Cogn Comput (2011) 3:525–538
DOI 10.1007/s12559-011-9108-1

Modeling the Development of Goal-Specificity in Mirror Neurons

Serge Thill · Henrik Svensson · Tom Ziemke

Received: 22 February 2011 / Accepted: 26 August 2011 / Published online: 29 September 2011

Abstract Neurophysiological studies have shown that parietal mirror neurons encode not only actions but also the goal of these actions. Although some mirror neurons will fire whenever a certain action is perceived (goal-independently), most will only fire if the motion is perceived as part of an action with a specific goal. This result is important for the *action-understanding hypothesis* as it provides a potential neurological basis for such a cognitive ability. It is also relevant for the design of artificial cognitive systems, in particular robotic systems that rely on computational models of the mirror system in their interaction with other agents. Yet, to date, no computational model has explicitly addressed the mechanisms that give rise to both goal-specific and goal-independent parietal mirror neurons. In the present paper, we present a computational model based on a self-organizing map, which receives artificial inputs representing information about both the observed or executed actions and the context in which they were executed. We show that the map develops a biologically plausible organization in which goal-specific mirror neurons emerge. We further show that the fundamental cause for both the appearance and the number of goal-specific neurons can be found in geometric relationships between the different inputs to the map. The results are important to the action-understanding hypothesis as they provide a mechanism for the emergence of goal-specific parietal mirror neurons and lead to a number of predictions: (1) Learning of new goals may mostly reassign existing goal-specific neurons rather than recruit new ones; (2) input differences between executed and observed actions can explain observed corresponding differences in the number of goal-specific neurons; and (3) the percentage of goal-specific neurons may differ between motion primitives.

Keywords Mirror neurons · Action-understanding hypothesis · Computational model · Neural activation patterns · Self-organizing map

S. Thill (✉) · H. Svensson · T. Ziemke
Cognition and Interaction Lab, School of Humanities and Informatics, University of Skövde, P.O. Box 408, 541 28 Skövde, Sweden
e-mail: serge.thill@his.se

Introduction

Functional Roles of Mirror Neurons

Since the mid-90s, mirror neurons have received a considerable and increasing amount of attention. Indeed, the possibility that neurons that are active both when agents observe a goal-directed action and when they execute the same action [1] has lead to the hypothesis that they may be the crucial link between perceiving and understanding actions. Such a mechanism is attractive to many fields: neurophysiologists, for instance, may be chiefly interested in understanding the precise functioning of the underlying neural mechanisms, while cognitive scientists are provided with a possible pathway for social understanding and interactions. The interest in mirror neurons also extends to the fields of artificial intelligence and robotics [2–4], where mechanisms that allow understanding and even learning the intentions or actions of others (for instance, through imitation) are a very active research topic.

Consequently, the literature is now swamped with papers on mirror neurons. A recent review paper [5] lists no less than 125 references even though it focuses only on one aspect of mirror neuron research (their functional role). Nevertheless, the debate over what cognitive functions

mirror neurons actually underlie is still going strong. The most important claim is the aforementioned linkage between action perception and action-understanding [5–9] but even that does not go without criticism. For example, it has been pointed out that there is not actually any conclusive evidence that mirror neurons are necessary for action understanding [10].

In terms of neurophysiological data to support the action-understanding hypothesis, a large proportion of recorded parietal mirror neurons in macaque monkeys, all shown to encode grasping, only fire if the goal of the overall observed action is either to eat the grasped object or to place it in a container [6]. Moreover, only a small proportion of neurons fire in both goal scenarios, whereas the remaining neurons appear to be selective. There is evidence that a similar mechanism may be at play in the human brain [7]. Another study presented an experiment in which monkeys grasped objects with pliers [8]. These pliers were either "normal," requiring a closing of the hand to close them or "inverted," requiring an opening of the hand to close them. Mirror neurons in area F5 were found to respond in a similar fashion to grasping irrespective of the type of pliers used, indicating that the neurons were encoding the "grasping" concept rather than the underlying motor commands.

Action understanding aside, mirror neurons are also thought to play a role in learning by imitation [11] as well as in the sensorimotor grounding of language (see [12] for a discussion). A more radical argument that mirror neurons are essential for the evolution of language is made largely by philosophical reasoning [13], supported by the hypothesis that Broca's area, involved in human language, may have evolved from area F5 in monkeys [14]. Again, these hypotheses do not go without criticism; it is for instance repeatedly pointed out that macaque monkeys, from which most of the currently existing neurophysiological data have been obtained, neither use language nor imitate [11]. Mirror mechanisms are also thought to be a key to understanding the close relationship between perception, action and social cognition [15]. Further, there are indications that mirror mechanisms are also involved in understanding the emotions of others. For example, it was found that observing facial expressions of disgust activated similar brain areas (anterior insula and to some extent anterior cingulate cortex) as when being exposed to different disgusting odors [16]. Mirror mechanisms can also extend to the sensation of, for instance, touch or pain [17, 18].

Hence, it is clear that the hypothesized functions of mirror neurons are all fundamental to the interaction between agents. It thus comes as no surprise that mirror neurons are also of high interest to the field of robotics [19, 20], in particular humanoid robotics, as they may hold the key to solving current challenges in designing robots that can interact robustly with humans and use, for instance imitation learning to survive in unknown environments. Robots may even be able to learn rudimentary forms of language based on computational mirror system models [21].

Mirror Neuron Models

In a recent review of computational mirror neuron models [11], a taxonomy based on the methodology underlying such models has been proposed, identifying four main categories: "data driven," "reason for existence," "assume existence" and "evolutionary algorithm," which are in some cases further subdivided. The main motivation underlying this division is that different methodologies can investigate different aspects of mirror systems. For instance, models in the "reason for existence" and "assume existence" categories try to, one way or another, determine the functions of the mirror system and replicate this in artificial agents to endow them with the same capability. Imitation is the chief example of such a functionality [3, 22, 23], but language understanding in robots [21] is another.

Data-driven models on the other hand attempt to collect available information on mirror neurons into a general model in order to produce new predictions. The chief examples in this category are the MNS [24] and MNS2 [25] models. For instance, although originally focusing more on the anatomy of the mirror system (and thus primarily based on monkey data), the MNS2 model has recently [26] been used to model behavior observed in a cat, hypothesizing that mirror neurons may be able to "reflect" on an agent's own actions and allow rapid reorganization of motor programs. Based on the model, it is possible to propose several mechanisms that may be underlying observed behaviors even in humans, but it remains to be verified to what degree these are actually biologically plausible. At the same time, as briefly discussed before, uncertainty about biological plausibility does not prevent the use of the proposed mechanisms in robots.

It is particularly interesting to note that when the review [11] was published, models driven specifically by physiological data were still missing from the scene. This is beginning to change; *Chain models* [4, 27, 28] for instance build upon the previously mentioned neurophysiological findings [6]. They have been used, for instance, to model the development of intention understanding [4, 27] and to explain conflicting data regarding facilitation and interference effects during sentence processing [12].

Developmental Models of Mirror Neurons

Relatively few models address the actual development of the mirror system (but see, for instance, [4, 30] and to some

extent [25]). However, from a robotics as well as an embodied cognition perspective, this is a highly interesting issue. In robotics, a developmental account removes, or at least significantly reduces, the need for a hard-coded system, opening instead the way for online, adaptive and human-interactive implementations. Further, such an account can increase our understanding of how the body, in particular bodily differences (for example, between humans and humanoid robots or between adults and children) and bodily experience (both individual and social), might shape the mirror system. Although Erlhagen et al. [4] address this in part by providing an online mechanism for learning chains of motion primitives, the use of a Hebbian learning approach may ultimately be limiting.

