters within PMC_1 with a value of 0.2 (this activation is obtained by opportunely setting the max value of the C1 connection weights). This means that the object can pre-activate several actions based on the seen object affordances. The signal from PFC_1
allows the full activation, and hence the selection, of one cluster (one desired posture) according to the organism’s goal, in our case the target image within PFC_1 (cf. Section 2.2). The training process is implemented using the Hebbian learning rule (Equation 1).

**Learning action selection (C6):** The training pattern is formed by two series of rotating images, which have a 15° different orientation per step. Each image is loaded into the PFC_1 map as the activity level of a set of neurons in the map. An important difference with respect to the pre-activation of affordances training phase discussed above, is that here the aim of the training process is to create a mapping (through Equation 1) between the specific target image (PFC_1) and specific wrist posture of the robot encoded by clusters of activities (Gaussian tuning curve) within PMC_1. In this way, the signal from PP pre-activates within PMC_1 all the seven possible desirable wrist postures related to the seen object, whereas the signal from PFC_1 supplies the crucial bias signal to select the desired wrist posture related to the target object.

**Learning the matching and the answering processes (C3, C4, C5):** The connections from PP and PFC_1 to the SOM PFC_2 (C3, C4) are responsible for the matching process. When the network generates a mental image in the PP having the same orientation of the target image encoded by PFC_1, then the process of learning is triggered. The connections link two maps: one is PFC_1 (target image), which is set at the beginning of each mental rotation and then kept fixed, and another is PP (the current mental image). A training set for PFC_2 is a combination of all the possible neural representations of PFC_1 and PP. PFC_2 forms a winning cluster of neuron for each two specific inputs. As there are 14 possible images in each input map, 196 clusters will be formed. To train PFC_2 the SOM learning rule (Equation 2) was used.

The answer triggering process uses the connection C5 from PFC_2 to PMC_2. When two images fed to PFC_2 are similar the robot chooses the ‘YES’ answer, otherwise it chooses the ‘NO’ answer (the term ‘similar’ meaning ‘approximately the same’). The mental rotation ends when the cluster of alternative neurons in M1_2 is close to the ‘stay still’ cluster (N2). When this happens, the most salient cluster in PFC_2 is used to produce the answer. Given the 196 possible combinations of inputs in the matching process, half of them are responsible for a ‘Same’ answer, while the remaining half for the ‘Mirror’ answer. Therefore, 98 regions in PFC_2, with respect to the same image from the PP and PFC_1, activate one cluster in PMC_2, while the other 98 regions represent different images of the two input maps and so activate a second cluster. In the current version of the model, the PMC_2 motor command is still not used to supply a control signal for the iCub but is directly interpreted as the response of the system.

After learning, an action potential of each neuron in the PMC_2 map is calculated by using a dynamic competition method (Erlhagen & Schöner, 2002). For this purpose, PMC_2 is endowed with within-map all-to-all connections. The connections follow the rule of long-range inhibition and short-range excitation. This pattern of connections causes a dynamic competition process within the map. Neighbouring neurons that are activated with a high input will receive excitatory signals and tend to form a winning cluster of activity. In contrast, other neurons far from the winning cluster in the neural space will receive an inhibition signal and their activity will be depressed.

The dynamic competition is used as a method to calculate the agent’s RTs to compare the model results with RT data in psychology experiments (Caligiore et al., 2010). Unlike a simple feed-forward process in layered neural networks, the dynamic competition process will be repeated until the action potential of at least one neuron in the neural map reaches a specific threshold. The number of cycles needed to achieve this threshold is used as simulated RTs (one cycle is assumed to correspond to 1 real-time millisecond).

### 3 Results

The two stimuli of the simulated mental rotation task were varied in seven angular positions in the range [−90°; 90°] with a step of 30°. Therefore, the maximum angular disparity between the two stimuli was 180° and required six rotational steps mentally to overlap the left stimulus to the right target one. When the number of rotation cycles was equal to 10, this indicated that the model could not correctly perform the mental image rotation of the left stimulus and so it was forced to do the matching process by using the last image (see Section 2.1).

