In the target article, Cook et al. suggest that the evolutionary origins and maintenance of the mirror neuron system (MNS) lie in "domain-general processes of associative learning in the course of individual development, and, although they may have psychological functions, their psychological functions do not necessarily have a specific evolutionary purpose or adaptive function" (target article, Abstract). We agree that the excitement surrounding the discovery of mirror neurons (MNs) has led to an inordinate focus on their role in social-cognitive functions and how these functions might play a role in the evolution of the MNS. However, we strongly disagree with the authors' claims that the known social-cognitive roles the MNS plays in primate cognitive and behavioral functioning have not and do not affect the MNS in an evolutionary context and that the associative account "separates questions about their origin and function" (sect. 1, para. 4).

The target article describes a lower-level biological mechanism (associative learning) that, as Cook et al. argue, fully accounts for the phylogenetic and ontogenetic development of mirror neurons. We assert that the initial evolution, further evolution, and evolutionary maintenance of the MNS is likely jointly influenced by such lower-level biological mechanisms *and* by the well-documented role that the MNS plays in social-cognitive functions. One example of this joint influence can be observed in individuals with an intact versus an impaired MNS who are more able to attract reproductive partners, reproduce, and protect and provide for their offspring within the complex social structures of primate societies (e.g., Howlin & Moss 2012).

We agree that associative learning is likely a critical mechanism for both the development and the evolution of mirror neurons. However, given that associative learning mechanisms exist in species that do not have a MNS, some alternative mechanism *must* interact with associative learning in order to produce the evolutionary pressure required for the origin and maintenance of the MNS in humans. To avoid directly addressing the evolutionary advantages the social-cognitive functions of the MNS confer, Cook et al. use a "straw man" argument. They attack the most extreme proposal of the role of social-cognitive functions in the evolution of the MNS – evolutionary selection via action understanding. The "associative learning in vivo" and "evolutionary selection based upon action understanding" accounts represent polar extremes, both of which are unlikely to reflect reality. Simultaneously, however, the adaptive advantages of the social-cognitive capacities (e.g., action perception, processing, and prediction) ascribed to the MNS enhance individuals' reproductive fitness, creating precisely the evolutionary pressure that the authors propose has not, and does not, exist.

Cook et al.'s depiction of the role of developmental research in elucidating biological/genetic versus environmental/learning influences on the MNS is concerning. We agree that evidence for neonatal imitation is limited and, even if it is present, is unlikely to be driven by MNS mechanisms since cortical regions that contain MNs are not fully developed at birth. However, the postnatal developmental timeline of the MNS neither rules out genetic/biological and evolutionary processes nor demonstrates the role of associative learning. It is well known that frontal and association cortices that house MNs undergo striking synaptic development and myelination between 8-months and 3-years of age (Huttenlocher 2002; Imada et al. 2006; Locke et al. 1995). Developmental EEG evidence similarly indicates protracted cortical development in these regions (Hagne 1968; Southgate et al. 2009), with continuing maturation until late childhood or adolescence (Martineau & Cochin 2003). Therefore, biological factors may explain protracted MNS development.

Cook et al. also dismiss EEG mu suppression as an index of MNS functioning too quickly. The strong relationship between the mu rhythm and action observation/execution can be traced back to 1954, when Gastaut and Bert reported that the mu rhythm was consistently reduced when stationary subjects "identified themselves with an active person represented on a screen" (see also Pineda 2005). We also recently published a re-analysis of pooled data from four published studies, including a total of 66 individuals with autism spectrum disorders (ASD), demonstrating that, across the age-span from 6–17 years, there was significantly less mu suppression in individuals with ASD compared with matched controls during action observation, but not during self-movement (Oberman et al. 2013). Although source estimation indicates that the generator of the mu rhythm is in the postcentral gyrus rather than premotor or primary motor cortex (Hari & Salmelin 1997), the possible downstream modulation of motor cortex by the MNS is tangential to their mirror properties. Cook et al. also ignore recent studies showing that the same stimuli that elicit mu suppression also activate MN regions (as indicated by BOLD response; Perry & Bentin 2009) and modulate a TMS-induced motor evoked potential (Lepage et al. 2008), suggesting that all three indices are likely capturing the same underlying cortical mechanism.