Our previous developmental model [29] uses self-organizing maps to illustrate that those can organize into region, or "pools" of nodes encoding specific motion primitives, in a way that is hypothesized by the chain models [28] based on the work of Fogassi et al. [6]. Although this model is able to learn new inputs during runtime, it does not take into account the goal-specific aspects of the neurophysiological data [6], nor does it model the formation of chains previously addressed by, for instance, [4].

Finally, a bio-robotic approach illustrating that mirror-like representations, as observed in brain area F5, can develop simply from the interaction of an agent with its environment has been presented by Metta et al. [30]. Although the overall structure is similar and largely in agreement with the MNS model, particular efforts were made to use unsupervised learning algorithms and to realize a partial robotic implementation.

Modeling the Development of Goal-Specificity in Mirror Neurons

The main purpose of the model presented in this paper is to provide an explanation of how a mirror system may develop goal-specific neurons as described by [6], an interesting feature of mirror neurons that has so far not received much attention by computational modelers. Specifically, we investigate (1) whether such an organization can develop without assumptions on the functional role of mirror neurons and (2) which aspects of the inputs to the mirror system may influence the development of such goal-specific neurons. We build upon our previous work [29] to detail how a "blank" structure (representing parietal mirror neurons that have seen some previous attention by modelers [28, 29]) can develop a representation of its inputs in line with the finding that most but not all neurons encoding a given motion primitive are also sensitive to the goal of the overall action [6]. We show that these findings can be reproduced without making any assumptions regarding functional roles of mirror neurons. We are further able to detail the mechanism underlying the emergence of goal-specific neurons and determine which aspects of the model input control their number. This has two important implications. First, it provides computational modelers and those interested in biologically inspired robotics with a simple, controllable way of reproducing and implementing the firing patterns observed in biology, so that they may be used in the implementation/development of an artificial agent's behavioral and cognitive capacities. Second, it provides a testable hypothesis that similar mechanisms might be at work in biological mirror systems, which in turn may increase our understanding of the evolution of the mirror system.

In terms of Oztop's taxonomy [11] then, the present paper spans several categories. It is data driven in the sense that the main aim is to reproduce a biologically observed firing pattern (albeit at a necessary level of abstraction). It falls into the "reason of existence" category in the sense that it aims to elucidate possible fundamental causes of this firing pattern. Although not using an evolutionary algorithm, the model does thus also address whether or not the mirror system would necessarily have evolved to represent goals and could, to some extent, be included in the "evolutionary algorithms" category. On the other hand, the model does not assume a particular cognitive role for the mirror system, thus not falling into the "assume existence" category.

Assumptions Made in the Present Model

A computational model necessarily comes with some assumptions and simplifications. It is therefore important to discuss these before describing the model in detail. First, we choose to model the mirror system using a self-organizing map paradigm [31]. The idea that mirror neurons can be seen as a form of associative network is not new; it is supported by, for instance, the hypothesized organization of mirror neurons into functional groups or pools, each encoding a specific type of motion [6, 8, 28]. In computational models, associative networks are a popular approach for reproducing at a minimum a certain functionality (e.g., multimodal integration) of mirror neurons (see [11] for a section reviewing a number of such models). In particular, it has been argued that the "reason of existence" of mirror neurons in such models is phenomenological rather than functional [11]. Since a point of the present paper is precisely to illustrate that goal-specific activity [6] can emerge even without assuming a functional role of mirror neurons (which does not preclude a functional use of the resulting activity later on), this is a good modeling approach for the present purposes.

It should also be noted that some associative network models of mirror neurons have in fact specifically used self-organizing maps [2, 21] and illustrated that they can be used for endowing robots with certain cognitive functions (in this particular case, a form of language comprehension). The approach in the present paper is somewhat different since, rather than a hierarchy of maps, we simply use one.

The second assumption, or more precisely set of assumptions, concerns the model inputs. The mirror system is known [25] to receive input from, among other brain areas, the superior temporal sulcus (STS) and the anterior intraparietal area (AIP) and prefrontal cortex (PFC). The former is thought to deliver information on observed movements of others (whereas proprioceptive feedback comes from the premotor areas, in particular the canonical neurons), while the latter two transmit information about the affordances of an object in view and the action goal if the action is executed by the agent itself (see also [32] for a more detailed modeling of affordances). These are of course not the only inputs into the relevant parts of the premotor areas, other information can come for instance via area 7b [25] or the prefrontal cortex [28, 32]. For the present purposes, however, we will limit ourselves to the two inputs mentioned above: (1) an encoding of observed or executed actions (obtained from the STS and canonical neurons) and (2) an encoding of the affordances of an object, which provide information on the context and likely goal of the actions.

Our main concern is showing how modifying certain aspects of the inputs affects a possible goal-specific activity in the map. This is in a sense a theoretical argument and requires the liberty to manipulate the inputs at will. It would thus not be beneficial to implement the model in an actual agent (robotic or simulated) in order to obtain the inputs. At the same time, however, the inputs must retain some relation to those that a sensorimotor system might receive and generate. We thus limit our analysis to features that are likely to hold true for any such implementation, irrespective of the actual sensorimotor system and neural circuitry. Aside from providing freely manipulable inputs, this approach has the added advantage that it does not accidentally tie the results to a specific implementation only. Rather, the results will ideally be relevant for any such system.

The main disadvantage of course is that the model does not address in detail how inputs can be obtained. However, these are non-trivial but solvable problems that have a strong research community behind them. In robotics, for example, even the perception of human motion is a difficult challenge, often relying on motion capture systems that may require the observed human to wear markers [33, 34], although markerless approaches exist [35]. Generally, recognition and segmentation of motion have been dealt with many times, both in the modeling of mirror neurons and, generally, in the field of robotics [3, 24, 25, 33, 34, 36, 37]. The typical common ground (in modeling terms) is that motion recognition is based on an evolution of spatial coordinates over time. These can simply be coordinates of the end-effector, whether Cartesian [25] or joint angles [36]. Recently, it has been shown that human action segmentation of object-centered hand and arm movements is related to the kinematics of the wrist, particularly its change in direction [38]. Other approaches consider the body in its entirety [34], for instance, demonstrated an online motion-segmentation algorithm of whole-body motion based on hidden Markov models.

Recognition of the context in which an action is taking place is likewise a problem under heavy investigation. Since mirror neurons only fire when an action toward an certain object is executed or observed, this recognition is likely to involve the processing of said object's affordances. Reviewing the literature on affordances is beyond the scope of this paper, but a helpful discussion can be found for instance in [39] and, more recently, [32].

In humans and monkeys, several brain areas are involved in the processing of affordances, including the visual cortex (VC, for object edge detection), the AIP (detection of object shape), the parietal reach region (PRR, detection of object position) and, importantly, the ventral occipitotemporal cortex (VOC, encoding the object identity) [32]. In earlier work, the FARS model [40] considers the AIP to determine different types of grasps afforded by an object, of which area F5 (which includes mirror neurons) then selects the most appropriate one. Related work [24] extends this principle. From a robotics point of view, for instance, [39] formulates a psychologically motivated definition of affordances as, broadly speaking, a relationship between an agent, an object and the environment. The final encoding of context, as relevant for the present work, can therefore be seen as a set of affordances currently present in the environment. Taking the setup by [6] as an example, this set could for instance be "placeable" for a solid object and "eatable, placeable" for food.

In sum, it is clear that the generation of the inputs may involve non-trivial processing of sensory data which depends on the agent under consideration, whether living or artificial. However, our main concern in this paper is to show how the *processed* data reaching the mirror neurons may affect their organization. We therefore leave out a detailed modeling of a neural pathway from sensory input to mirror neurons and only make two assumptions: irrespective of the form (e.g., a vector of values, neural

activation) or the detailed source (e.g., joint angles or end-effector coordinates in the case of motion inputs), both the motion and the context inputs can be represented in a space that is (1) finite and (2) wherein different motions (or contexts encodings involving objects with different affordances) form distinct, mostly separable subspaces.

