

for the associative account, are necessarily relevant to the development of mirror mechanisms. This evidence is questionable because the training studies involve behavioral assessments which are inferred to reflect mirror mechanisms, but there is no direct evidence that the underlying mirror mechanisms are modified nor is there evidence that these observed short-term changes can translate into more permanent long-term effects. Also, at the neural level, there are two types of experiential learning (Greenough et al. 1987). One type is limited to sensitive periods during early development and is characterized by an overproduction of new synapses in anticipation of specific experiences that will contribute to the development of species typical behaviors, such as locomotion and language development (Bertenthal & Campos 1987). By contrast, experience-dependent processes are associated with the formation of new synapses that develop in response to unique experiences of the individual organism throughout development. If Cook et al. are correct that mirror mechanisms are exclusively a function of an inductive process involving sensorimotor learning, their development would correspond to an experience-dependent process. Currently, this claim is not defensible given that it is just as likely that the early behaviors associated with mirror mechanisms, such as imitation, are species-typical behaviors, and thus just as likely to be mediated by an experience-expectant process which predisposes infants to develop mirror mechanisms.

In conclusion, the target article raises legitimate reasons to question an extreme nativist position regarding the development of mirror mechanisms, but errs in the opposite direction by claiming a strong empiricist position. If a more probabilistic than pre-determined view of epigenesis is considered (Gottlieb 2007), it is difficult to imagine how a genetic predisposition could not contribute to the development of mirror mechanisms.

Associative learning is necessary but not sufficient for mirror neuron development

doi:10.1017/S0140525X13002227

James Bonaiuto

Division of Biology, California Institute of Technology, Pasadena, CA 91125.

bonaiuto@caltech.edu

<http://www.caltech.edu>

Abstract: Existing computational models of the mirror system demonstrate the additional circuitry needed for mirror neurons to display the range of properties that they exhibit. Such models emphasize the need for existing connectivity to form visuomotor associations, processing to reduce the space of possible inputs, and demonstrate the role neurons with mirror properties might play in monitoring one's own actions.

The primary hypothesis set forth by Cook et al. is that mirror neurons (MNs) are the result of generic associative learning processes, rather than the result of evolutionary selection for action understanding. They claim that the standard view of mirror neurons, what they call the “genetic account,” suggests that the predisposition to develop MNs is heritable and was selected for on the basis of their role in action understanding. However, in their characterization of the genetic account, Cook et al. do allow for the role of experience in shaping MNs. Computational models that simulate the development of MNs through experience show that this is possible through associative learning mechanisms, but that the connectivity to form these associations must already be in place and that this connectivity must be somewhat specialized for control of hand actions.

Cook et al. describe the “exaptation hypothesis” as claiming that MNs require a special kind of sensorimotor learning. However, a closer look at several of the computational models developed

under this hypothesis, such as the Mirror Neuron System (MNS) model (Oztop & Arbib 2002), reveals that they do in fact use standard learning algorithms completely compatible with the associative learning account. What makes these models work is the structure of their input representations and their connectivity. The pure associative learning account seems to assume that every neuron is either directly or indirectly connected with every other neuron in the brain. Such architecture would require significantly more trials of action and observation in order to correctly associate visual stimuli with the relevant motor representations.

The simplest version of the genetic account would predict that MNs would be found in different areas of the brain, depending on the unique history of each individual. This is not the case, at least in monkeys, and this seems to be due to a genetic influence on the patterns of connectivity expressed by each brain region. Indeed, as Cook et al. claim, there is a “wealth of the stimulus” – so much that the space of possible hand–object interaction representations in the visual and motor domains makes the associative learning account computationally intractable. What makes the “exaptation hypothesis” models able to handle such a space is the fact that the inputs are constrained to represent the hand–object relationships appropriate for performing manual actions. This is thought to occur throughout motor development as the infant learns to extract the relevant features from visual stimuli for controlling the hand relative to the object (Oztop et al. 2004). Once the inputs are restricted to those necessary to control transitive actions, “domain-general learning processes” can proceed to associate the visual representation with the motor program at various levels of abstraction.