In summary, we argue, contrary to Cook et al., that the origins and evolution of mirror neurons are unlikely to be driven by associative learning alone, but, rather, to be a consequence of a combination of evolutionary, biological, developmental, social-cognitive, and experience-based influences. Indeed, we speculate that the MNS is not functionally fixed, but rather a currently evolving, flexible, semi-modular neural network that interacts with multiple other neural systems, including the motor and social-motivation systems (Oberman et al. 2008). The functioning of such a system at any point in development should be viewed as a snapshot of a dynamic system that is constantly modulated by these influences and interactions with other systems (Johnson 2011; Johnson et al. 2002). Environmental and biological influences unfold simultaneously and interactively, not separately and sequentially, and their relative roles can only be disentangled with careful measurement and calculation (Dobkins et al. 2009; Smit et al. 2012). Cook and colleagues attack theories that argue for the evolution of the MNS based upon its proposed role in action understanding (Rizzolatti & Fadiga 1998; Rizzolatti et al. 1996), but we believe that the theory proposed by Cook et al. arguing that associative learning mechanisms alone can account for the origins and development of the MNS is equally as unlikely. Both models ignore the reciprocal relationships between evolutionarily adaptive psychological mechanisms and the lower-level biological mechanisms that are required for their existence.

Testing key predictions of the associative account of mirror neurons in humans using multivariate pattern analysis

doi:10.1017/S0140525X13002434

Nikolaas N. Oosterhof,[a,b,c] Alison J. Wiggett,[d] and Emily S. Cross[d,e]

[a]*Centro Interdipartimentale Mente/Cervello (CIMeC), University of Trento, 38068 Rovereto, Trento, Italy;* [b]*Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755;* [c]*Department of Psychology, Harvard University, Cambridge, MA 02138;* [d]*Wales Institute of Cognitive Neuroscience, School of Psychology, Bangor University, Bangor, Gwynne, LL57 2AS, United Kingdom;* [e]*Behavioural Science Institute and Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, 6525 EN Nijmegen, The Netherlands.*

nikolaas.oosterhof@unitn.it a.wiggett@bangor.ac.uk e.cross@bangor.ac.uk

http://www.unitn.it/en/cimec/22589/nikolaas-oosterhof

http://www.bangor.ac.uk/psychology/people/profiles/alison_wiggett.php.en

www.soba-lab.com

Abstract: Cook et al. overstate the evidence supporting their associative account of mirror neurons in humans: most studies do not address a key property, action-specificity that generalizes across the visual and motor domains. Multivariate pattern analysis (MVPA) of neuroimaging data can address this concern, and we illustrate how MVPA can be used to test key predictions of their account.