Methods

Mirror Neurons as a SOM

The present work is based on modeling the mirror neuron system as a self-organizing map (SOM, [31]). Using such maps is appropriate here since the focus is on *organizational* principles of the modeled system. Although more neurophysiologically plausible approaches exist [12], they typically rely on hard-coding a large number of free parameters in order to show the desired effect and thus lack the autonomous self-organizing aspects that are important when modeling *developmental* phenomena. In the present case, it is preferable to have a modeling approach that can address the developmental aspects even though it comes at the expense of precise neurophysiological detail. The qualitative nature of the results is, however, not affected by such a lack of detail.

Generally, the viability of a SOM-based approach to the modeling of mirror systems has also been shown previously [21, 29]. The basic modeling approach presented here follows that of the latter paper. The total input vector for the SOM is thus composed of two parts—one part that encodes a motion primitive either observed or executed by the agent in which the mirror system is embedded and a second part encoding the context in which this primitive is observed. As discussed previously, it is simply assumed that vectors representing motion primitives are sampled from clusters of data points and that clusters for different primitives are distinct. There is thus some variability between different vectors representing the same motion primitive. We further postulate that data points representing motion primitives executed by different types of limbs (e.g., arms and legs) are also separated in the input space. Inputs defining the context are likewise sampled from distinct clusters of their own.

Thus, if $\vec{m}(m_1, \ldots, m_n)$ is a vector representing an encoding of a motion primitive and $\vec{c}(c_1, \ldots, c_m)$ is similarly a vector representing contextual encoding, then the resulting input vector to the SOM is given by $\vec{i}(m_1, \ldots, m_n, c_1, \ldots, c_m)$. We will show below that the results presented here do not depend on specific lengths of these vectors. In other words, the results are not affected by the dimensionality of the input data.

SOM Initialization and Training

As in our previous paper [29], the SOM is initialized through an infancy phase, roughly corresponding to a motor babbling phase, in which it is merely exposed to points randomly sampled from input spaces representing two limbs (e.g., arm and leg), which ultimately results in two regions, each representing one type of limb, within the map (see Fig. 1a). During this phase, the neighborhood function n_t and the learning rate α_t decrease from their maximal values to low (but nonzero) final values (here set to $n_{\min} = 1$ and $\alpha_{\min} = 0.2$), according to the following equations:

$$n_t = n_{\min} + \lfloor (s - n_{\min})\tau_t \rfloor \quad (1)$$

$$\alpha_t = \alpha_{\min} + (1 - \alpha_{\min})\tau_t \quad (2)$$

with s being the side length of the map, $\tau_0 = 1$ and $\tau_t = \max(\tau_{t-1} - 1/t_{\inf}, 0)$, where $t_{\inf} = 5{,}000$ defines the duration of the infancy phase. The SOM thus remains capable of online learning (e.g., it can learn to represent novel motion primitives) but will not dramatically change in organization (see [29] for details). Following the infancy phase, the network is specifically exposed to data points representing five different motion primitives shown in two different contexts for a period equal in length to the infancy phase. During this phase, the map develops regions that specifically represent these primitives in the area corresponding to the related limb (see Fig. 1 for an example). The overall nature of this layout mimics the hypothesized organization of the corresponding premotor areas [6, 28, 29]. Here, the main focus is on the detailed organization of these regions under different conditions. For the remainder of the paper, we therefore only consider motion primitives from one of the possible limbs.

Neural Activity in SOMs Based on Input Distance

In traditional SOMs [31], the weight vector of every node is compared to the input vector and a winning node, which with the shortest distance to the input vector, is determined. Here, however, we are more interested in the behavior of all nodes than the mere location of the winning node. Conceptually, one could thus see the nodes in the SOM as neurons whose activity is inversely correlated with the distance d to the input vector. For the purposes of this paper, the values of d for all SOM nodes are therefore the main parameter of interest (see Fig. 1 for an example). Although it might be tempting to translate d into proper neural activity for all subsequent analysis, doing so would merely introduce freely tunable parameters (e.g., in the

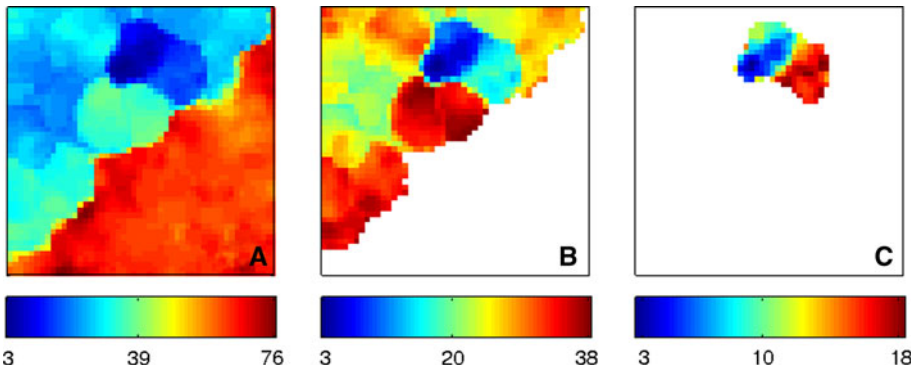

Fig. 1 SOM visualization emphasizing distance between input vector and nodes in a trained map. The figure shows distance (in color-coding) from a target vector for every node. In this case, the target vector has been sampled from a region of the total input space that represents a motion primitive related to the first trained limb. **a** Distances for all neurons, separation of the map into two body part regions is clearly visible, with roughly half the nodes (in *blue*) much closer to the input vector than the other half (*red*). **b** Same distances but recolored after limiting the visible nodes to those encoding the first limb (i.e., whose weight vectors fall into the first limb's cluster within the input space, corresponding to the *blue* region from **a**). Several regions stand out, but only one is characterized by small distances. **c** Visible nodes now further restricted to those whose weight vectors indicate an encoding of the motion primitive presented in the input vector (corresponding to the *blue* region in **b**). Again, a separation into two classes is clearly visible, this time based on the neuron's goal preference

usually non-linear neural response function or connectivity weights). Since these parameters could in principle be tuned to highlight any behavior of interest, it is preferable to avoid these extra (but unnecessary for the present purposes) parameters and focus on the neural inputs in the form of d here.

Parameters Affecting the Distance

The immediate question is which input features determine the distance d between input and weight vector. To answer this, we identify the maximal value this distance can have under the constraint that both the input and the weight vector encode a given motion primitive in a given context. Remembering that the input vector $\vec{i}(i_1,\ldots,i_{n+m})$ is really a composition of two vectors (of length n and m, representing motion primitive and context, respectively) and calling the corresponding parts of the weight vector $\vec{x}(x_1,\ldots,x_n)$ and $\vec{y}(y_1,\ldots,y_m)$, d is given by:

$$d = \sqrt{\sum_{j=1}^{n}(m_j - x_j)^2 + \sum_{j=1}^{m}(c_j - y_j)^2} \tag{3}$$

where maximal values for d are simply determined by the argument to the square root. Thus:

$$\max^2(d) = \max\left(\sum_{j=1}^{n}(m_j - x_j)^2 + \sum_{j=1}^{m}(c_j - y_j)^2\right) \tag{4}$$

where it should be noted that there is a slight abuse of notation in that the argument to the $\max(\cdot)$ function represents not a scalar function but a list of all possible combinations of choices for $\vec{m}, \vec{c}, \vec{x}$ and $\vec{y}$. Since all terms are positive, this is equivalent to:

$$\max^2(d) = \max\left(\sum_{j=1}^{n}(m_j - x_j)^2\right) + \max\left(\sum_{j=1}^{m}(c_j - y_j)^2\right) \tag{5}$$

Since both components of the input vector are sampled from delimited clusters and assuming, since the present model uses a SOM, that the weight vector of trained neurons will also fall within these clusters, the highest possible value for each term in Eq. 5 is simply the distance between the two most distant points in the respective cluster. It is always possible to surround the clusters by a hypersphere whose diameter is given by those two points (and whose center is given by their average coordinates). Calling the radii of these hyperspheres r_m and r_c for the motion primitive and context clusters, respectively, and noting that $r_c = r_m/\beta$ for some β gives after simplification:

$$\max(d) = 2r_m\sqrt{1 + \frac{1}{\beta^2}} \tag{6}$$

The maximal possible distance d between an input vector encoding a given motion in a given context and a weight vector encoding the same motion and context is thus given by Eq. 6 (while the minimal distance is of course 0), which illustrates that the only input space parameters determining d are the radii of the clusters from which the input components are sampled (and their relative lengths) but not the dimensionality or the location of other clusters, which is expected given the mechanics of a SOM.

