Neural and Computational Mechanisms of Action Processing: Interaction between Visual and Motor Representations

Martin A. Giese[1,*] and Giacomo Rizzolatti[2,*]

[1]Section on Computational Sensomotorics, Hertie Institute for Clinical Brain Research & Center for Integrative Neuroscience, University Clinic Tübingen, Otfried-Müller Str. 25, 72076 Tübingen, Germany
[2]IIT Brain Center for Social and Motor Cognition, 43100, Parma, Italy; Dipartimento di Neuroscienze, Università di Parma, 43100 Parma, Italy

*Correspondence: martin.giese@uni-tuebingen.de, giacomo.rizzolatti@unipr.it

Abstract

Action recognition has received an enormous interest in the field of neuroscience over the last two decades, with a strong impact also in many other disciplines such as philosophy and robotics. In spite of this interest and impressive numbers of publications on this topic, the knowledge in terms of fundamental neural mechanisms that provide constraints for underlying computations remains rather limited. This fact stands in contrast with a wide variety of speculative theories about how action recognition might work, and how it might interact with other cognitive brain functions. This review focuses on new fundamental electrophysiological results in monkeys, which provide constraints for the detailed underlying computations, where we focus particularly on mirror mechanisms and interactions between visual and motor processing. In addition, we review models for action recognition with concrete mathematical implementations, as opposed to purely conceptual models. We think that only such implemented models can be meaningfully linked quantitatively to physiological data and have a potential to narrow down the many possible computational explanations for action recognition. In addition, only concrete

implementations allow to judge whether postulated computational solutions are feasible and can be implemented with real cortical neurons.

Introduction

Action recognition and its relationship to other cognitive functions has been one of the core topics in cognitive neuroscience over the last decade (Keysers, 2011; Keysers and Perrett, 2004; Rizzolatti and Fogassi, 2014; Rizzolatti et al., 2001; Schutz-Bosbach and Prinz, 2007). The discovery of *mirror neurons* in the premotor cortex of the monkey (Gallese et al., 1996; Rizzolatti et al., 1996) has initiated a wide interest in the neuroscience community for action processing and understanding, with implications in many other disciplines of neuroscience, including social neuroscience, motor control, body and self-representation, body motion perception, and emotion processing. At the same time, action processing and understanding in biological systems have become a topic of high interest in other disciplines outside neuroscience. This includes, for example, computer vision, robotics (e.g. Demiris, 2002; Schaal et al., 2003), and philosophy (e.g. Petit, 1999; Sinigaglia, 2013). In spite of the outstanding interest for this topic, the number of publications on the electrophysiological basis of action recognition that provide *precise constraints* for the underlying neural and computational mechanisms is still rather limited (cf. e.g. Kilner and Lemon, 2013).

This lack of strongly constraining data, combined with the vivid interest in the problem of action recognition and understanding motivated the development of a broad spectrum of, partly extremely speculative theoretical accounts of action processing. Many of these theories have never been concretely implemented, and have served only as frameworks for conceptual discussions. However, considering the complexity of the

underlying neural and dynamical processes and the high dimensionality of the underlying visual and motor patterns, the establishment of valid theories without the help of concretely implemented models is very difficult. Likewise, it is almost impossible to falsify such conceptual accounts by comparing them with specific experimental results in a conclusive manner.

Action-selective neurons are found in a number of brain structures, including the *superior temporal sulcus* (STS), the parietal, the premotor, and the motor cortex. We will briefly review here mainly the recent relevant results, focusing especially on a number of novel studies on mirror neurons. Much more detailed information about previous studies and other action-selective neurons without mirror properties can be found in other reviews (Nelissen et al., 2011; Puce and Perrett, 2003; Rizzolatti and Fogassi, 2014; Rizzolatti et al., 2001).

Early studies on mirror neurons (Gallese et al., 1996; Rizzolatti et al., 1996) reported that the responses of some mirror neurons to visual stimulus might depend on specific characteristics and modalities of the visual stimulation. However, only recent neurophysiological studies have studied these aspects systematically. They have investigated how spatial parameters of observed actions influence the activity of mirror neurons, including the distance of the action from the observer as well as the perspective or stimulus view, i.e. from which direction the action is observed. Moreover, these studies show that the mirror neuron discharge intensity is influenced by the value that is associated with objects on which the action is performed. We think that such parametrically well-controlled studies of the different aspects that influence the activity of action-selective neurons, and especially of mirror neurons, are absolutely essential for the development of solid computational theories of action perception in the primate cortex. In

addition, the work of Lemon and his co-workers (see Kraskov et al. 2014) showed that the mirror mechanism is not limited to parieto-frontal circuit, but also includes pyramidal tract neurons originating from areas F5 and F1 (primary motor cortex).

As a step towards a deepening of the understanding of the biologically-relevant neuro-computational mechanisms of action recognition, we provide also overview of the existing computational and neural models that are implemented in a sufficiently concrete manner to allow meaningful comparisons with such experimental data.

This overview of the existing work reveals several gaps in terms of critical experiments that might help to decide between different computational accounts, as well as between the available theoretical frameworks, all of which fail to capture some essential properties of the neural data. We hope that this analysis will help to set the goals for future research in experimental as well as in theoretical neuroscience.

Preliminary remark: different classes of actions

Before reviewing neurophysiological data on action recognition and discussing related models, it is important to stress that actions made by other individuals fall into *two main categories*. One category is constituted by actions that are present in the motor repertoire of the observers, and the other by actions that are extraneous to their motor abilities. The processing of these two different classes of actions involves partially different neural substrates. Both categories of actions activate visual action-selective areas located in the superior temporal sulcus (STS), while they differ with respect to the involvement of motor structures. To give an example, the observation of biting done by a dog, a monkey or a human being activates the same cortical parieto-frontal network in human observers. In

contrast, the observation of a dog that barks activates visual, but not motor areas (Buccino et al., 2004).

A psychological explanation of these finding has been proposed by Jeannerod. He suggested that "mere visual perception, without involvement of the motor system would only provide a description of the visible aspects of the movements of the agent, but it would not give precise information about the intrinsic components of the observed action which are critical for understanding what the action is about, what is its goal, and how to reproduce it". This implies that perception of actions without motor involvement is in some sense incomplete. Others have interpreted the motor activation triggered by others' actions in a more mechanistic way, suggesting that the motor activation of the parieto-frontal network results in a "direct recognition" of the observed action through the similarity between the observed and the executed action, not requiring additional complex inference processes (direct matching hypothesis) (Rizzolatti et al., 2014). A more recent interpretation is that motor activation during action observation represents a prediction triggered by the observed stimuli, which is necessary to disambiguate the sensory representations emerging during action observation (Kilner, 2011; Kilner et al., 2007; Wilson and Knoblich, 2005). Conceptually, this view minimizes, in part, the role of motor system in action processing, while stressing instead interactions between visual and motor areas for action understanding.

Another important distinction from a theoretical point of view is the one between *transitive* actions, which are directed towards goal objects, and *non-transitive* ones without such goal objects. It turns out (see Section 'Example-based visual recognition models') that the processing of transitive actions is computationally more difficult. It requires not only the recognition of the effector movement (e.g. the moving hand) but also a processing of

the relationship between the effector and the goal object (e.g. whether hand and object match spatially, or if the correct type of grip is applied to a specific object). This necessitates additional computational mechanisms that relate the movements of the effector to the properties of goal objects (e.g. Oztop et al., 2004).

Electrophysiological results

Due to space limitations the following review of electrophysiological results focuses on a few recently established novel aspects of mirror neurons, and properties of action-selective neurons that likely provide input to the classical mirror neurons system. An overview of the anatomy of the action observation system is given in Figure 1. With respect to a more elaborate treatment of previous results on the mirror neuron and action processing system we refer to several previous reviews (Puce and Perrett, 2003; Rizzolatti and Craighero, 2004; Rizzolatti and Fogassi, 2014).

Basic motor properties of canonical and mirror neurons

Area F5 contains two main types of neurons responding to visual stimuli: canonical neurons and mirror neurons. *Canonical neurons* are neurons that respond to the presentation of three-dimensional objects. Typically, there is congruence between the size of the objects that trigger the neuron and the type of grip encoded by that neuron (Murata et al., 1997). More recently, Fluet et al. (2010) recorded canonical neuron activity in monkeys, instructed by an external context cue to grasp a handle with a precision grip or a power grip. In addition, object orientation was varied. The neurons showed a context-dependent grasp planning activity after cue presentation, and a motor grasp-related activity during movement execution.

Contrasting with this class of neurons, mirror neurons are a specific set of neurons originally described in area F5 in the premotor cortex of the monkey. As all other types of neurons in area F5, mirror neurons discharge during *goal directed* actions such as grasping, holding, placing. Their main characteristic is that they respond to the observation of actions done by others. This property differentiates them not only from mere motor neurons, but also from canonical neurons. The relative proportion of these neuron types was investigated in a recent study in which a large number of neurons of F5 were recorded using multi-electrode linear arrays. The study reported that out of 479 recorded grasping neurons, 221 were purely motor, 197 mirror neurons, including 60 that also responded to object presentation, and, finally, 46 were canonical neurons (Bonini et al., 2014).