Figure 3(a) shows the mental rotation steps (PP) and the matching (PFC_2) and answering (PMC_2) processes for a successful trial. In this example the mental rotation process takes five steps to rotate an image of a stimulus at −60° to match it to an image of a target at 90°; both stimuli are object-A. The mental rotation process ends when the rotated image reaches 90°. After that, the matching process within PFC_2 is performed by using as input the neural activity of target image in PFC_1, and the rotated image in PP. The neural activation representing the matching process within PFC_2 is shown in the third column of the last row on Figure 3(a). The neural activity within PFC_2 shown in the figure is the level of action potential of each particular neuron (within the range 0.0–1.0). A salient cluster that is indicated by the black spot is the answer of the map. After applying a filtering process, the cluster with most activity in PFC_2 is used as an input to PMC_2. The answering process of PMC_2 is indicated in the fourth
column of the last row on Figure 3(a). The cluster of activity formed in the left side of the map will cause the answer ‘YES’ to be chosen. The blank panels in the figure indicate that the rotational steps needed in this sample are less than 10.

In contrast, Figure 3(b) shows one case in which the model cannot rotate the left stimulus of 0° into the 60° position of the target stimulus: as indicated by the panels ‘Mental’ and ‘Target’ in Figure 3(b), the final rotated object image is incorrect. The model fails to rotate the image within 10 cycles, and so is forced to trigger an action by using the last image in PP. This scheme is similar to a guessing process in human subjects when the time to perform the mental rotation task is over. The model failure of this case might be caused by a mismatch cluster in SS caused by a noisy cluster position in M1_1.

Possible failures in rotation and response of the model mainly come from the map M1_1 and connection C9 and C10. Because there are many possible patterns in M1_1, these might overlap in part so generating incorrect or noisy activations in M1_2 and SS. This property of the model simulates the error responses found in human subjects as a consequence of wrong working memory reconstructions of the rotated object images.

After testing the model with all possible pairs of stimuli used in the training set, the model achieves a 97.95% (192 out of 196) success rate (rotation of the left stimulus to match the target). The overall percentage of correct responses is 85.7% (168 out of 196).

As indicated by the RTs profiles showed in Figure 4(a), when the angular disparity is high the required number of cycles of rotation and RTs also increases. The angular disparity (x-axis) is calculated by using the difference in orientation between the two stimuli. A 0° disparity corresponds to the left stimulus orientation equal to the one of the right target object (but can be from a different type of object). As indicated by the RTs profile, there is no significant effect from the different types of object that are used in rotation.

There are three types of errors incurred by the model. The first is from the situation where the model cannot rotate the left stimulus to match the right one within 10 rotational cycles. An error of this type causes a higher RT than normal cases, and also an incorrect response. Secondly, as the connections from SS to PP underlie mental imagery, a possible error in SS directly affects a mental image in PP. In some cases, this leads to a successful rotation by chance. In detail, when active neurons in M1_1 cause an incorrect cluster in
SS, this might be a cluster that causes the image of the target. In this case, the number of rotations will be less than usual. Last, even when the model can successfully rotate the left stimulus the answer might not be always correct. The first two errors are caused by a wrong neuron activity within M1_1 while the last error is caused by PFC_2.

In the introduction, we mentioned that important empirical evidence shows that mental rotation processes are embodied in that they involve the same brain structures involved in overt sensorimotor processes. In particular, various experiments show that the performance of overt actions interferes or facilitates mental rotation processes (Wexler et al., 1998; Wohlschläger, 2001; Wohlschlager & Wohlschlager, 1998). We ran an experiment with the model to start to investigate these phenomena, illustrated in the following.

The proprioceptive signal in SS has been simulated by using the current wrist angle of the robot. This process acts as a cluster pre-activating the map SS. When the position of the pre-activated cluster and of the cluster caused by M1_1 are the same, or overlapped, this should support the rotational processes and so the RTs are expected to be reduced. In contrast, if they are different the dynamic competition process should take a longer time to activate the most salient cluster within the map.

The results, illustrated in Figure 4(b) indicate that the model produced different RT profiles when matching, mismatching, or no proprioceptive signals were supplied (as in the simulations presented above) to SS. In the matching conditions, the signal from the current wrist posture pre-activated the same cluster in SS as the one proprioceptive signal. In contrast, in the

Figure 4. The comparison of response time profiles with different proprioceptive signals. (a) Difference of response time profiles between different pairs of stimuli; AA denotes that the left stimulus is object-A and the target object-A, while AB, BA, and BB denote the other possible combinations. (b) Difference of response times when supplying a matching, mismatching or no proprioceptive signal (corresponding to the normal operation of the model) to the somatosensory (SS) cortex.
mismatching condition a random cluster is sent to SS, so the competition within SS has to use more cycles to form a salient cluster and this slows down the RTs. However, the perturbed proprioceptive input does not affect the accuracy of the response. Therefore, in the current setting the signal from the wrist proprioception affects only RTs.