Determining Neural Preference

Once the SOM is trained, 100 data points per motion primitive are generated. This input is presented to the SOM with the context components sampled from the first possible cluster, and the average distances of all neurons in the SOM to all inputs are calculated. This is then repeated for the same primitive but with context components now chosen from the second possible cluster. The entire procedure is repeated for other motion primitives. Of fundamental interest are neurons that react to one motion primitive either (1) independently of context or (2) only if the primitive is shown in one of the two possible contexts. The latter class of neurons corresponds to the context-specific neurons observed by [6], while the first class is context unaware.

The neurons in each class are thus determined as follows: For every neuron reacting to a motion primitive, if its mean distance to input vectors given in the context of one goal is lower than the mean minus one standard deviation of the distance to input vectors from the second context, then the neuron is said to be specifically encoding the first goal. In all other cases, the neuron is said to have no goal-specific preference. The advantage of this approach is that it ensures a clear separation between what is and what is not considered a goal-specific neuron. The disadvantage is that the separation is somewhat arbitrary. However, explorations with differently defined separations have shown that the qualitative nature of the results in the present paper are not affected by this, although the precise numerical values will of course vary. As with any computational model, stable qualitative results are more informative than precise numerical outputs, and thus, we do not discuss the effect of varying the definition of goal-specific vs. non-goal-specific in more detail here.

Formally, the definition above can be expressed in a cumbersome but general way as follows: For all contexts j in which a given motion primitive is observed and any neuron n, we can compute the mean distance $\mu_{j,n}$ between n and input vectors sampled from some j as well as the associated standard deviation $\sigma_{j,n}$. For a specific context k, we can then define a set $\mathcal{M}_{k,n}$ containing all contexts j that satisfy:

$$\mu_{j,n} - \sigma_{j,n} < \mu_{k,n} \tag{7}$$

The cardinality of $\mathcal{M}_{k,n}$, written as $|\mathcal{M}_{k,n}|$, determines how preferentially n encodes k. If C is the number of tested contexts, the preference $\text{pref}_{n,k}$ of the neuron n for context k is given by:

$$\text{pref}_{n,k} = 1 - \frac{|\mathcal{M}_{k,n}| - 1}{C - 1} \tag{8}$$

In the case of two contexts (as used here), the preference can only be 1 (high preference, $|\mathcal{M}_{k,n}| = 1$ and $C = 2$) or 0 (no preference at all, $|\mathcal{M}_{k,n}| = 2$ and $C = 2$). However, if more contexts are used, this formulation implicitly permits more graded evaluations. While there are, at present, no neurophysiological data that define neural preferences in more than two contexts, this definition of preference may remain useful in future work.

Simulations

Initially, the two parameters affecting d, namely r_m and β, are varied across a range of values (discussed below) to illustrate their effect on the goal-encoding within the SOM. To obtain a representative dataset, the following is repeated 100 times for every choice of parameters: (1) An input space (containing 5 subspaces encoding motion and two subspaces encoding contextual information) is generated randomly; (2) a map is initialized and trained on the five motion primitives as discussed before; and (3) the map's response to each motion primitive in both possible contexts is measured by presenting 100 randomly chosen input vectors per goal context per primitive to the map and computing the distance of every node to these vectors. Results are then calculated from the entire dataset thus obtained.

Statistics

Most statistical tests reported have a standard 2-way layout. However, unless indicated otherwise, a Jarque–Bera test [41] rejected the null hypothesis that the data are normally distributed, which prohibits the use of a 2-way ANOVA. In these cases, we therefore use the nonparametric Friedman test [42], which is an appropriate substitute if the data are not normal.

Results

Model Parameter Effects

Given the mechanics of a SOM, one would not expect r_m to have a significant effect on the results, since the organization of the map should merely depend on the relative distance between the input vectors. On the other hand, β should affect the organization of the SOM at least somehow since it regulates the relative size of the clusters from which the components of the input vector are sampled. It will thus determine how much of the total input variability can be explained by variation in one component only. Whether or not β will actually affect goal-specificity of neurons in the SOM as defined in "Methods" is, however, less obvious.

A quick exploration of SOM organization at a crude granularity, varying r_m from 10 to 150 in increments of 20 and β from 1 to 5 in increments of 1, confirms the above expectations, showing that r_m has no effect ($df = 7$, $\chi^2 = 4.13$, $P > .74$) while, interestingly, β indeed affects the goal-specificity of the neurons ($df = 4$, $\chi^2 = 3{,}711.57$, $P = 0$).

At this point, it is important to remember that β, as it is used here, is the ratio between the radii of the clusters representing *one* primitive and *one* goal. However, the complete space from which the inputs for the SOM are sampled contains several such clusters—five for the primitives and two for the goals. One could therefore also compute a ratio γ between the radii of the clusters encompassing *all* primitives and *all* goals. Due to the way the input space is randomly generated here, there is no strict 1:1 relationship between values for β and γ in actual data sets but the variability is small and the relationship between the two is mostly linear. It is therefore possible that the goal-specificity is really determined by γ and that β is merely a good indicator. Whether or not this is indeed the case can be tested by increasing the minimum distance between clusters representing individual primitives by some factor f while not modifying the minimal distance between clusters representing individual goals. This increases γ while keeping β constant. When varying β as before and f from 1 to 20 in increments of 5 (4 in the first step), we find that the f has no significant effect ($df = 4$, $\chi^2 = 6.77$, $P > 0.14$) on the percentage of non-goal-specific neurons, even though it affects γ as expected, while β does ($df = 4$, $\chi^2 = 2{,}292.38$, $P = 0$). This confirms β as the parameter of interest here.

Overall, this is therefore an important result: It has been shown that, in the context of this model, the main parameter with an effect on the behavior of interest is the relative size of the clusters representing specific primitives and goals. The detailed effects of β are investigated further in the next subsection.