Mirror neurons are also present in monkey parietal areas connected with area F5 (see below). Their properties appear to be similar to those of mirror neurons in area F5. However, detailed comparative studies that assess possible differences between the functional properties of parietal and premotor mirror neurons have still to be undertaken.
In humans, mirror neurons were recorded in mesial motor areas and the hippocampus (Mukamel et al., 2010). The recordings were made in surgical patients with drug resistant epilepsy. The type of electrodes used (large linear electrodes with low impedance sites on the shaft and a bundle of tiny wires at the electrode end for single neuron recording) biased, inevitably, the single neurons database towards the medial part of the brain. A large body of evidence (including EEG and MEG, TMS, and brain imaging experiments) shows, however, that human parietal and premotor areas became active during action observation (see Rizzolatti and Fogassi, 2014). These areas closely correspond to those active in the monkey during action observation in fMRI experiments (Nelissen et al., 2011).

These are also the areas where mirror neurons were recorded. Thus, there is little doubt that the action execution/action observation circuit of humans houses mirror neurons.

Some newly established properties of mirror neurons in area F5

a) Influence of the observed action location relative to the observer

Early studies of mirror neurons were focused on demonstrating *congruence* between motor and visual responses of the recorded neurons (Gallese et al., 1996; Rizzolatti et al., 1996). Already in those studies it was reported, however, that mirror neurons form different subcategories according to the visual stimuli that are most effective in triggering them, and not all of them showed strong congruency between visual and motor tuning.

Different aspects of the visual tuning properties of mirror neurons were addressed in the last few years. One of these aspects was how the spatial location of the observed actions influences mirror neurons discharge (Caggiano et al., 2009). The results showed that the response of about half of the mirror neurons of area F5 discharged differently according to the location in space of the observed motor acts. Half of them discharged more strongly or exclusively to stimuli presented in the monkey peripersonal space, half preferred the extrapersonal space.

In the same study it was investigated whether space-selective neurons encode space in a "metric" or in an "operational" format. "Metric format" indicates that the location of effective stimuli was defined in terms of the true geometrical position or distance from the monkey. In contrast, space encoding in an "operational format" refers to the fact that the effective stimulus location is dependent on the possibility of the monkey to interact with the objects, and not on the true physical distance between the monkey and

observed action. The experiment by Caggiano et al. (2009) showed that about half of the tested space-selective mirror neurons were "operational mirror neurons" while the other half encoded the space in a metric way ("Cartesian mirror neurons").

To our knowledge, no computational or neural models exists that would capture these observed transformations of spatial tuning properties dependent on the operational space of the monkey.

b) Modulation of mirror neuron responses by the perspective view of observed actions

A very interesting recently investigated issue was whether mirror neurons provide information concerning the *perspective* from which the motor acts of others are observed (Caggiano et al., 2011). Three perspectives were tested: subjective view (0°), side view (90°) and frontal view (180°). The results showed that most tested mirror neurons (74%) were view-dependent, their responses being tuned either to one or, more frequently, to two specific points of view. Only a minority of the studies neurons (26%) exhibited view-independent responses, that is their response did not vary significantly with the perspective.

The observation of view-dependence fits nicely with example-based visual recognition mechanisms (see the Section 'Theoretical models'). However, it has to be noted that the same population of neurons can be simultaneously tuned to multiple parameters, e.g. to the view and different grip types. In addition, individual neurons can show different degrees of invariance with respect to these parameters (see also Singer and Sheinberg (2010)). This type of multi-dimensional tuning is not captured by most existing theoretical models, which typically make the simplifying assumption that individual

modules encode only a specific set of parameters instead of mixing many apparently unrelated computational functions.

There are two further issues that deserve some discussion here. The first is the origin of the input that may determine the properties of view-dependent and view-independent F5 mirror neurons. The second is what might be the functional role of these two types of mirror neurons. The main input to F5 arises from parietal areas PFG and AIP (antero interpatietal area) (see Figure 1). However, there is no detailed information available about the properties of these neurons in terms of their view-dependence properties. PFG and AIP receive input from various areas located in the superior temporal sulcus region. In this region neurons with view-dependent and view-independent properties have been described before (Perrett et al., 1985).

A common explanation for the computational function of view-tuned neuros in the visual pathway is that they represent an intermediate step towards view-invariant representations (e.g. Perrett and Oram, 1993). View-invariant neurons might pool the responses from view-variant ones with selectivity for different views. However, given that mirror neurons by definition have well-defined motor tuning properties, and thus are *motor* neurons, this explanation captures only a part of their possible computational role. It seems likely that such neurons combine information about the visual perspective of perceived actions with associated motor behavior.

An interesting possibility is, , that view-dependent mirror neurons might be helpful within an architecture that combines forward and backwards streams of information processing, in order to support feedback from motor representations to purely visual areas, e.g. via parietal cortex. Such modulation of bottom-up processing by an interpretation on more abstract levels higher up in the processing hierarchy has been repeatedly

conceptualized, e.g. in the context of Reverse-Hierarchy theory by (Ahissar and Hochstein, 2004) several years ago. Likewise, this idea forms a central element in theories in computer vision (Ullman, 1996) and plays a central role in predictive coding theories on action recognition (see Section 'Bayesian models'). More empirical data are needed, however, to confirm this appealing hypothesis.

c) Mirror neurons are sensitive for the value of an observed action

It was originally suggested that mirror neurons describe exclusively the goal of the observed action, and that their discharge is not influenced by the properties of the objects on which the action is performed, or by the value that this object may have for the monkey. A series of recent findings indicate that is not always true. In fact, a set of mirror neuron in area F5 has been observed whose discharge was modulated by the value that the grasped object had for the monkey (Caggiano et al., 2012). Two experiments demonstrated this point.

In the first, the discharge of mirror neurons during the observation of an agent, who was grasping food, was contrasted with that of the same agent grasping objects devoid of any meaning and value for the monkey. It was found that the large majority of tested neurons were more strongly activated in the "food" condition. In the second experiment, the responses of mirror neurons were studied in response to the observation of an agent grasping the same objects, which were either associated with a reward given to the monkey or were not rewarded. About 50% of the tested neurons responded more strongly when the observed motor acts were performed on rewarded objects, while a small percentage showed a stronger response for non-rewarded objects. Finally, the discharge of about 40% of neurons was not influenced by the reward conditions.

At first glance, the influence of the object value on mirror neuron responses is rather surprising. However, there is evidence (see below) that one of the nodes of the mirror system (parietal area AIP) receives information not only from the lower bank of the STS, but also from the inferotemporal lobe, a region that likely encodes the semantics of objects. Furthermore, neurons in both orbitofrontal cortex and cingulate sulcus are more strongly activated when the monkey anticipates a larger reward (Maunsell, 2004; Roesch and Olson, 2003, 2007). It is likely that areas which associate object and reward determine the value-related responses of mirror neurons through their output to the premotor areas.

This again demonstrates that cortical levels of processing cannot be easily mapped onto distinct computational steps, like action recognition, motor planning, or the decision between different alternative motor programs. Present theoretical frameworks do not provide a systematic approach to deal with such fuzzy assignments of computational functions to anatomical levels.

Action observation circuit: the input to the premotor cortex

The functional properties of STS neurons strongly suggest that these neurons provide the fundamental cortical visual input to mirror neurons. This hypothesis was recently confirmed by (Nelissen et al., 2011) using fMRI techniques, complemented by neuroanatomical tracing. Monkeys were presented with different types of hand grasping actions. Activations were found in three cortical regions: STS, inferior parietal lobule, and the premotor region. A subsequent analysis, carried out using as *region of interest (ROI)* the parietal cytoarchitectonically defined areas PF, PFG, PG and AIP, showed activation only in areas PFG and AIP. No action-specific activation was found in the other parietal areas. A subsequent connectivity study showed that the two parietal "mirror" areas are linked

with different sectors of STS. Area PFG is connected with the *upper* bank of STS, and in particular with area STPm. In contrast, AIP is mostly connected with the *lower* bank of STS, and in particular with its most rostral subdivisions (see Figure 1). Note that the temporal input to AIP originates not only from STS lower bank, but also from cortex that is part the inferotemporal lobe. These finding is of great interest because it indicates that the mirror network has access to information concerning object semantics. Such semantics defines, for example, classes of objects that are associated with the same type of grip.

Unlike mirror neurons the neurons in the STS do not have motor properties, but respond to visual action stimuli (Oram and Perrett, 1996; Perrett et al., 1989). Recent studies on the neural encoding of observed actions in the STS showed that this area contains many view-dependent neurons (Barraclough et al., 2009; Vangeneugden et al., 2011). This seems consistent with the idea that the view-dependence in mirror neurons might result from their afferent visual inputs. Finally, many STS neurons show temporal sequence selectivity and seem to associate the information of stimulus patterns over time (Barraclough et al., 2009; Singer and Sheinberg, 2010; Vangeneugden et al., 2011). Some of these neurons show tuning for actor identity (Singer and Sheinberg, 2010). In addition, the similarity of the neural activation patterns of STS neurons match closely the physical similarity between the encoded action patterns (Vangeneugden et al., 2009). STS neurons encode thus many aspects of actions regardless of whether those actions belong to the observer motor repertoire. They do not show, however, the motor properties that characterize premotor and parietal mirror neurons.