Although the learning algorithm in the MNS model was compatible with the associative learning account, the network required extensive pre-processing of its input. Mirror neurons respond to observation of dynamic hand actions and therefore must process trajectories in the space of hand–object relationships. Mirror neurons will often respond to observation of a grasp before the hand contacts the object. In order to predict the outcome of a grasp before its completion, the MNS model transformed a temporal sequence of hand–object relations into a spatial pattern of neural activity for input to the network. A subsequent version of the MNS model, MNS2, discarded this preprocessing step by using a recurrent neural network and a modified learning algorithm to handle raw input sequences (Bonaiuto et al. 2007). These models show that although MNs may acquire their properties through associative-style learning processes, extra circuitry is required to perform the computations necessary for processing dynamic visual input from object-directed hand actions.

The MNS2 model additionally proposed that audiovisual MNs develop their auditory properties through simple associative learning. However, in this model, extra mechanisms such as working memory and dynamic remapping were required to handle the case where MNs correctly predict the outcome of a grasp when the final portion was obscured. It is not clear how these functions could be developed through pure associative learning.

Giese and Poggio (2003) present a model of visual tuning in the mirror system that is the most compatible with the associative learning account. This model currently does not include a learning mechanism, but it does address the existence of view-dependent and -independent mirror neurons and does not require reconstruction of the arm and hand shape. However, it still requires extensive processing to transform visual input into a reduced space such that it can be associated with motor signals.

The Augmented Competitive Queuing (ACQ) model embeds a network such as those in the MNS and MNS2 models in a larger network that learns self-actions (Bonaiuto & Arbib 2010). In this model, MN activity signals recognition of successful completion of one's own actions. Their output is used as an eligibility trace in reinforcement learning algorithms that modify the recognized action's desirability – how likely an action is to lead to a reward; and executability – how likely an action can be successfully

performed in the current context regardless of reward. This model shows how mirror systems can have evolved for the purposes of monitoring one's own actions and fit within a reinforcement learning framework for action selection.

A mechanistic model of MNs with random or full connectivity and pure associative learning has never been developed. Current computational models suggest that appropriate coarse-grained connectivity and input representations are required to make the space of possible hand-object relation trajectories tractable. While the associative learning account is compatible with these models at a first approximation, it does not offer any detailed explanations as to how networks of MNs acquire their properties in development and operate in the adult. Conceptual models such as the associative learning theory of mirror neuron origins which do not provide a proof of concept in the form of a computational model, are unconvincing.

More than associations: An ideomotor perspective on mirror neurons

doi:10.1017/S0140525X13002239

Marcel Brass and Paul S. Muhle-Karbe

Department of Experimental Psychology, Ghent University, 9000 Ghent, Belgium.

marcel.brass@ugent.be paulsimon.muhlekarbe@ugent.be
http://users.ugent.be/~mbrass/Web-Site/Dr_Marcel_Brass.html

Abstract: In this commentary, we propose an extension of the associative approach of mirror neurons, namely, ideomotor theory. Ideomotor theory assumes that actions are controlled by anticipatory representations of their sensory consequences. As we outline below, this extension is necessary to clarify a number of empirical observations that are difficult to explain from a purely associative perspective.

How often the battles of psychology have to be fought over again each time with heavier armies and bigger trains...

— William James (1890, *Principles of Psychology*, Vol. 2, p. 523, footnote)

Although William James wrote this sentence more than a hundred years ago in his ingenious paragraph on ideomotor action, it could have also been coined in the discussion about the origin and functional role of mirror neurons (MNs). One important contribution of the associative learning account outlined in the target article by Cook et al. has been to situate the finding of MNs in the historical context of psychological theorizing on the relationship of perception and action. Moreover, associative learning provides a powerful approach to explaining the ontogenesis of MNs based on general learning principles. However, a purely associative account of mirror neurons falls short in explaining a number of important findings regarding the modulation and control of the mirror system. In this comment, we therefore outline an extension of associative learning, namely, ideomotor theory that addresses several of these problems.