Goal-Specificity in the SOM and in Monkeys

The effect of β on goal-specific neurons Varying β, as discovered, has a strong direct effect on the number of goal-specific neurons. The larger the value of β, i.e., the larger the diameter of the clusters encoding motion primitives with respect to those encoding contextual information, the smaller the percentage of goal-specific neurons. (Fig. 2). To investigate this with a slightly finer granularity, values were sampled at intervals between 0.1 and 1 in increments of 0.3 as well as between 1 and 5 in increments of 0.5, which, although chosen arbitrarily, can be seen to cover the entire range from purely goal-specific neurons to no goal-specific neurons in those responding to a motion

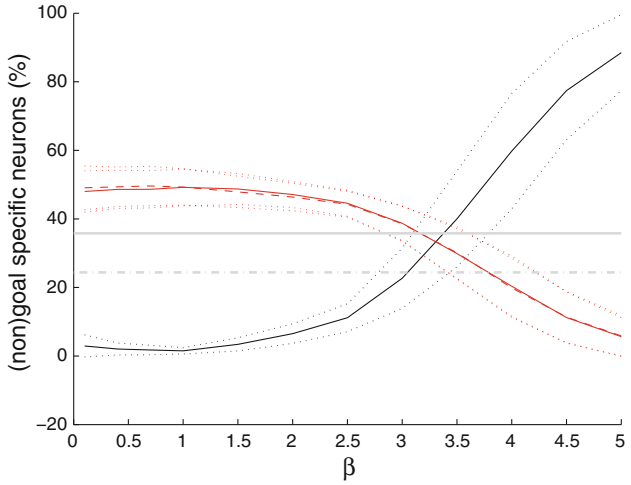

Fig. 2 Effect of β on the specificity of neurons. *Black increasing line* indicates percentage of non-goal-specific neurons, *solid (dashed) red decreasing line* is percentage of neurons preferring first (second) goal. *Dotted lines* indicate ±1 SD. *Solid (dashed) gray straight line* indicated the percentage of non-goal-specific neurons measured during action execution (observation) by [6]. Varying β from 0.1 to 5 can be seen to cover almost the entire percentage range. There is no significant difference between the percentage of neurons preferring the first goal compared to those preferring the second goal (Color figure online)

primitive in general (Fig. 2). Example SOM activations illustrating the difference in goal-specificity of neurons depending on β are shown in Fig. 3. It is also worth noting that we found no significant differences between the number of encoding the different primitives for $\beta \geq 1$ (lowest $P > 0.14$ highest $P > 0.9$) nor, as seen in Fig. 2, any difference between neurons encoding different goals. This is expected and merely serves as confirmation that the results are not due to abnormalities in the model.

A comparison with neurophysiological data can provide a rough estimate of possible β values in macaque monkeys The neurophysiological results [6], which showed that about a third of the measured neurons were not goal-specific, are thus also reproduced by our model. To determine the β value most likely to lead to a SOM reproducing those results, we calculate the probability density functions (PDFs) for the raw data underlying Fig. 2 for every value of β. This is done using a standard kernel smoothing density estimate [43], with a window of 2%. Figure 4 shows the value of every β's PDF at the points $P = 24.4\%$ and $P = 35.8\%$, which are the percentages of non-goal-specific neurons observed by [6] when the monkey was, respectively, observing or executing an action. We find that the probability of obtaining $P = 24.4\%$ is highest for $\beta = 3$, while that for obtaining $P = 35.8\%$ is highest for $\beta = 3.5$. However, it has to be kept in mind that β is technically a continuous parameter which we only sampled at a few intervals and that, likewise, the PDFs are

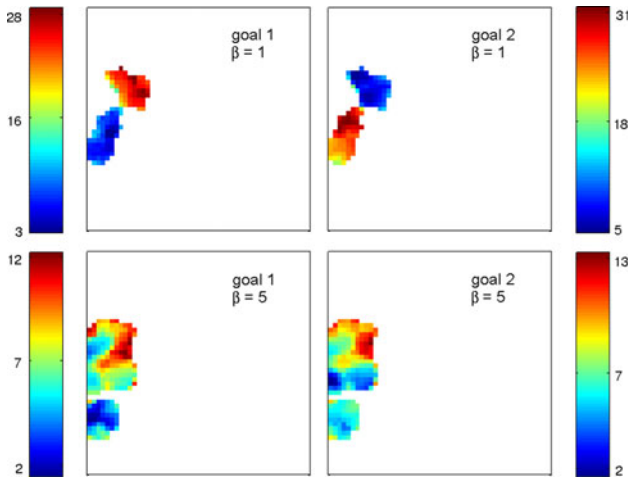

Fig. 3 Goal-dependent activity in neurons encoding one primitive for different β. All plots show the distance between neurons encoding a primitive and an input vector chosen from that primitive (as in Fig. 1c). For the same value of β, the primitive component of the input vector is kept identical and only the goal context component is varied, to clearly isolate the effect of different goal contexts. A small value for β (*top row*) results in markedly different activation patterns for both goals, with neurons clearly separated into two groups, each preferentially encoding one goal. A large value (*bottom row*) causes a slight variation in activation but no clear goal preferences

continuous. This makes strong statements about "realistic" β values very difficult, and the present results should therefore rather be seen as a good initial indication, excluding merely the βs whose PDFs are near zero at the points of interest.

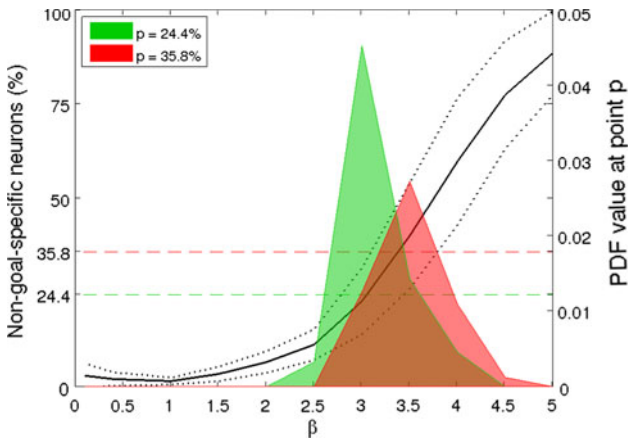

Fig. 4 Best β values for reproducing observations in monkeys. *Left axis*: percentage of non-goal-specific neurons as in 2, shown for easy reference. *Red (green) dashed line* indicates corresponding percentage value of 35.8% (24.4%) found by [6] for monkeys executing (observing) an action. *Right axis*: For every β, the value at points $P = 24.4\%$ and $P = 35.8\%$ of PDFs generated on the dataset used to obtain a mean value for the percentage of non-goal-specific neurons. Observations in monkeys are most likely to be reproduced for $\beta \in [3, 3.5]$

Preferential encoding of some goals can be explained through overrepresentation in training A second interesting aspect of the neurophysiological results [6] was a difference in percentage of neurons encoding each of the goals. Specifically, more neurons encoded "eating" than "placing." This is in all likelihood due to the fact that the monkeys were exposed to the first goal more often in their life than the second (placing is not an action that comes naturally to monkeys). In a SOM, one would expect an overrepresentation of inputs from a specific region of the total input space to affect the resulting organization of the map. Whether or not this would actually affect goal-specificity of the neurons as defined here is less obvious though. To test this, we vary β again at a higher granularity from 1 to 5 in increments of 1.0, while the proportion P of motor primitives shown in the context of the first goal (as opposed to the second one) is given by:

$$P = \frac{n}{n+1} \tag{9}$$

where n is varied from 1 to 5, also in increments of 1. The results show that P indeed affects the proportions of goal-specific neurons and the percentage of neurons preferentially encoding the first goal is correlated with the overrepresentation during training of the SOM (Fig. 5).