Action observation circuit: the mirror output form the premotor cortex

It is well known (Dum and Strick, 1991) that the hand representation of the primary motor cortex (areas M1 or F1) receives a strong input from area F5. However, early studies testing the mirror properties of neurons located in area F1 yielded negative results (Gallese et al., 1996). Recently, in a series of experiments on the mirror properties of the *cortico-spinal* tract neurons (Kraskov et al., 2009; Vigneswaran et al., 2013) demonstrated that many of these neurons respond to the observation of actions done by others.

A first study examined the activity of cortico-spinal neurons originating from area F5. They found that the discharge of about half of the tested neurons was modulated by grasping observation. Interestingly, the discharge rates of about 25% of these neurons were not increased, but rather suppressed during observation (Kraskov et al., 2009).

A second study investigated the responses of cortico-spinal neurons originating from area F1 (Vigneswaran et al., 2013) . About half of the tested neurons were modulated by action observation. Among these neurons, most increased their discharge rates during observation, while others reduced their discharge rates, or even stopped firing. A comparison between the properties of cortico-spinal F1 and F5 mirror neurons showed that the visual responses in F1 were much weaker than in F5. Thus, although many cortico-spinal F1 neurons fire during action observation, their input to spinal circuitry is weak and insufficient to produce movement.

These data are of great importance because they indicate that the understanding of goals of motor behaviour might not be simply a function of F5 mirror neurons, but rather is based on complex motor representations that involve even corticospinal tract neurons. This again indicates the weakness of the classical conceptualization of strictly hierarchical processing, here in terms of a separation between motor programming (in premotor

cortex) and processes of motor control that are associated with area F1 and the cortico-spinal tract.

Theoretical models with explicit mathematical implementations

While a wide spectrum of conceptual models for action processing exists, we focus here only on models with explicit mathematical implementations since we think that they will be most useful for narrowing down underlying computational mechanisms. In addition, space constraints do not allow us to extend the discussion to several interesting aspects that have been extensively discussed in the context of conceptual models. This includes: (i) the relationship between action processing, mirror neurons, and the representation of language (Arbib, 2005; Pulvermuller, 2005; Rizzolatti and Arbib, 1998); (ii) the issue of how mirror neurons emerge in terms of learning and evolution (Cook et al., 2014; Keysers and Perrett, 2004); (iii) the relationship between action processing and social cognition (Gallese et al., 2004; Rizzolatti and Sinigaglia, 2008; Spaulding, 2013); (iv) philosophical aspects, such as how mirror neurons are related to mind reading (Keysers and Gazzola, 2007), the awareness of self and others, or empathy (e.g. Oberman and Ramachandran, 2007; Rizzolatti and Sinigaglia, 2008). While such aspects are of broad general interest, the presently available neural data seems not sufficient to constrain mathematically implemented computational models on these aspects.

Before starting our review of existing models it seems important to discuss briefly the role of models in neuroscience, and specifically in the field of action recognition. There exists a heterogeneous spectrum of understandings about the function of theories in cognitive neuroscience, which ranges from quantitative, exactly-defined mathematical models (e.g. for biophysical processes in neurons, or about relationships between

psychophysical variables that can be accurately measured) up to conceptual post-hoc discussions and box-and-arrow models, summing up speculative claims with relevance for subsets of data. Since in action recognition already a vast number of speculative explanations exist we think that this field might profit more from theory concept that is similar to the one physics. According to this, a theory should link quantitatively different variables, for which one can specify an exact method how they are measured. This seems not to apply to a variety of popular concepts in the field of action perception (such as 'intention', 'empathy', 'mind', etc.), and for this reason we do not discuss them in this paper. The quantitative link to data might be made at different levels, e.g. at a behavioral level or even by explaining or predicting the behavior of individual neurons. Phenomenological models that relate different behavioral variables are important and often help to delineate fundamental computational problems and principles. Yet, usually they do not uniquely specify the neural mechanisms that implement such computations. For this, more detailed models including details about the processing in neurons or neuron populations are required. In addition, it is possible that certain computations cannot be efficiently implemented with real neurons. It is thus a nontrivial step to claim that a computationally efficient algorithm, e.g. in computer vision, really has a relevance for the brain. For this, at least at some point, one has to show how the postulated computations can be implemented by neurons with biophysically plausible properties, and it has to be verified if the resulting predicted behavior of neurons matches electrophysiological data. This conception of different levels of theory in neuroscience matches the classical distinction of levels of analysis that has been proposed by D. Marr (1982) for computational vision.

The existing theoretical approaches in the focus of this review fall in three main categories. The *first* tries to implement the "direct matching hypothesis" (Rizzolatti and Fogassi, 2014; Rizzolatti and Sinigaglia, 2010), thus the hypothesis that action recognition and understanding exploits motor-representations of the observed action (see above).

The *second* category tries to explain action recognition and understanding using predictive Bayesian models, assuming an interplay of sensory representations and their validation via top-down predictions from higher-level representations that encode underlying causes or motor states, including action goals.

The third class of approaches are visual pattern recognition models that accomplish action recognition by the identification of visual feature sequences, without making reference to the motor system. These approaches account for visual action recognition, but not for the observed interactions between vision and motor execution.

We try to focus here mainly on novel approaches and refer to classical models only to the degree that is necessary to understand the conceptual basis of the more recent approaches.-

Classical models based on artificial neural networks

The first implemented models for action processing and mirror systems were based on artificial neural networks. These models are important since they demonstrated that computational problems related to action processing and the mirror neuron system could be implemented and solved mathematically. However, these models typically do not claim that the proposed implementations reproduce detailed properties of cortical neurons. This puts these approaches between conceptual models and models that reproduce details of the physiology of the action processing system.

Among the first and most influential work using artificial neural networks are the seminal models by Arbib and co-workers (Bonaiuto et al., 2007; Oztop and Arbib, 2002). Using architectures that are coarsely inspired by the connectivity between the different parts of the action processing network (e.g. the STS, parietal areas such as AIP or IPS *(intraparietal sulcus),* or LIP *(lateral intraparietal cortex),* premotor areas such as F5, and primary motor cortex), and implementing individual computational modules using classical neural network techniques (including back-propagation), these models accomplish the recognition of grips and trajectory prediction. Action recognition is accomplished by the learning of mappings between visual features and features characterizing the goal object and its affordances (i.e. the way how it has to be manipulated), and between visual features and hand states, that characterize the hand configuration during grasping. The temporal sequence of hand states can then be exploited to recognize the action. The newer version of the model includes also an auditory pathway in order to model the multimodality of mirror neurons (Kohler et al., 2002), and it accounts for the fact that mirror neurons respond during partially occluded actions (Umilta et al., 2001). In addition, these architectures have been linked to controller models (Oztop et al., 2006). , and recently they have been extended by mechanism for monitoring of the possibility to execute actions and of the values of their outcomes (Bonaiuto and Arbib, 2010). A similar theoretical approach is also followed by models of other groups, such as the TROPICAL model (Caligiore et al., 2010), which exploits a variety of classical neural network techniques (including also Kohonen maps, neural fields, and hierarchical object recognition architectures) in order to model behavioral results on stimulus response compatibility. Others have applied classical neural network techniques to account for the problem how different action perspectives might be matched during the learning of mirror neurons (Schrodt et al., 2014).

Models based on controller architectures

Another class of classical models implementing the direct matching hypothesis has been derived from controller architectures that have been developed to account for motor contol. These behavioral models typically combine two types of dynamical internal models that model parts of the sensorimotor loop (Demiris and Khadhouri, 2006; Wolpert et al., 2003): 1) *Forward models* that compute predictions for state changes of the motor system and associated sensory signals from the motor command, where it is assumed that the brain sends a copy of the motor control signal as input to the forward model (reafference). 2) *Inverse models* that map sensory signals directly to appropriate control signals. This makes it possible to accomplish fast control in situations where the signals for feedback control are too slow. A prominent implementation of this approach is the MOSAIC architecture (Haruno et al., 2001), which combines multiple controllers for different behaviors (each consisting of a forward and inverse model) that are operating in parallel within a mixture of experts architecture (Figure 2A). The final control signal is determined by weighting of the outputs of the expert controllers, dependent on the sizes of their prediction errors with respect to the available sensory signals. As consequence, the outputs of the controllers that best predict the sensory inputs signals have the highest weights. It has been postulated that the MOSAIC model also explains the perception of actions and social behaviors, assuming that movements are classified by determining the controller module that produces the smallest prediction error (Wolpert et al., 2003). Controller-based models have been embedded in hierarchical architectures with a lower level that is formed by the expert controller modules for different actions and a top level that controls their contribution, allowing for the generation of action sequences (Haruno, 2003) .