Although preliminary, the results of the experiment represent an important starting point to design future extensions of the model directed to account fully for the relations existing between covert and overt mental rotation processes.

4 Discussion

Various studies support the view presented here on the basis of the model, for which mental rotation processes heavily rely on sensorimotor brain structures that play an important part in over action. In this respect, however, it is possible that the degree of interaction of the two classes of processes change during development. For example, in studies of mental rotation in human children an interesting finding indicates that the link between motor performance and mental rotation are more pronounced in children than in adults (Frick, Daum, Walser, & Mast, 2009; Funk, Brugger, & Wilkening, 2005). However, the empirical study by Krüker and Krist (2009) showed opposite findings in which the motor process was less pronounced in the participants aged 5–6years than in 7-year-old children and in adults. The speed of mental rotation also depends on age and improves with development (Kail, Pellegrino, & Carter, 1980). These phenomena might be addressed in future experiments testing the model at different phases of learning or considering other types of learning processes like reinforcement learning (e.g., Sutton & Barto, 1998), applicable in a modelling neurorobotic context as here (Herbert, Ognibene, Butz, & Baldassarre, 2007; Ognibene, Rega, & Baldassarre, 2006), which allow overcoming the limitations of associative forms of learning as those used here (Caligiore et al., 2008b).

Other important aspects not considered here are related to other types of feedback aside proprioception. In this respect, the visual input of seen hands, not modelled here, plays a central role. Indeed, it might be combined with the proprioceptive signal to produce a matching/mismatching effect as the one shown here for proprioception in SS. The role of seen gestures has not been studied in depth yet, so there is no evidence on whether people benefit from such input when dealing with mental rotation problems. Some researchers (e.g., Goldin-Meadow, 2005) claim that even blind people produce gestures when they talk. This might suggest the importance of motor processes over perceived image of the hands or objects. In addition, motor processes and visual perception of moving hands might be attended differently in different contexts. In particular, attention mechanisms may lead the subject engaged in the mental rotation task to neglect the seen hands to better focus on the task (Pannebakker et al., 2011). The model might face this problem by sending an additional visual input to PC, an important locus for the integration of proprioceptive and visual information (Hagura et al., 2009).

The model generates errors, but in its current version, it does not do so in relation to the angular disparity and hence the difficulty of the rotation task as it happens in human subjects. This limitation might be investigated in future work. For example, at the moment, the model can process only two types of objects and this might create mental rotation processes that do not degrade with the number of rotation steps. Endowing the model with the capacity to rotate any type of object might make it more prone to errors when the rotation task becomes more challenging. To permit the rotation of unseen objects, the object orientation detection function might be separated by the object identification one, e.g., using an inferotemporal cortex (IT) map whose neurons encode objects identity independently of their orientation (Goodale & Milner, 1992).

Due to individual differences, people can apply a variety of strategies to solve mental rotation tasks such as: using their own hand to indicate the movements of a stimulus, imagining rotation of the stimulus itself, or even using non-rotational strategies. There is no right or wrong strategy to solve mental rotation tasks. In future work, the integration of some of these abilities and strategies might be incorporated in the model to account for the variety of human performances in solving mental rotation tasks.

5 Conclusion

The neurorobotics model proposed in this paper accounts for mental rotation processes based on neural mechanisms involving visual imagery, bottom-up and top-down control, and mental imagery based on inverse and forward models. The model also highlights the importance of motor processes and proprioceptive inputs in the performance of mental rotation tasks. In this respect, the proposed approach agrees with the most recent theoretical and empirical findings on mental rotation (Lamm et al., 2007) and more in general mental simulation (Pezzulo, Barca, Bocconi, & Borghi, 2010).

Importantly, in addition to replicating the typical mental rotation data, the model is able to account for other data, which link overt movements and mental rotations (Wohlschläger, 2001; Wohlschläger & Wohlschläger, 1998; cf. Introduction). This recent
empirical evidence shows that the performance of mental rotation tasks can be improved by the assistance of hand movements, or gestures, called 'co-thought gestures' (Chu & Kita, 2008, 2011). Spontaneous gestures during the performance of mental rotation provide a rich sensorimotor experience. Following this evidence, the model includes proprioceptive areas that encode the proprioception resulting from wrist movements. This directly affects the mental rotation processes within the parietal–premotor circuits. On this basis, the model suggests an operational hypothesis on the specific mechanisms through which covert mental rotation processes might rely on overt ones on the basis of forward models.

The model was also validated with the simulated humanoid robot iCub engaged in solving a mental rotation task. This gave further support to the idea that the integration of mental rotation capabilities with affordance and embodied processes is at the basis of the successful performance of the mental rotation tasks.