Theoretical reasons for the β effect The main mechanism underlying SOMs is the organization of the neurons so that similar inputs activate neighbouring neurons. Since

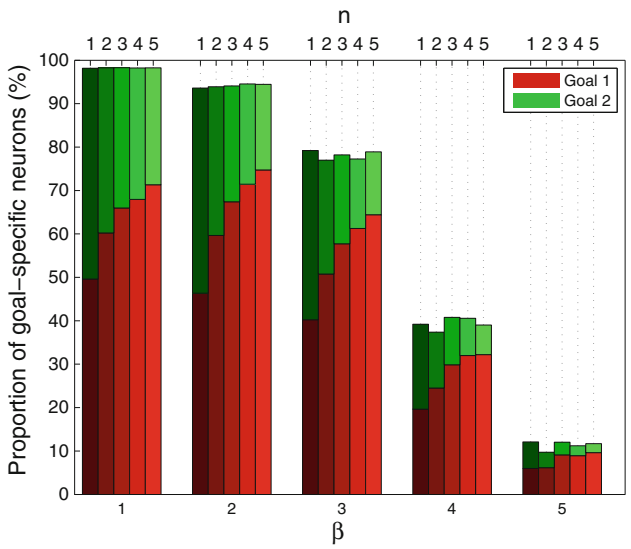

Fig. 5 Overrepresenting the first goal during training. *Stacked bars* show mean proportion of neurons representing the first and the second goal, respectively, for different values of n (*top axis*, as defined in Eq. 9) and β (*bottom axis*). Standard deviations are omitted for clarity. It can be seen that varying n has no effect on the overall proportion of goal-specific neurons but does affect their distribution over both goals

the input vector in this case is a concatenation of two vectors, the relative importance of the individual components for the organization of the SOM is a function of how much of the maximal distance between two input vectors (Eq. 6) can be explained by these components individually.

Since the β values of interest are larger than 1, the maximally possible distance between two input vectors composed of vectors encoding the same motion primitive but different contextual information is larger than the same distance between input vectors composed of vectors encoding the same contextual information but different motion primitives. This leads to the organization of the SOM into areas encoding a given motion primitive, as also shown previously [29].

Within the area encoding a certain motion primitive, whether or not goal-specific neurons emerge is then dependent on how necessary the contextual information is in representing the input data. In other words, the critical question is how much of the maximal distance between two input vectors can be explained solely by the maximal distance between the two motion-encoding components. In general terms, we can answer this by considering the ratio ρ_m between the diameter of the cluster from which a given motion primitive is sampled and the maximal distance between input vectors:

$$\rho_m = \frac{2r_m}{\max(d)} = \frac{1}{\sqrt{1+\frac{1}{\beta^2}}} \tag{10}$$

For the sake of completeness, we can also calculate ρ_c, which is the equivalent ratio for the contextual information cluster:

$$\rho_c = \frac{2r_c}{\max(d)} = \frac{2r_m}{\beta\max(d)} = \frac{1}{\sqrt{\beta^2+1}} \tag{11}$$

From Eq. 11, we find $\rho_m > 0.92$ for a β value of 2.5 (for which Fig. 2 begins to show a steady decrease in goal-specific neurons). Thus, goal-specific neurons begin to disappear when around 92% of the possible distance between two input vectors can be explained from the motion component alone.

As previously stated in the "Methods" section, the assumption in this paper is not that the mirror system can be equalled to a SOM, but merely that some of the principles that govern plasticity in a SOM may also apply to the mirror system. The effect of interest here is therefore the fundamental cause of the emergence of goal-specific neurons within the SOM. We have shown that this cause is the fact that the action encoding inputs can explain most of the possible variability in the overall input (leading to primitive-encoding areas) but not all of it (leading to goal-specific neurons within these areas).

Model Predictions

β Regulates the proportion of goal-specific neurons but not their organization An important question related to overrepresenting one goal in the training data is whether this affects the number of goal-specific neurons or merely their organization. This is tested, as before, using the nonparametric Friedman test on the newly generated data, and it is found that P has no significant effect ($df = 4$, $\chi^2 = 8.1$, $P > 0.08$, also seen in Fig. 5) on the percentage of goal-specific neurons while β, of course, still does ($df = 4$, $\chi^2 = 2{,}312.61$, $P = 0$). This is a rather interesting result since it indicates that, although β regulates the proportion of goal-specific neurons, it does not by itself determine how these neurons are then further organized. In the present model, the proportion of neurons and their use are therefore determined by separate parameters: The proportion is defined by characteristics of the input (specifically the difference in maximal variability of both components of the input vector), while the use of this proportion is determined during training. This leads to the prediction that, as the monkey learns a new goal to a level sufficient to cause its representation in parietal mirror neurons, some of the existing goal-specific neurons are reassigned to encoding this new goal (as opposed to new neurons being recruited to this effect). A possible way to verify this would be through the recording of parietal mirror neurons as a monkey learns a new goal (for instance, placing rather than eating [6]) for an action.

Differences in firing patterns when monkeys were executing actions (compared to observing them) may be caused by different input encodings Slightly different percentages of goal-specific neurons were reported, depending on whether the monkey was executing or observing the action [6]. It is hard to judge whether this difference is a real difference in encoding or merely an artifact due to the limited number of neurons that were measured (see Fig. 4), and more experimental data would be needed to resolve this. However, if it was confirmed that the difference is real, the cause would need to be investigated. The present model allows us to hypothesize that there is in fact a difference in the variability in input depending on whether an own movement or that by another agent is observed. Rather interestingly, if the difference is real, it would, in terms of the present model, indicate a smaller value of β for inputs resulting from the observation of the actions of others.

The model presented here predicts two possible causes for such a difference in β: (1) There may be a higher variability in the encoding of the proprioceptive information about one's own movement than in the encoding of observed motions and/or (2) the variability in the context-

encoding input might be smaller in the first case. Both of these predictions should be testable through brain imaging studies.

The first prediction implies that one's neural encoding of movements in STS during the observation of another's movement should show lower variability than the corresponding encoding within canonical neurons when executing the action oneself. Although not trivial to test, since it requires the ability to measure such a variability from neurophysiological recordings or perhaps imaging studies, the prediction should in principle be verifiable.

The most likely cause for a smaller variability in contextual encoding when executing own actions, on the other hand, is that the PFC may play a more prominent role in the case of own actions, for instance by providing accurate and certain knowledge of the goal of the action. One would therefore expect less variability in the contextual input (potentially accompanied by a weaker influence of the AIP) when executing own actions. Additionally, said input should remain similar for actions that have the same goal but may involve different objects when executing own actions (for example, coconuts and bananas are different objects that are both likely to turn up in actions whose goal is eating). On the other hand, observing actions would be accompanied by inputs, likely involving the AIP to a higher degree, which show higher variability during the observation of one action and between actions, particularly when the objects could also afford other actions with differing goals (for example, a coconut but not a banana could be used as a weapon).

The value of β may vary between actions Finally, a related observation (implicit in the model but not explored in detail here) is that it may be possible that β values are different for different actions (e.g., reaching vs grasping) as a consequence of the fact that β is defined based on the radius of a cluster within the input space representing *one* motion primitive. Clusters for different movements may have different sizes, which would affect β. This leads to the prediction that, if such a difference in encoding really exists in, for instance the STS (for observed actions), then neurophysiological recordings such as those presented by [6] should find different proportions of goal-specific vs goal-independent neurons for the same goal but different actions.

Discussion

A Developmental Model of Mirror Neuron Organization

The present paper has presented a model of a developmental process that can result in a SOM whose organization mimics that found in mirror neuron systems [6, 28]. The model builds on earlier approaches [4, 28, 29] that have addressed different aspects of the mirror system. It goes beyond the previous work by extending the modeled aspects to the origin of goal-specific firing in mirror neurons [6]. At the same time, it is a developmental model in the sense that the organization of the SOM only emerges after repeated exposure to the different inputs.

The model in the present paper thus presents the first developmental account of how a simple SOM can organize into a structure that reproduces features of (parietal) mirror systems to a remarkable degree. It additionally provides a grounded hypothesis of the reason that goal-specific neurons may exist in the first place. Specifically, it has been possible to show in this paper, through a systematic variation of model parameters, that the relation between geometric features of sensory inputs encoding observed or executed motion and inputs encoding an observed context is a key factor affecting not only the existence of goal-specific neurons but also their proportion.