Mirror neurons have been associated specifically with the implementation of the forward models in such control architectures (Oztop et al., 2006), and the parietal cortex has been proposed as being involved in the representation of the inverse models (Miall, 2003). In addition, it seems likely that subcortical structures, such as the cerebellum, are involved in the implementation of the relevant internal models (Caligiore et al., 2013).

Controller-based approaches have been very successful in accounting for behavioral data and have motivated many behavioral experiments. The way how such controllers are implemented in terms of cortical and subcortical circuits is not entirely clear. However, the idea of predictive control and of internal models that predict sensory consequences is dominating the present discussion about action encoding in cognitive neuroscience. The same concepts are presently frequently discussed in the context of predictive coding theories (see also Section 'Bayesian Models'). Another important concept is the idea of hierarchical representation of actions, which also has been postulated on the basis of human imaging data (Grafton and Hamilton, 2007).

A computational problem is that controller models often assume an internal simulation of motor programs in joint angle space, without specifying how such motor-relevant variables can be efficiently extracted from retinal image sequences. In robotics this difficult computational vision problem is typically bypassed by use of computer vision systems or special sensors that are not biologically plausible. Another problem for models that try to identify motor control policies from observation (Schaal2003) is that the dynamics of observed actors often does not match the one of the observer, for example because the observer has a different body geometry or masses of the body segments. This leads to a nontrivial *correspondence problem*, where one has to find a mapping between the

movements or control policies of agents with different physical properties (Dautenhahn, 2001).

Dynamic recurrent neural networks and neural field models

Another class of action processing models with closer relationships to brain functions is based on recurrent neural networks and neural fields. Neural fields are space-continuous recurrent neural network models that describe the dynamics of distributed activation patterns in the nervous system. Opposed to large-scale models with discrete neurons that in certain cases they permit a mathematical analysis and understanding of the emerging activity patterns. Recently such models are also often discussed under the term 'neural mass models' (e.g. Deco et al., 2011). They describe the dynamic variation of the average activity of ensembles of cortical neurons with similar tuning properties (mean-field approximation). This makes them suitable to establish links to detailed mechanisms at the level of real neuron ensembles. Neural fields have been used to account for action recognition and for mirror representations.

Influential models based on recurrent neural networks have proposed by Tani and colleagues (Tani et al., 2004). The networks are trained in a supervised manner with pairs of sensory and motor signals. The trained networks can predict trajectories from incomplete sensor information. This class of models has been extended towards dynamic hierarchical representations, where the top level represents the sequential order of actions, while the lower levels represent the trajectories of the individual actions. More recently, such models have been extended by inclusion of neuron pools with multiple time scales, forming a hierarchy (Yamashita and Tani, 2008). The models of the Tani group have been

tested extensively for movement generation and recognition in humanoid robots, showing that they scale to up for complex real-world problems. This shows the feasibility of action representation with such recurrent network architectures. The developed networks are not aiming at reproducing detailed properties of cortical neurons, while it appears that by appropriate modification of these models this might be possible. By linking perceptual and motor representations such models address mirror representations.

A second class of biologically-motivated dynamical network models is based on dynamic neural fields (Amari, 1977; Wilson and Cowan, 1972). Dynamic neural fields have been proposed as physiologically-inspired models for the distributed representation of motor programs, as well as for the self-organization of perceptual patterns, e.g. in low-level vision (Dayan and Abbott, 2001; Erlhagen and Schöner, 2002; Giese, 1999). For appropriate choice of the lateral connectivity, neural fields can have stable solutions that correspond to temporally propagating localized activity pulses. Such propagating pulse solutions can be used to model the sequential activation of neurons that encode different motor states, or instances along a trajectory. This mechanism has been proposed for the encoding of motor programs as well as for the encoding of perceived visual pattern sequences (Cisek and Kalaska, 2010; Giese and Poggio, 2003; Zhang, 1996). Neural fields have also been used to account for the interaction between movement recognition and action planning in robotics (Erlhagen et al., 2006; Sauser and Billard, 2006) For example, the STS, and cortical areas PF and F5 have been modelled by dynamically coupled neural fields with highly simplified inputs (Erlhagen et al., 2006). In addition, it was demonstrated that the required coupling between such neural fields can be learned with a Hebbian learning rule, providing a possible implementation for the hypothesis that mirror

circuits might result from Hebbian plasticity during self-observation (Keysers and Perrett, 2004).

Neural field models have been directly compared to neural data in motor and premotor cortex (e.g. Cisek and Kalaska, 2010; Erlhagen et al., 1999). A major limitation of the discussed neural field model for mirror representations is that they use highly simplified low-dimensional input and motor patterns. This leaves open whether the proposed architectures, and the associated nonlinear dynamics, scales up to perceptual and motor patterns with realistic dimensionality. (See below an application of neural fields for action perception from real videos.)

Closely related to neural field models are models for 'motor chains' (Chersi et al., 2011). These models are directly motivated by electrophysiological experiments in premotor and parietal cortex. They consist of ensembles of spiking neurons models that encode different phases of actions (e.g. reaching, grasping, and bringing to the mouth), and which are dynamically coupled in a way that results a sequential activation, representing the sequential order of the temporal phases of the action. This property reproduces the fact that neurons, e.g. in premotor cortex, often are tuned for individual grip phases and show activation only during the relevant phase. It is assumed that the ensembles receive input from motor as well as from visual structures, and that the first ensemble is excited by an external ensemble that encodes intention (potentially represented in prefrontal cortex). The activity of this external ensemble initiates the sequential activation of the chain of ensembles. This mechanism is thus very similar to the propagation of a localized activation pulse in a neural field (see above). The intention ensemble receives feedback input form the corresponding motor chain, resulting in an

activation of the corresponding intention representation when a motor behavior is observed or executed.

Chain models have a high degree of physiological plausibility and are thus suitable for a detailed comparison with real neural data. So far the models have been tested only with idealized low-dimensional peak-shaped input signals, which leaves the question open whether such models and their dynamics generalize to pattern spaces with realistic complexity.

Bayesian models

Another extremely popular class of models is based on Bayesian probabilistic inference. A first simple Bayesian model for action classification has been proposed in the context of a robotics system (Metta et al., 2006). It postulates that different parts of the mirror neuron system correspond to the components of a Bayesian action classifier and accomplishes classifications of hand actions from video input, using a computer-vision input model that is not biologically plausible.

Another very influential Bayesian action recognition model has been proposed by Friston and colleagues (Friston et al., 2011; Kilner et al., 2007), based on the idea of 'predictive coding' (Friston, 2010). This models picks up a variety of principles from the theories discussed before: (i) formulation of parts of the sensorimotor loop as predictive dynamical systems; (ii) the minimization of the prediction error in sensory space; (iii) use of nonlinear dynamical systems, and specifically of sequentially activated chains of neurons (called 'stable heteroclinic channels' in this literature) for the encoding of the sequential time structure of actions; (iv) dynamical hierarchies, specifying actions at different levels of abstraction, and with a bottom-up and top-down exchange of

information. These elements are combined within a brain theory, derived from machine learning and theoretical physics, which postulates that circuits in the brain realize a special form of belief-propagation algorithm. According to this interpretation, the brain estimates hidden variables (including internal dynamical states and intentions) based on sensory observations, where it exploits a probabilistic hierarchical dynamical generative model in order to specify how the sensory signals depend on external causes, and on the internal state variables of the brain. The parameters and variables of this probabilistic model are estimated using a Bayesian approach, combining prior distributions for the estimated variables with likelihoods. The likelihoods specify how the variables at individual hierarchy levels are statistically related to prediction errors for the variables in the next-lower level of the hierarchy. The likelihoods define thus a bottom-up stream of information within the hierarchy. In addition, it is assumed that the priors at the different levels are estimated by top-down predictions from the next higher level in the hierarchy (implementing a special form of 'empirical Bayes', where prior distributions are also estimated by maximizing consistency of the model with the data). This defines a top-down stream of information within the hierarchy. The approach makes it in principle possible to estimate all model parameters by minimizing the prediction error (more precisely the surprise or entropy) in the space of the sensory signals.

The underlying parameter estimation problem cannot be solved exactly because it becomes intractable, even for relatively low-dimensional problems (a frequent problem in Bayesian inference). This problem can be circumvented by minimizing not the real prediction error, but an upper bound that depends on it ('free energy'). This bound is formulated using an approximative distribution for the hidden variables that results in a tractable problem ('variational Bayesian inference'). A key assumption in the theory is that

this approximative distribution is Gaussian (which can be motivated by a 'Laplace approximation' for the underlying true distributions). With these assumptions it is possible to derive an algorithm that minimizes the free energy bound by a gradient descent. The gradient descent algorithm can be implemented as message passing procedure, where signals are exchanged within a network that consists of hierarchically connected 'nodes'. The idea is that these nodes can be mapped onto neurons or neuron ensembles. The proposed message passing procedure specifies exactly which signals are exchanged between the nodes, and how these signals are computed within the nodes (Friston, 2005).