While the origins of ideomotor theory can be traced back to the beginning of the nineteenth century, the most prominent proponent of ideomotor theory was William James (1890). In its modern form (Greenwald 1970), ideomotor theory assumes, just like associative learning, that learning promotes the association of sensory and motor codes. However, ideomotor theory states that in the course of learning, additional ideomotor representations are formed that resemble anticipations of the to-be-produced sensory consequences of an action (see our Fig. 1, a–c, for a model of how ideomotor representations are formed).

According to ideomotor theory, these representations primarily serve a motor control function. We control our actions by anticipating their sensory consequences. Moreover, ideomotor theory



Figure 1 (Brass & Muhle-Karbe). Acquisition of an ideomotor representation (adapted from Greenwald 1970). (a) A stimulus (S) triggers a specific response (R) that leads to a sensory effect (E). (b) After learning, the stimulus will activate an anticipation (e) of the effect that precedes the response. (c) This anticipation (e) becomes conditioned to the response and allows for control of the response. (d) Priming by action observation: a stimulus that resembles the effect of the action (S_e) primes the ideomotor representation (e) which activates the response.

predicts a specific form of sensorimotor compatibility, namely, *ideomotor compatibility*. A stimulus that resembles the anticipation of a sensory action-effect activates the corresponding ideomotor representation (see Fig. 1d). For example, the image of another person opening their hand strongly overlaps with the representation that is used to control the hand-opening movement. Consequently, ideomotor-compatible stimuli can to some degree bypass response selection by directly activating motor programs (Brass et al. 2001).

As ideomotor representations are conceived as neither uniquely sensory nor motor, they should be localized in dedicated motor control structures that are distinct from primary sensory or motor areas. Such representations can be activated without necessarily leading to overt behavior and are thus likely used for motor planning and prediction (see also the commentary by Keysers, Perrett, and Gazzola for the idea of mirror neurons being involved in predictive coding). This property of ideomotor theory is consistent with human brain imaging studies showing an overlap of brain areas involved in action planning, movement observation, and motor imagery (Grezes & Decety 2001).

Another important consequence of ideomotor compatibility is that it can lead to self-other confusion. Because ideomotor-compatible stimuli directly activate representations that are used for motor control, confusion can arise between externally and intentionally triggered motor representations. Accordingly, controlling imitative behavior has been related to brain areas that are involved in the sense of agency and self-other distinction, and dissociated from brain areas involved in controlling interference from overlearned stimulus-response associations (Brass et al. 2003; 2005).

One crucial difference between ideomotor representations and simple stimulus-response associations relates to the underlying learning mechanisms. Ideomotor representations evolve from learning the relationship between responses and subsequent sensory effects (R-E learning). In contrast, classical associative learning theories, although concerned with action-outcome contingencies, primarily focus on learning the relationship of responses to those stimuli that precede them (S-R learning). Importantly, most experiments demonstrating that imitative response tendencies can be easily reversed use S-R learning paradigms rather than R-E learning paradigms (e.g., Catmur et al. 2007). From an ideomotor theory perspective, these findings may reflect that rapid learning strengthens the corresponding S-R associations to such a degree that they temporarily overrule existing ideomotor representations, leading to an advantage of ideomotor-incompatible over compatible mappings (Catmur et al. 2007).

Another difference between associative learning and ideomotor theory lies in their capacity to deal with specific forms of contextual modulation. Cook et al. outline how associative learning can explain the influence of contextual information on MN responses. In human studies, however, it has been demonstrated that the response of the mirror system is not only sensitive to contextual cues but also to high-level beliefs about the intentionality of the