In real-world terms, the values of β for which goal-specific neurons were found (i.e., values above 1) imply that a specific motion primitive can be encoded in more diverse ways than a specific context. It is important to note that this does not affect the number of "features" taken into account as the β effect has been proven to be independent of the dimensionality of the data. It merely implies, for example, that the variability in the encoding of "grasps" (executed or observed) is larger than the variability in the encoding of contexts that imply an overall "eating" goal.

A putative explanation for this difference in variability can be obtained by considering the possible encodings involved. Information about executed grasps is fed back via proprioception, while observed action information is relayed via the STS. Further, the action recognition hypothesis implies that the action is only identified within the mirror neurons. The information reaching these neurons is thus not likely to be a consistent abstraction representing the action, and one would therefore expect some variability in the encodings of different instances of, for instance, a grasp. This variability would be caused partly by the nature of the grasp (observed or executed) and partly by the details of the motion. Contextual information on the other hand, provided mainly through the AIP, has likely been heavily processed already [25, 32, 40], which may cause it to be more consistent between instances of the same goal.

It also is interesting to note that the model has been able to produce the goal-specific neurons without explicitly implementing a hypothesized cognitive function of the mirror neurons. That this was possible does not invalidate, for instance, the action-understanding hypothesis (which sees goal-specific neurons as important supporting evidence) but it does provide a developmental account in

which the emergence of such neurons is possible without a specific cognitive requirement driving the evolution. Rather, it appears plausible that goal-specific neurons are, in fact, merely picking up on specific properties of the encoding of information received by, for instance, the AIP and PFC as well as STS and canonical neurons. Thus, the cognitive ability to infer the goals of actions based on this mirror neuron activity, should such a mechanism exists, may have evolved on top of a mirror system organization which already "accidentally" produced apparent goal-specific firing patterns.

These findings further have implications in the design of artificial systems endowed with mirror systems. It is likely, as argued above, that the goal-specific mirror neurons play a significant supporting role in aspects of higher-level embodied cognition and social interactions. As such, it may be desirable to include such a mechanism in artificial systems, even though this has not been explicitly addressed by previous models. The important consequence of the work presented here is that this can be achieved simply through an appropriate modeling of information delivered to the mirror system rather than via an explicit mechanism to represent goals.

Relevance to Embodied Cognition

In the present paper, we have illustrated a modeling approach that, rather than tying the model to a specific embodiment, systematically modifies relevant parameters affecting the artificial sensory inputs received by the model. While the current model is disembodied, except in the minimal sense of being physically implemented on a computer, it can nonetheless provide insights to situated and embodied cognition by providing a model that is able to mimic aspects of the neurophysiological properties observed in mirror systems. In particular, it provides a meaningful approach to studying processes of abstraction from an embodied cognition perspective.

According to [44], most categorization is automatic and unconscious, and it is part of what makes up our conscious experience. Concepts, such as color concepts, are neural structures that allow us to mentally characterize categories and reason about them. Every conceptual structure is realized as a neural structure. The very structure of human reason, the way it is encoded in the underlying neural circuitry, comes from the properties of the embodiment and situatedness of humans: "Our abilities to move in the ways we do and to track the motion of other things give motion a major role in our conceptual system. The fact that we have muscles and use them to apply force in certain ways leads to the structure of our system of causal concepts" [44, p. 19].

Barsalou [45, 46] argued that several misunderstandings of embodied/grounded cognition have arisen because of a lack of computational models for these kind of cognitive processes, e.g., grounded theories being viewed as "recording systems that only capture images (e.g. cameras) and are unable to interpret these images conceptually" [45, p. 620]. The type of model put forward in this paper might provide a less problematic way of explaining how the sensorimotor system (especially within the premotor areas) organizes itself to represent more abstract aspects of an action, e.g., the goal of the action.

Although the current model does not explain how a system acquires conceptual content or tests the theories of Barsalou [45, 46], it provides a mechanism for developing a multimodal "representation" (e.g., goal-specific coding of movements), beyond simple recordings, in a self-organizing fashion in the course of agent–environment interaction. Two aspects of the model are of particular interest in the context of grounding embodied cognition. Firstly, the resulting goal-specific "representations" are the result of dynamic interactions between specific model inputs. Although the main point here was to investigate the β-effect rather than model the input provenance, these inputs provide, via the STS and AIP [25], both sensory and proprioceptive information. The "representation" of actions and their contexts are thus a joint property of the input characteristics, ultimately defined by sensory as well as motor information, and the organization of the SOM. Secondly, the model, although itself disembodied, suggests that the development of goal-specificity is dependent on the particular embodiment of the body part involved (in the sense that the nature of the limbs may influence the encoding of observed or executed actions involving it).

Further, as discussed above, the model is able to predict how simple features of affordance and observed/executed motion processing may give rise to the observed neural structure. In other words, rather than detailing which cognitive features arise from a specific embodiment, the model is able to suggest which embodiment is required for a specific feature to arise. If one subscribes to the idea that the organization of neural structures is an important aspect of the cognitive and behavioral capacities they underlie plus that the body and sensorimotor perceptions affect the neural organization, then an account such as the one presented in this paper is highly desirable as it provides a clear description of the required "format" of sensory inputs for a desired neural structure. In robotics, such knowledge can for instance be used in the design of a robot's sensorimotor mechanisms based on a specification of the desired cognitive capabilities.

Summary of Predictions

The model presented here has led to several predictions that need to be investigated in further work. Here, we summarize them briefly. We have shown that goal-specific neurons can emerge in a SOM if the inputs have certain properties (discussed in more detail above). The theoretical reason for this, discussed in "Goal-Specificity in the SOM and in Monkeys", was found to be that under these conditions, the motion-encoding inputs alone were almost, but not quite, sufficient for an adequate explanation of the distances (and thus similarities) between different inputs. Consequently, the predictions of the model center around the relative properties of the inputs to the mirror system in different conditions. These include the following: (1) Learning a new goal reassigns existing goal-specific neurons rather than recruiting new ones; (2) differences in percentages of goal-specific neurons when responding to executed and observed actions respectively can be due to either lower variability in input from the STS compared to that from canonical neurons and/or the involvement of the PFC in defining contextual information to a higher degree when executing own actions; and (3) pools of neurons encoding different actions may show different β values. Suggestions for testing these predictions were given in "Model Predictions".

In terms of mirror neuron research in general, the importance of the model presented here is the emphasis placed on the role that the model inputs have on the final organization within the SOM. This has led to the previously discussed predictions of how mirror neurons may be integrating the inputs they receive as well as how motion- and context-encoding inputs may relate to each other. Hypotheses regarding the roles that such mirror neurons might play in, for instance, action understanding [5, 6] need to take into account corresponding influences of brain areas such as STS, AIP, PFC and canonical neurons on the development of mirror neurons rather than solely considering the "finished product." For this reason, developmental models in general are useful in furthering our understanding of the emergence of the mirror system. The present work adds to previous models [4, 25, 30] by addressing the goal-specificity in parietal mirror neurons.

Conclusion

We have presented a computational model detailing a developmental process that results in pools of neurons in a SOM encoding specific motion primitives as postulated by the chain model [28]. We have shown that the neurons within these pools can self-organize/develop into goal-specific neurons as discovered by [6], and we were further able to detail the precise mechanisms underlying the emergence of such an organization and provide empirically testable predictions of the model. This work thus provides further support for the hypothesis that goal-specific mirror neurons (in parietal areas) may not have specifically evolved to support action understanding. Later cognitive processes may well have evolved to make use of the firing patterns which can facilitate action understanding, in what could be an instance of *neural reuse* [47]. A similar argument has been presented previously for the development of social roles of prefrontal mirror neurons [24]. Furthermore, the work presented here extends previous modeling work on premotor mirror neurons [30] and can be used in the design of robotic systems dealing with the initial processing of affordances and observed/executed motions in order to facilitate a natural emergence (rather than a hard-coded design) of a parietal mirror system complete with goal-specificity within the agent's controller.