It has been speculated that the mirror neuron system and action perception might be understood within this framework (Kilner et al., 2007). Action control and recognition using this idea have been implemented for simple examples with a two-degree of freedom arm (Friston et al., 2011; Friston et al., 2010). In addition, it has been demonstrated (for artificial simulated birdsongs) that the framework allows the learning of hierarchical models that represent dynamical signals at multiple time-scales (Kiebel et al., 2008; Kiebel et al., 2009).

The theory proposed by Friston clearly addresses mirror mechanisms since it specifies how sensory and motor representations are interacting, and since intermediate levels of the hierarchy combine visual and motor signals. Opposed to speculative frameworks, the implemented versions of the free-energy framework by Friston have the advantage that they make precise predictions about the exchanged neural signals and the computations in neurons ensembles that correspond to individual nodes. Such predictions can be tested by comparison with electrophysiological data. However, the detailed evaluation of such predictions is far from accomplished. Some aspects of the free energy

framework are also common to many other theories (e.g. hierarchy, asymmetry of bottom up and top-down connections, sequence encoding by neural state dynamics). Other aspects are more specific, and some predictions about the signal flow across cortical layers seem to match observations in real neurons (Bastos et al., 2012). Other aspects seem not to be in agreement with electrophysiological data, at least in action selective neurons. An example is the prediction that cortical neurons ensembles only encode unimodal distributions (because of the necessity to assume Gaussian distributions for the message passing algorithm) (Friston, 2008). A further issue is that it is not clear if for realistically complex pattern spaces the belief propagation iteration can be finished sufficiently fast in order to account for the observed rather low neural latencies. Action-selective neurons in area F5 can have latencies as short as 60 ms (Maranesi et al., 2014), and the average latencies are not much larger than 100 ms. This is not so much more than the sum of the synaptic delays from the retina to these neurons, leaving not much time for convergence of complex inference algorithms or optimization schemes. It has been discussed that a part of this convergence processes might happen during evolution. However, in order to go beyond speculation, it would be important to implement the proposed theory for patterns with realistic complexity and then to show that by pre-training one can build a system that can realize the remaining inference steps sufficiently fast. It remains thus an exciting and challenging topic to verify if the predictions of predictive coding account are really consistent with the properties of real cortical neurons.

Example-based visual recognition models

A further class of action recognition models takes a completely different approach from previously discussed theories, conceptualizing action recognition as a purely visual

pattern recognition process. Such approaches account also for action recognition in cases where the observer is lacking of motor representations of the observed behavior (e.g. observation of a flying bird, or a dog barking). These theories were motivated by learning-based object recognition models (review see Tarr and Bulthoff, 1998) that accomplish invariant object recognition by learning example views. Likewise, action recognition can be accomplished by learning to recognize sequences of visual patterns that are derived from retinal image sequences. The key for the efficiency of such systems is to accomplish invariance of the learned representations against parameters that are not relevant for the recognition (e.g. position, view, unimportant kinematic details, etc.). Almost all technical approaches for the robust visual detection and recognition of actions in computer vision and robotics are based on the learning of visual patterns (e.g. Moeslund et al., 2006), demonstrating the feasibility of this approach for real-world problems.

Example-based physiologically-inspired neural action recognition models have first been proposed for the recognition of non-transitive actions without goal objects. The developed models consist of hierarchies of feature detectors that mimic the properties of neurons at different levels of the visual pathway, from primary visual cortex, over intermediate levels such as area V4, up to the STS (Giese and Poggio, 2003; Lange and Lappe, 2006). Invariance is accomplished by pooling the responses from size-, position-, or view-specific detectors at higher levels of the hierarchy (e.g. Riesenhuber and Poggio, 1999). The recognition of actions can be accomplished either by recognizing sequences of body shapes, or by the recognition of sequences of associated optic flow patterns, where likely the brain integrates both feature types in recognition. Neural fields, implementing a predictive dynamics with a stable travelling pulse solution, have been proposed as physiologically plausible mechanism for the recognition of feature sequences (Giese and

Poggio, 2003), next to other schemes for spatio-temporal integration. While such example-based hierarchical visual recognition models have been originally developed as models for cortical processing, they have been extended to applications in computer vision, achieving competitive performance with state-of-the-art computer vision algorithms (Escobar and Kornprobst, 2008; Jhuang et al., 2007; Schindler et al., 2008).

Very recently, computer vision has developed a strong interest in such hierarchical neural network architectures, discussed under the labels 'deep learning architectures' or 'convolutional neural networks'. This interest was initiated by the fact that such architectures outperformed other algorithms, first for object recognition (LeCun et al., 2015), but later also for many other computer vision problems including action classification (Karpathy et al., 2014; Le et al., 2011). Such deep architectures are presently among the best performing solutions for action classifications from real videos in computer vision. Recently, also deep architectures including dynamical neurons which learn hierarchical spatio-temporal representations have been proposed (Jung et al., 2015). Many details of the existing deep architectures (applied filter kernel, training schemes, regularization by 'drop out', etc.) are not biologically plausible, so that it would be a superficial conclusion to link them directly to real cortical neurons. However, it has been shown that appropriately trained and constrained deep learning architectures can develop neurons during learning whose tuning properties resemble the ones of neurons in areas V4 and IT (*inferotemporal cortex*) (Yamins et al., 2014). In addition, new methods for the analysis of the tuning properties of neurons on the intermediate layers of such hierarchies have been developed that might become interesting for the analysis of neural data (Karpathy et al., 2015; Zeiler and Fergus, 2014). Finally, some recent work also tries to

develop a theoretical understanding why hierarchical (deep) architectures have so favorable generalization properties (Anselmi et al., 2015).

Recently, physiologically-inspired example-based recognition models have also been extended for the recognition of transitive goal-directed actions, such as the manipulation of objects with the hand. This requires a modeling of additional computations, potentially realized parietal areas, that analyze the shape and position of goal objects and the relationship between object and the effector movements. One model of this type reproduces a variety of electrophysiological results from action-selective neurons in the STS and premotor area F5, at the same time accomplishing recognition from real videos (Fleischer et al., 2013) (see Figure 2B). This model reproduces, for example, the view-dependence of action-selective neurons (see Section 'Some newly established properties of mirror neurons in area F5'), their temporal sequence-selectivity, and predicts are relationship between mirror neurons and mechanisms for causality perception (Fleischer et al., 2012). Other models of this type have been tested in robotics, accomplishing for example the recognition of grip apertures, affordances, or hand action classification (Prevete et al., 2008; Tessitore et al., 2010).

Since this class of model was derived, taking into account the tuning properties of cortical neurons, they make predictions about the behavior of individual neurons and motivate novel electrophysiological experiments. The fact that these models have been successfully tested with real videos provides evidence that they are computationally powerful. A shortcoming of most models of this type is that they have a pure feed forward architecture and largely do not include top-down effects. However, it has been shown that top-down connections can be added to such architectures, and under appropriate circumstances can improve recognition performance (Layher, 2013). Example-based-vision

models do not account for the well-established interactions between action vision and action execution (review Schutz-Bosbach and Prinz, 2007) since they do not include motor representations. However, it seems straight-forward to extend these architectures by layers that represent motor programs (e.g. using neural fields) and to implement dynamic couplings between neural vison-based and such motor representations in order to account for mirror properties. However, presently no electrophysiological data is available that would allow to constrain the exact form of this coupling.

Conclusions

This paper reviewed recent neurophysiological experiments on action recognition that define constraints for the development of future computational and neural theories of action recognition. Most importantly, the combination of anatomical and fMRI studies has established homologies between the action recognition system in humans and monkeys that justify comparisons between both species. In addition, this work has delineated the major pathways that are involved in action recognition, including specifically the STS, parietal cortex and premotor cortex.

Mirror representations with neurons that combine defined visual and motor tuning properties have been found specifically in parietal and premotor cortex, and to a less degree also in other structures, like the primary motor cortex. What is less clear and sets a challenge for future theory-experiment collaborations, is the characterization of the specific computational roles of these different regions. This problem is non-trivial because the correspondence between different necessary computational steps and these neural structures is typically fuzzy, and often neurons with similar tuning properties are found in

different regions. In addition, electrophysiological studies, trying systematically to establish the communality and differences between the properties of action-selective neurons, and especially mirror neurons, in different areas are largely lacking. Such studies will be essential to make substantiated assignments between different cortical areas and computational steps postulated in computational and neural models.

It is of interest to stress that recent research in neurophysiology shows a shift towards a much more quantitative characterization of the underlying representations, for example using movie stimuli with controlled timing, or specific parametric variations,. Such well-controlled parametric manipulations are an important step towards the generation of data that can be linked to detailed quantitative computational and neural models.