For its embodied nature, the model presented here also sets the basis for investigating the role of co-thought gestures (Chu & Kita, 2008, 2011) to support mental rotation tasks, as well as other cognitive capabilities such as the use of communicative gestures and verbal language.

Overall, the proposed neurorobotic model provides a useful computational framework to study the integration between mental rotation capabilities and embodied cognition, in particular to demonstrate the role of motor processes and forward models in mental simulation tasks.

Funding

This research received funds from EU funded projects RobotEra (FP7-IP-288899), Poeticon + + (FP7-STREP-288382), a grant from the Thai government, and ‘IM-CLeVeR-Intrinsically Motivated Cumulative Learning Versatile Robots’, contract no. FP7-IP-231722.

References


About the Authors

Kristsana Seepanomwan received an MSc in computer science from King Mongut's University of Technology North Bangkok (KMUTNB), Bangkok, Thailand, in 2007. In 2010, he received a scholarship from the Thai Government to study a doctoral program in the field of robotics. Currently, he is a PhD research student at the centre for robotics and neural systems, Plymouth University, UK. His research interests include brain-like processing mechanisms and focus on formation of planning, language, and thoughts in humanoid robots.

Daniele Caligiore obtained a Masters in electronics engineering from the Universita’ degli Studi di Catania, Catania, Italy, in 2003 and a PhD in biomedical engineering from the Universita’ Campus Bio-Medico di Roma, Roma, Italy, in 2011. During his PhD, he was visiting scholar at the University of Plymouth, Plymouth, UK, to collaborate with professor Angelo Cangelosi and professor Rob Ellis, and at the Universita’ di Bologna, Bologna, Italy, to collaborate with professor Anna M Borghi. Since 2004, he has worked with the Institute of Cognitive Sciences and Technologies, which is part of the National Research Council, based in Rome, Italy (ISTC-CNR). During this period, he has participated as an investigator in several projects funded by the European Commission in the field of cognitive robotics and embodied cognition (Mind RACES – from Reactive to Anticipatory Cognitive Embodied Systems; ROSSI – Emergence of communication in RObots through Sensorimotor and Social Interaction; IM-CLeVeR – Intrinsically Motivated Cumulative Learning Versatile Robots). He has published more than 40 peer-review international papers. He is currently a post-doc researcher at Laboratory of Computational Embodied Neuroscience (LOCEN), which is part of the ISTC-CNR. His research interests include motor development, computational embodied neuroscience, affordances, brain cortical and sub-cortical hierarchies; reinforcement and hebbian learning. Recently, he was guest-editor for a special issue of the international journal Psychological Research on ‘Embodiment of Vision, Action and Language’.

Gianluca Baldassarre In 2003, he received a PhD in computer science at the University of Essex, UK, and then he did a post-doc at the Italian Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy, working on swarm robotics. From 2006, he has been a researcher at this same institute and coordinates the research group that he formed called ‘LOCEN – Laboratory of Computational Embodied Neuroscience’. In 2006–2009, he was the team leader of the EU project ‘ICEA – Integrating Cognition Emotion and Autonomy’ and since 2009 he has been the coordinator of the European integrated project ‘IM-CLeVeR – Intrinsically-Motivated Cumulative-Learning Versatile Robots’. His research interests are on cumulative learning of sensorimotor skills driven by extrinsic and intrinsic motivations in animals and humans. He studies these topics with both models constrained on the basis of data from psychology and neuroscience, and with embodied/robotic models. He has produced about 80 international peer-review publications. Email: gianluca.baldassarre@istc.cnr.it Website: http://www.istc.cnr.it/people/gianluca-baldassarre

Angelo Cangelosi is professor of artificial intelligence and cognition, and the director of the centre for robotics and neural systems at Plymouth University (UK). His main research expertise is on language and cognitive modelling in humanoid robots, on language evolution in multi-agent systems, and the application of bio-inspired techniques to robot control (e.g., swarm of UAVs). He is the coordinator of the Marie Curie ITN ‘RobotDoC’ (2009-2013), the UK EPSRC project ‘BABEL’ (2012-2016) and the FP7 project ‘ITALK’ (completed in 2012). Cangelosi has produced more than 200 scientific publications, is editor-in-chief of the journal Interaction Studies, and has chaired numerous workshops and conferences including the IEEE ICDL-EpiRob 2011 Conference (Frankfurt, August 2011). In 2012, he was nominated chair of the International IEEE Technical Committee on Autonomous Mental Development.