Acknowledgments The authors thank India Morrison for helpful comments on the manuscript. ST and TZ are supported by the European Commission FP7 project ROSSI (*emergence of communication in RObots through Sensorimotor and Social Interaction*), Grant agreement no. 216125 (web: http://www.rossiproject.eu).

References

1. Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. Brain Res. 1996;119:593–609.
2. Wermter S, Weber C, Elshaw M. Associative neural models for biomimetic multimodal learning in a mirror-neuron based robot. Prog Neural Process. 2005;16:31–46.
3. Tani J, Ito M, Sugita Y. Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB. Neural Netw. 2004;17:1273–89.
4. Erlhagen W, Mukovskiy A, Chersi F, Bicho E. On the development of intention understanding for joint action tasks. In: Proceedings of the 6th IEEE international conference on development and learning. London: Imperial College; 2007. p. 140–5.
5. Rizzolatti G, Sinigaglia C. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. Nat Rev Neurosci. 2010;11(4):264–74.
6. Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G. Parietal lobe: from action organization to intention understanding. Science. 2005;308:662–7.
7. Cattaneo L, Fabbri-Destro M, Boria S, Pieraccini C, Monti A, Cossu G, Rizzolatti G. Impairment of actions chains in autism and its possible role in intention understanding. Proc Natl Acad Sci USA. 2007;104(45):17825–30.
8. Umiltà MA, Escola L, Intskirveli I, Grammont F, Rochat M, Caruana F, Jezzini A, Gallese V, Rizzolatti G. When pliers

become fingers in the monkey motor system. Proc Natl Acad Sci USA. 2008; 105(6):2209–13.
9. Bonini L, Rozzi S, Serventi FU, Simone L, Ferrari PF, Fogassi L. Ventral premotor and inferior parietal cortices make distinct contributions to action organization and intention understanding. Cereb Cortex. 2010;20:1372–85.
10. Hickok G. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. J Cogn Neurosci. 2008; 21(7):1229–43.
11. Oztop E, Kawato M, Arbib MA. Mirror neurons and imitation: a computationally guided review. Neural Netw. 2006;19:254–71.
12. Chersi F, Thill S, Ziemke T, Borghi AM. Sentence processing: linking language to motor chains. Front Neurorobot. 2010; 4(4). doi:10.3389/fnbot.2010.00004.
13. Arbib MA. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. Behav Brain Sci. 2005;28:105–67.
14. Rizzolatti G, Arbib MA. Language within our grasp. Trends Neurosci. 1998;21(5):188–94.
15. Gallese V, Keysers C, Rizzolatti G. A unifying view of the basis of social cognition. Trends Cogn Sci. 2004;8(9):396–403.
16. Wicker B, Keysers C, Plailly J, Royet J, Gallese V, Rizzolatti G. Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. Neuron. 2003;40(3):655–64.
17. Keysers C, Kaas JH, Gazzola V. Somatosensation in social perception. Nat Rev Neurosci. 2010;11:417–28.
18. Morrison I, Löken LS, Olausson H. The skin as a social organ. Exp Brain Res. 2010;204:305–14.
19. Wermter S, Weber C, Elshaw M, Panchev C, Erwin H, Pulvermüller F. Towards multimodal neural robot learning. Robot Auton Syst. 2004;47(2–3):171–5.
20. Inamura T, Toshima I, Tanie H, Nakamura Y. Embodied symbol emergence based on mimesis theory. Int J Robot Res. 2004; 23(4–5):363–77.
21. Wermter S, Elshaw M, Farrand S. A modular approach to self-organization of robot control based on language instruction. Connect Sci. 2003;15(2–3):73–94.
22. Demiris Y, Johnson M. Distributed, predictive perception of actions: a biologically inspired robotics architecture for imitation and learning. Connect Sci. 2003;15(4):231–43.
23. Haruno M, Wolpert D, Kawato M. MOSAIC model for sensorimotor learning and control. Neural Comput. 2001;13(10): 2201–20.
24. Oztop E, Arbib MA. Schema design and implementation of the grasp-related mirror neuron system. Biol Cybern. 2002;87(2): 116–40.
25. Bonaiuto J, Rosta E, Arbib MA. Extending the mirror neuron system model, I. Biol Cybern. 2007;96:9–38.
26. Bonaiuto J, Arbib MA. Extending the mirror neuron system model, II: what did i just do? A new role for mirror neurons. Biol Cybern. 2010;102(4):341–59.
27. Erlhagen W, Mukovski A, Bicho E. A dynamic model for action understanding and goal-directed imitation. Brain Res. 2006;1083: 174–88.
28. Chersi F, Mukovskiy A, Fogassi L, Ferrari PF, Erlhagen W. A model of intention understanding based on learned chains of motor acts in the parietal lobe. In: Proceedings of the 15th annual computational neuroscience meeting, Edinburgh, UK, 2006.
29. Thill S, Ziemke T. Learning new motion primitives in the mirror neuron system: a self-organising computational model. In: S Doncieux et al. editors. SAB 2010, LNAI 6226. Heidelberg: Springer; 2010. p. 413–23.
30. Metta G, Sandini G, Natale L, Craighero L, Fadiga L. Understanding mirror neurons: a bio-robotic approach. Interact Stud. 2006;7:197–232.
31. Kohonen T. Self-organizing maps. Heidelberg: Springer; 1997.
32. Caligiore D, Borghi AM, Parisi D, Baldassarre G. TRoPICALS: a computational embodied neuroscience model of compatibility effects. Psychol Rev. 2010;117:1188–228.
33. Kulić D, Nakamura Y. Scaffolding on-line segmentation of fully body human motion patterns. In: IEEE international conference on intelligent robots and systems. 2008. p. 2860–6.
34. Kulić D, Nakamura Y. Incremental learning and memory consolidation of whole body human motion primitives. Adapt Behav. 2009;17(6):484–507.
35. Rittscher J, Blake A, Hoogs A, Stein G. Mathematical modelling of animate and intentional motion. Phil Trans R Soc Lond B. 2003;358:475–90.
36. Yamashita Y, Tani J. Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment. PLoS Comput Biol. 2008;4(11):1–18.
37. Pomplun M, Matarić MJ. Evaluation metrics and results of human arm movement imitation. In: IEEE-RAS international conference on humanoid robotics, 2000.
38. Hemeren PE, Thill S. Deriving motion primitives through action segmentation. Front Psychol. 2011;1(243). doi:10.3389/fpsyg.2010.00243.
39. Şahin E, Çakmak M, Doğar MR, Uğur E, Üçoluk G. To afford or not to afford: a new formalization of affordances toward affordance-based robot control. Adapt Behav. 2007;15(4):447–72.
40. Fagg AH, Arbib MA. Modeling parietal-premotor interaction in primate control of grasping. Neural Netw. 1998;11:1277–303.
41. Jarque CM, Bera AK. A test for normality of observations and regression residuals. Int Stat Rev. 1987;55(2):163–72.
42. Corder GW, Foreman DI. Nonparametric statistics for non-statisticians: a step-by-step approach. New York: Wiley; 2009.
43. Bowman AW, Azzalini A. Applied smoothing techniques for data analysis. Oxford: Oxford University Press; 1997.
44. Lakoff G, Johnson M. Philosophy in the flesh: the embodied mind and its challenge to western thought. New York: Basic Books; 1999.
45. Barsalou LW. Grounded cognition. Annu Rev Psychol. 2008; 59(1):617–45.
46. Barsalou LW. Perceptual symbol systems. Behav Brain Sci. 1999;22(4):577–660.
47. Anderson ML. Neural reuse: a fundamental organizational principle of the brain. Behav Brain Sci. 2010;33(4):245–313.