In addition, we tried to give an overview of the theoretical models that are implemented mathematically so that their behavior can be meaningfully compared with electrophysiological experiments. The comparison between physiological data and existing implemented models reveals several shortcomings of the existing theoretical frameworks, as well as of the available experimental approaches. Some of these shortcomings are listed in Box 2.

Coarsely speaking, the major limitation of the presently available theories is that they are either only computational, not giving details about the neural implementation of specific mechanisms, or they are limited to specific functions in action recognition (e.g. accounting only for the visual tuning of relevant neuron populations). In the field of visual object recognition, meanwhile, very developed neural theories exist that have been associated with detailed neural data at multiple levels.. The field of action recognition is much less developed in this regard. No physiologically plausible model that integrates visual and motor representations by biophysically plausible mechanisms exists. While the number of

hierarchical models for action classification (including deep architectures) is growing, it is less clear how to link such architectures to semantic aspects of actions that seem to be important for the brain (e.g. the matching between classes of objects and effector movements, or to different actions subserving the same goal). Also no systematic theoretical understanding exists how to control the learning processes in such architectures. A further problem is how cortical neural mechanisms for action recognition interacts with subcortical processes, e.g. in the cerebellum or oin the basal ganglia, and what are the computational advantages of this interaction. These problems set novel theoretical challenges, including the questions how the relevant computations can be implemented neurally.

Bridging the existing gaps between theory and experiments seems to necessitate new developments in experimental as well as in theoretical neuroscience. On the theoretical side, the clarification of the detailed neuro-computational mechanisms requires the development of theories that are close to real neurons. Purely computational models are not sufficiently constraining to verify such mechanisms in detail. This obviously makes it necessary to take into account and understand the mathematical details of such models. Just using theoretical approaches as black-box and speculating about how they might explain data will not be sufficient to find out whether they are really implemented by neurons. In addition, models have to be tested with realistic stimulus sets to verify their computational limits. Two decades of experience in machine learning and brain-inspired computing demonstrate that often apparently exciting algorithms did not scale up for real-world problems, which makes it unlikely that the brain exploits these algorithms, since it faces the same computational challenges as technical applications..

On the side of experimental research, more experiments need to be developed that test specific computational mechanisms and operations. Novel methodological approaches, such as techniques for the simultaneous recordings of large number of neurons and optogenetic approaches for the causal manipulation of the activity in specific classes of neurons likely will be helpful for this purpose. Such techniques might help to clarify central questions, like how the different types of invariance properties (e.g. with respect to view, action, actor, or the affordances of different objects) are jointly encoded in ensembles of action-selective neurons, and how different areas contribute to these invariances. Likewise, multi-unit recordings and simultaneous recordings in multiple relevant areas, and specific cortical layers, might help to clarify the information flow between layers. This will help to understand the role of different pathways in the action processing system (e.g. between the STS and area F5 (Nelissen et al., 2011)), and it will help to validate specific models, such as message passing accounts for bottom-up and top-down processing.

For the near future, it appears clearly feasible to further characterize the tuning properties of neurons (e.g. with respect to space, timing, and in terms of parameters characterizing the object-effector relationship) in the action processing network, using parametrically highly-controlled stimuli generated by computer graphics. In addition, machine learning techniques (including classifiers or visualization techniques for neurons at intermediate levels of hierarchies) might help to analyze the distributed representation of relevant parameters in populations of action-selective neurons. Likewise, it seems possible to test specific predictions about local computations according to predictive coding theories in comparison with electrophysiological data. Another interesting question, which might be addressed by pharmacological or optogenetic techniques in combination with theoretical modelling, is how critical different parts of the cortical action

processing network are for different tasks, and how cortical and subcortical structures interact during action recognition (Caligiore et al., 2013).

In terms of theoretical approaches, it seems feasible to extend existing hierarchical and deep-learning approaches by inclusion of layers that represent semantic aspects, such as the relationships between different classes of actions and goal objects. In addition, the realization of bottom up and top-down connections within such hierarchies is important. While predictive coding theory suggests one approach to implement such connections, there are many other possibilities, including ones directly derived from neural data. Such architectures will be suitable for the modeling of the interaction between action recognition and attention, and also to capture the observed modulation of neuron activity by reward expectations.

Summarizing, it appears that only progress along both lines, experiments and the development of new theoretical frameworks, will ultimately result in a satisfying understanding of the mechanisms of action processing. Once such a deeper understanding for the core problem of action processing is accomplished, one might start to explore the interesting question whether relevant computational principles also apply to other cognitive phenomena, as postulated in existing speculative accounts about the relevance of mirror mechanisms for emotion processing, empathy, or social cognition.

Acknowledgements

We thank D. Endres, S. Kiebel, and T. Poggio for helpful discussions and comments. We thank M. Angelovska for help with the editing. MG was supported by DFG GI 305/4-1,

DFG GZ: KA 1258/15-1, and by the European Union FP7-Flagship no 604102 (HBP), FP7-ICT-2013-10/ 611909 (Koroibot), FP7-PEOPLE-2011-ITN(Marie Curie) ABC PITN-GA-011-290011 (ABC), and H2020 ICT-23-2014 /644727 (CogIMon). Further support was provided by BMBF, FKZ: 01GQ1002A. GR was supported by the European Union under FP7-ERC-2009-AdG / 250013 (Cogsystems).

Box 1: What we know

- The recognition and understanding of actions of others that belong to the motor repertoire of the observer involves an activation of the observer's motor representations in addition to a visual analysis of the observed actions.
- The recognition and understanding of others' actions, which are not part of the motor repertoire of the observer, is based on visual analysis of the stimuli and subsequent inferential processes.
- Mirror representations of others' action are not limited to simple hand, mouth and leg actions but include complex actions (e.g. climbing) and emotional behaviors (e.g Abdollahi et al. (2013)).
- Visual processing of actions can be accomplished robustly by the learning of hierarchies of detectors with appropriate mechanisms for the encoding of invariance.
- Several computational approaches addressing mechanisms and relevant processes in action recognition and understanding are available.

Box 2: What we need to know / what is missing

- We do not have a clear understanding how computations are distributed between different cortical areas. (For example, are there different computational roles of mirror neurons in parietal, premotor and motor cortex?)
- How do cortical and subcortical structures interact in action recognition?
- A variety of models implements necessary computational steps by non-biological algorithms. Future work needs to focus on verifying implementations in terms of physiologically plausible mechanisms by detailed comparison with physiological data.
- Some theories make strong assumptions about available preprocessed inputs, bypassing non-trivial computational problems. Unless it can be found out how the brain solves these non-trivial computational problems, these approaches might not be relevant for the brain.
- Some theories have been implemented only for toy examples for highly simplified pattern spaces and ignoring constraints in terms of computation time. Valid theories need to scale up to sensory and motor patterns of realistic complexity and need to result in solutions with realistic computation times, taking into account the processing constraints of real neurons.
- Some models cover only partial aspects of action recognition and understanding, such as purely visual processing. These approaches have to be extended by working out possible links to motor and other related brain structures.
- While some of the discussed experimental results, such as view-tuning, activation of mirror neurons by occluded stimuli, etc., are captured by some of the existing

models, most of the discussed new experimental results, such as the modulation of activity by operational distance or expected reward, are not.

- The existing experimental results often are not appropriate to distinguish sufficiently between different possible theoretical explanations. Experiments have to be designed that aim at distinguishing different theoretical explanations (for example message passing vs. activity changes as predicted by recurrent neural network models).

.

Figure captions:

Figure 1. Action observation network: A) Lateral view of a macaque brain showing the locations of three region (STS, Intraparietal sulcus region- IPS, Inferior arcuate sulcus region- IAS) involved in action observation. B) Flattened representation of STS, IPS, and IAS. FEF (frontal eye fields). Visual information on observed actions is sent from STS through parietal cortex to area F5 along two functional routes indicated with red and blue arrows, respectively. Area 45B receives parietal input from LIP and also has direct connections with the lower bank of STS (green arrows). For further abbreviations, see text. (Modified from Nelissen et al. 2011).

Figure 2. Examples of models for the visual recognition of goal-directed actions that illustrate different theoretical principles. (A) Model based on a motor control architecture, such as the MOSAIC model. Different controllers are responsible for the different actions, such as walking and kicking. Forward models compute the predicted sensory signals from the corresponding motor commands. The control model with the smallest prediction error in the sensory domain determines the classified actions ('kicking') (modified from Wolpert et al., 2003). (B) Example-based visual recognition model for hand actions. The model consists of neural detectors that mimic properties of cortical neurons. It comprises three modules: (a) shape recognition hierarchy that recognizes hand and object shapes; (b) 'affordance module' that analyses the matching between grip type and objects shape and their spatial parameters; (c) recognition module that consists of neurons that are selective for goal-directed hand actions. View-independence is generated at the highest level of the hierarchy by pooling the output signals from view-specific modules (modified from Fleischer et al., 2013).

References

Abdollahi, R.O., Jastorff, J., and Orban, G.A. (2013). Common and segregated processing of observed actions in human SPL. Cerebral cortex 23, 2734-2753.

Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. Trends in cognitive sciences 8, 457-464.

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields, Vol 27.

Anselmi, F., Rosasco, L., Tan, C., and Poggio, T. (2015). Deep Convolutional Networks are Hierarchical Kernel Machines. arXiv preprint arXiv:1508.01084.

Arbib, M.A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. Behav Brain Sci 28, 105-124; discussion 125-167.

Barraclough, N.E., Keith, R.H., Xiao, D., Oram, M.W., and Perrett, D.I. (2009). Visual adaptation to goal-directed hand actions. Journal of cognitive neuroscience 21, 1806-1820.

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical microcircuits for predictive coding. Neuron 76, 695-711.

Bonaiuto, J., and Arbib, M.A. (2010). Extending the mirror neuron system model, II: what did I just do? A new role for mirror neurons. Biol Cybern 102, 341-359.

Bonaiuto, J., Rosta, E., and Arbib, M. (2007). Extending the mirror neuron system model, I - Audible actions and invisible grasps. Biol Cybern 96, 9-38.

Bonini, L., Maranesi, M., Livi, A., Fogassi, L., and Rizzolatti, G. (2014). Space-dependent representation of objects and other's action in monkey ventral premotor grasping neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience 34, 4108-4119.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., Porro, C.A., and Rizzolatti, G. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: an FMRI study. Journal of cognitive neuroscience 16, 114-126.

Caggiano, V., Fogassi, L., Rizzolatti, G., Casile, A., Giese, M.A., and Thier, P. (2012). Mirror neurons encode the subjective value of an observed action. Proc Natl Acad Sci U S A 109, 11848-11853.

Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J.K., Thier, P., Giese, M.A., and Casile, A. (2011). View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. Curr Biol 21, 144-148.

Caggiano, V., Fogassi, L., Rizzolatti, G., Thier, P., and Casile, A. (2009). Mirror neurons differentially encode the peripersonal and extrapersonal space of monkeys. Science 324, 403-406.

Caligiore, D., Borghi, A.M., Parisi, D., and Baldassarre, G. (2010). TRoPICALS: a computational embodied neuroscience model of compatibility effects. Psychological review 117, 1188-1228.

Caligiore, D., Pezzulo, G., Miall, R.C., and Baldassarre, G. (2013). The contribution of brain sub-cortical loops in the expression and acquisition of action understanding abilities. Neuroscience and biobehavioral reviews 37, 2504-2515.

Chersi, F., Ferrari, P.F., and Fogassi, L. (2011). Neuronal chains for actions in the parietal lobe: a computational model. PLoS One 6, e27652.

Cisek, P., and Kalaska, J.F. (2010). Neural Mechanisms for Interacting with a World Full of Action Choices. Annu Rev Neurosci 33, 269-298.

Cook, R., Bird, G., Catmur, C., Press, C., and Heyes, C. (2014). Mirror neurons: from origin to function. Behav Brain Sci 37, 177-192.
Dautenhahn, K., Nehaniv, C. L. (2001). Like Me? - Measures of Correspondence and Imitation' Cybernetics and Systems. Cybernetics and Systems 32, 11-51.
Dayan, P., and Abbott, L.F. (2001). Theoretical Neuroscience (Cambridge, MA: MIT Press).
Deco, G., Jirsa, V.K., and McIntosh, A.R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. Nature reviews Neuroscience 12, 43-56.
Demiris, J., Hayes, G.M. (2002). Imitation as a dual-route process featuring predictive and learning components: a biologically plausible computational model. . In Imitation in animals and artifacts, K. Dautenhahn, Nehaniv, C.L., ed. (Cambridge, MA, USA: MIT Press), pp. 327-361.
Demiris, Y., and Khadhouri, B. (2006). Hierarchical attentive multiple models for execution and recognition of actions. Robotics and Autonomous Systems, 361-369.
Dum, R.P., and Strick, P.L. (1991). Premotor areas: nodal points for parallel efferent systems involved in the central control of movement. In Motor Control: Concepts and Issues, D.R. Humphrey, Freund H.J., ed. (Chichester: Wiley), pp. 383–397.
Erlhagen, W., Bastian, A., Jancke, D., Riehle, A., and Schoner, G. (1999). The distribution of neuronal population activation (DPA) as a tool to study interaction and integration in cortical representations. Journal of neuroscience methods 94, 53-66.
Erlhagen, W., Mukovskiy, A., and Bicho, E. (2006). A dynamic model for action understanding and goal-directed imitation. Brain research 1083, 174-188.
Erlhagen, W., and Schöner, G. (2002). Dynamic field theory of movement preparation. Psychological review 109, 545-572.
Escobar, M.-J., and Kornprobst, P. (2008). Action Recognition with a Bio-inspired Feedforward Motion Processing Model: The Richness of Center-Surround Interactions. In Computer Vision - Eccv 2008, Pt Iv, Proceedings, D. Forsyth, P. Torr, and A. Zisserman, eds., pp. 186-199.
Fleischer, F., Caggiano, V., Thier, P., and Giese, M.A. (2013). Physiologically inspired model for the visual recognition of transitive hand actions. The Journal of neuroscience : the official journal of the Society for Neuroscience 33, 6563-6580.
Fleischer, F., Christensen, A., Caggiano, V., Thier, P., and Giese, M.A. (2012). Neural theory for the perception of causal actions. Psychological research 76, 476-493.
Fluet, M.C., Baumann, M.A., and Scherberger, H. (2010). Context-specific grasp movement representation in macaque ventral premotor cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 30, 15175-15184.
Friston, K. (2005). A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360, 815-836.
Friston, K. (2008). Hierarchical models in the brain. PLoS Comput Biol 4, e1000211.
Friston, K. (2010). The free-energy principle: a unified brain theory? Nat Rev Neurosci 11, 127-138.
Friston, K., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. Biol Cybern 104, 137-160.
Friston, K.J., Daunizeau, J., Kilner, J., and Kiebel, S.J. (2010). Action and behavior: a free-energy formulation. Biol Cybern 102, 227-260.
Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain : a journal of neurology 119 (Pt 2), 593-609.
Gallese, V., Keysers, C., and Rizzolatti, G. (2004). A unifying view of the basis of social cognition. Trends in cognitive sciences 8, 396-403.

Giese, M. (1999). Dynamic Neural Field Theory for Motion Perception (Dordrecht: Kluwer).
Giese, M.A., and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. Nature Reviews Neuroscience 4, 179-192.
Grafton, S.T., and Hamilton, A.F. (2007). Evidence for a distributed hierarchy of action representation in the brain. Human movement science 26, 590-616.
Haruno, M., Wolpert D.M., Kawato, M. (2003). Hierarchical MOSAIC for movement generation. International Congress Series 1250, 575-590.
Haruno, M., Wolpert, D.M., and Kawato, M. (2001). Mosaic model for sensorimotor learning and control. Neural computation 13, 2201-2220.
Jhuang, H., Serre, T., Wolf, L., Poggio, T., and Ieee (2007). A biologically inspired system for action recognition. In 2007 Ieee 11th International Conference on Computer Vision, Vols 1-6, pp. 1253-1260.
Jung, M., Hwang, J., and Tani, J. (2015). Self-Organization of Spatio-Temporal Hierarchy via Learning of Dynamic Visual Image Patterns on Action Sequences. PloS one 10, e0131214.
Karpathy, A., Johnson, J., and Li, F.-F. (2015). Visualizing and Understanding Recurrent Networks. arXiv arXiv:1506.02078.
Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., and Fei-Fei, L. (2014). Large-Scale Video Classification with Convolutional Neural Networks. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (IEEE Computer Society), pp. 1725-1732.
Keysers, C. (2011). The Empathic Brain. How the Discovery of Mirror Neurons Changes Our Understanding of Human Nature (Lexington: Ky. Social Brain Press).
Keysers, C., and Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. Trends in cognitive sciences 11, 194-196.
Keysers, C., and Perrett, D.I. (2004). Demystifying social cognition: a Hebbian perspective. Trends in cognitive sciences 8, 501-507.
Kiebel, S.J., Daunizeau, J., and Friston, K.J. (2008). A hierarchy of time-scales and the brain. PLoS Comput Biol 4, e1000209.
Kiebel, S.J., von Kriegstein, K., Daunizeau, J., and Friston, K.J. (2009). Recognizing sequences of sequences. PLoS Comput Biol 5, e1000464.
Kilner, J.M. (2011). More than one pathway to action understanding. Trends Cogn Sci 15, 352-357.
Kilner, J.M., Friston, K.J., and Frith, C.D. (2007). The mirror-neuron system: a Bayesian perspective. Neuroreport 18, 619-623.
Kilner, J.M., and Lemon, R.N. (2013). What we know currently about mirror neurons. Curr Biol 23, R1057-1062.
Kohler, E., Keysers, C., Umilta, M.A., Fogassi, L., Gallese, V., and Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. Science 297, 846-848.
Kraskov, A., Dancause, N., Quallo, M.M., Shepherd, S., and Lemon, R.N. (2009). Corticospinal neurons in macaque ventral premotor cortex with mirror properties: a potential mechanism for action suppression? Neuron 64, 922-930.
Lange, J., and Lappe, M. (2006). A model of biological motion perception from configural form cues. Journal of Neuroscience 26, 2894-2906.
Layher, G., Giese, M. A. & Neumann, H. (2013). Learning representations of animated motion sequences-a neural model. Top Cogn Sci 6, 170-182.

Le, Q.V., Zou, W.Y., Yeung, S.Y., and Ng, A.Y. (2011). Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (IEEE Computer Society), pp. 3361-3368.
LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature 521, 436-444.
Maranesi, M., Livi, A., Fogassi, L., Rizzolatti, G., and Bonini, L. (2014). Mirror neuron activation prior to action observation in a predictable context. The Journal of neuroscience : the official journal of the Society for Neuroscience 34, 14827-14832.
Marr, D. (1982). Vision. A Computational Investigation into the Human Representation and Processing of Visual Information (San Francisco,CA: W. H. Freeman and Company).
Maunsell, J.H. (2004). Neuronal representations of cognitive state: reward or attention? Trends in cognitive sciences 8, 261-265.
Metta, G., Sandini, G., Natale, L., Craighero, L., and Fadiga, L. (2006). Understanding mirror neurons - A bio-robotic approach. Interact Stud 7, 197-232.
Miall, R.C. (2003). Connecting mirror neurons and forward models. Neuroreport 14, 2135-2137.
Moeslund, T.B., Hilton, A., and Krüger:, V. (2006). A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding 104, 90-126
Mukamel, R., Ekstrom, A.D., Kaplan, J., Iacoboni, M., and Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. Curr Biol 20, 750-756.
Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., and Rizzolatti, G. (1997). Object representation in the ventral premotor cortex (area F5) of the monkey. Journal of neurophysiology 78, 2226-2230.
Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., Rizzolatti, G., and Orban, G.A. (2011). Action observation circuits in the macaque monkey cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 31, 3743-3756.
Oberman, L.M., and Ramachandran, V.S. (2007). The simulating social mind: the role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. Psychological bulletin 133, 310-327.
Oram, M.W., and Perrett, D.I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. Journal of neurophysiology 76, 109-129.
Oztop, E., and Arbib, M.A. (2002). Schema design and implementation of the grasp-related mirror neuron system. Biol Cybern 87, 116-140.
Oztop, E., Bradley, N.S., and Arbib, M.A. (2004). Infant grasp learning: a computational model. Experimental brain research 158, 480-503.
Oztop, E., Kawato, M., and Arbib, M. (2006). Mirror neurons and imitation: a computationally guided review. Neural networks : the official journal of the International Neural Network Society 19, 254-271.
Perrett, D.I., Harries, M.H., Bevan, R., Thomas, S., Benson, P.J., Mistlin, A.J., Chitty, A.J., Hietanen, J.K., and Ortega, J.E. (1989). Frameworks of analysis for the neural representation of animate objects and actions. J Exp Biol 146, 87-113.
Perrett, D.I., and Oram, M.W. (1993). Neurophysiology of shape processing. Image and Vision Computing 11, 317-333.
Perrett, D.I., Smith, P.A., Potter, D.D., Mistlin, A.J., Head, A.S., Milner, A.D., and Jeeves, M.A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. Proc R Soc Lond B Biol Sci 223, 293-317.

Petit, J.-L. (1999). Constitution by movement: Husserl in light of recent neurobiological findings. In Naturalizing Phenomenology, J. Petitot, F. Varela, B. Pachoud, and J.M. Roy, eds. (Stanford, CA: Stanford University Press), pp. 220-244.
Prevete, R., Tessitore, G., Santoro, M., and Catanzariti, E. (2008). A connectionist architecture for view-independent grip-aperture computation. Brain research 1225, 133-145.
Puce, A., and Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. Philosophical transactions of the Royal Society of London Series B, Biological sciences 358, 435-445.
Pulvermuller, F. (2005). Brain mechanisms linking language and action. Nature reviews Neuroscience 6, 576-582.
Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature neuroscience 2, 1019-1025.
Rizzolatti, G., and Arbib, M.A. (1998). Language within our grasp. Trends in neurosciences 21, 188-194.
Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., and Rozzi, S. (2014). Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. Physiological reviews 94, 655-706.
Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. Annu Rev Neurosci 27, 169-192.
Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. Brain research Cognitive brain research 3, 131-141.
Rizzolatti, G., and Fogassi, L. (2014). The mirror mechanism: recent findings and perspectives. Philos Trans R Soc Lond B Biol Sci 369, 20130420.
Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. Nature reviews Neuroscience 2, 661-670.
Rizzolatti, G., and Sinigaglia, C. (2008). Mirrors in the brain : how our minds share actions and emotions (Oxford ; New York: Oxford University Press).
Rizzolatti, G., and Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. Nature reviews Neuroscience 11, 264-274.
Roesch, M.R., and Olson, C.R. (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. Journal of neurophysiology 90, 1766-1789.
Roesch, M.R., and Olson, C.R. (2007). Neuronal activity related to anticipated reward in frontal cortex: does it represent value or reflect motivation? Annals of the New York Academy of Sciences 1121, 431-446.
Sauser, E.L., and Billard, A.G. (2006). Parallel and distributed neural models of the ideomotor principle: an investigation of imitative cortical pathways. Neural networks : the official journal of the International Neural Network Society 19, 285-298.
Schaal, S., Ijspeert, A., and Billard, A. (2003). Computational approaches to motor learning by imitation. Philos Trans R Soc Lond B Biol Sci 358, 537-547.
Schindler, K., Van Gool, L., and de Gelder, B. (2008). Recognizing emotions expressed by body pose: a biologically inspired neural model. Neural networks : the official journal of the International Neural Network Society 21, 1238-1246.
Schrodt, F., Layher, G., Neumann, H., and Butz, M.V. (2014). Modeling Perspective-Taking upon Observation of 3D Biological Motion. In Proceedings of the Joint IEEE International

Conferences on Development and Learning and Epigenetic Robotics (ICDL-Epirob) (Los Alamitos: IEEE), pp. 305-310.
Schutz-Bosbach, S., and Prinz, W. (2007). Perceptual resonance: action-induced modulation of perception. Trends in cognitive sciences 11, 349-355.
Singer, J.M., and Sheinberg, D.L. (2010). Temporal cortex neurons encode articulated actions as slow sequences of integrated poses. The Journal of neuroscience : the official journal of the Society for Neuroscience 30, 3133-3145.
Sinigaglia, C. (2013). What type of action understanding is subserved by mirror neurons? Neuroscience letters 540, 59-61.
Spaulding, S. (2013). Mirror Neurons and Social Cognition. Mind & Language 28, 233-257.
Tani, J., Ito, M., and Sugita, Y. (2004). Self-organization of distributedly represented multiple behavior schemata in a mirror system: reviews of robot experiments using RNNPB. Neural networks : the official journal of the International Neural Network Society 17, 1273-1289.
Tarr, M.J., and Bulthoff, H.H. (1998). Image-based object recognition in man, monkey and machine. Cognition 67, 1-20.
Tessitore, G., Prevete, R., Catanzariti, E., and Tamburrini, G. (2010). From motor to sensory processing in mirror neuron computational modelling. Biological cybernetics 103, 471-485.
Ullman, S. (1996). High-Level Vision.
Umilta, M.A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., and Rizzolatti, G. (2001). I know what you are doing. a neurophysiological study. Neuron 31, 155-165.
Vangeneugden, J., De Maziere, P.A., Van Hulle, M.M., Jaeggli, T., Van Gool, L., and Vogels, R. (2011). Distinct mechanisms for coding of visual actions in macaque temporal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 31, 385-401.
Vangeneugden, J., Pollick, F., and Vogels, R. (2009). Functional differentiation of macaque visual temporal cortical neurons using a parametric action space. Cerebral cortex (New York, NY : 1991) 19, 593-611.
Vigneswaran, G., Philipp, R., Lemon, R.N., and Kraskov, A. (2013). M1 corticospinal mirror neurons and their role in movement suppression during action observation. Current biology : CB 23, 236-243.
Wilson, H.R., and Cowan, J.D. (1972). Excitatory And Inhibitory Interactions In Localized Populations Of Model Neurons, Vol 12.
Wilson, M., and Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. Psychol Bull 131, 460-473.
Wolpert, D.M., Doya, K., and Kawato, M. (2003). A unifying computational framework for motor control and social interaction. Philos Trans R Soc Lond B Biol Sci 358, 593-602.
Yamashita, Y., and Tani, J. (2008). Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment. PLoS Comput Biol 4, e1000220.
Yamins, D.L., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., and DiCarlo, J.J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. Proceedings of the National Academy of Sciences of the United States of America 111, 8619-8624.
Zeiler, M., and Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In Computer Vision – ECCV 2014, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds. (Springer International Publishing), pp. 818-833.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. The Journal of neuroscience : the official journal of the Society for Neuroscience 16, 2112-2126.