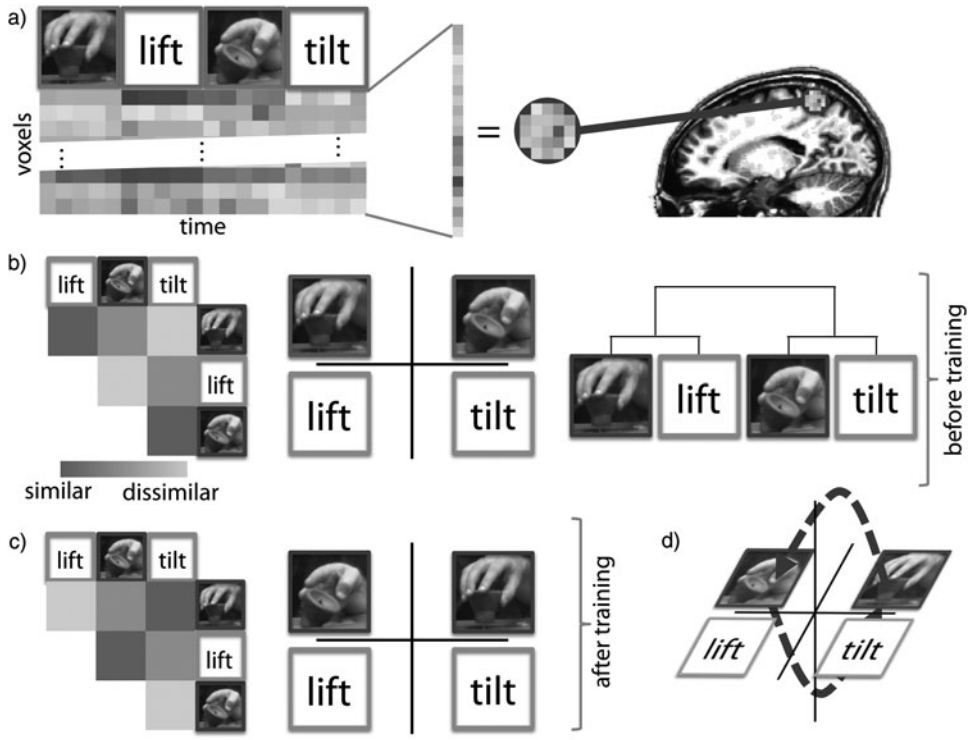

Figure 1 (Oosterhof). Hypothetical example illustrating neural effects of visuomotor "counter-mirror" training. (a) Neural responses are measured in a region of interest using fMRI while participants observe (red boxes) or are instructed to perform (green boxes) two actions (lifting or tilting a cup-shaped object). Each trial is associated with a spatially distributed pattern of responses over voxels. (b) Before training, congruent actions across the visual and motor domain are represented more similarly, as illustrated by a similarity matrix (higher pattern similarity indicated in dark orange; left), multi-dimensional scaling in two dimensions (more similar patterns depicted nearer in space; centre), and a dendrogram ("leaves" representing patterns are connected by shorter paths if patterns are more similar; right). (c) After counter-mirror training, where participants learn to lift (or tilt) the object after observing a tilt (or lift, respectively) action, the neural similarity structure might have changed, where non-matching actions are represented similarly across the visual and motor domain. (d) When visualized using multi-dimensional scaling, here in three dimensions, the similarity "trajectories" (dark purple arrows) during learning and unlearning of the associations can be visualized, allowing for assessment of the temporal dynamics of changes in visuomotor associations. A color version of this image is available at http://dx.doi.org/10.1017/S0140525X13002434.

The discovery of mirror neurons (MNs) in macaques has undoubtedly had a major impact on the field of social neuroscience by providing a potential mechanism for associating visual and motor aspects of actions, which could be a core component in action understanding. The timely article by Cook et al. addresses one of the fundamental questions on this topic, namely, how such visuomotor associations emerge in the first place.

While we agree with much in Cook et al.'s article, our concern is that some of the interpretations and conclusions the authors draw about mirror neurons in humans generally, and in favour of the associative account specifically, risk being overstated based on the evidence reviewed. Specifically, most inferences (but see Mukamel et al. 2010) about MNs in humans are based on a combination of proposed homologies between macaques and humans, and on studies employing less invasive methods to study the brain, including transcranial magnetic stimulation, magneto- and electroencephalography, and most notably, functional magnetic resonance imaging (fMRI). The former approach assumes a mapping between brain regions across monkey and human species, but this homology is imperfect (Sereno & Tootell 2005). The latter ostensibly provides a wealth of evidence for human MNs; for example, numerous fMRI studies have shown common areas demonstrating an increased response during observation and execution of actions (compared to a baseline condition), but such effects could be explained by task engagement or attention processes (Oosterhof et al. 2013).

In contrast, we would argue that stricter requirements are needed to infer that a brain region may contain mirror neurons. In macaques, MNs have been shown to code specific, individual manual actions that generalize across the visual and motor domains. In contrast, most fMRI studies with humans do not test for this key property: action-specificity. To demonstrate action-specificity in the human brain, the same action, whether observed or executed, should elicit more similar neural responses than dissimilar actions (Oosterhof et al. 2013).

In our view, the most promising approach addressing this limitation is the application of multivariate pattern analysis (Edelman et al. 1998; Haxby et al. 2001; Haynes & Rees 2005; Norman et al. 2006), which considers neural responses across a group of voxels. The logic behind this method is that the spatially distributed pattern, across voxels, of a specific observed (or executed) action should be more similar to the pattern associated with executing (or observing, respectively) the same action than a different action. Although fMRI cannot be used to measure individual neurons, MVPA complements single-cell recording approaches by considering spatially distributed responses at a system level. Importantly, this sensitive approach allows for dissociating responses of spatially overlapping neural populations (Peelen & Downing 2007).

Relevant to the topic of the present article, the human mirror system, recent MVPA studies have provided evidence for cross-modal action-specific representations of manual actions in anterior parietal and lateral occipito-temporal cortex (Oosterhof et al. 2010; 2012). These findings, together with evidence from single-cell recordings in humans (Mukamel et al. 2010), indicate that regions consistent with MN properties can also be found

outside the canonical fronto-parietal network, consistent with Cook et al.'s associative account.

Particularly relevant for the target article is that MVPA – in particular, representational similarity analysis (RSA; Kriegeskorte 2009; Kriegeskorte et al. 2008) – can be used to test key predictions, at a neural population level, of the associative account of the human mirror system. Cook et al. write "the properties of MNs can be changed in radical ways by relatively brief periods of sensorimotor experience" (sect. 3.4, para. 4), based on evidence from several behavioural studies showing that priming effects (e.g., automatic imitation effects) can be reduced or even reversed after counter-mirror training. Although we agree that these findings are interesting and consistent with an associative learning account, we do not believe that results from such studies alone provide strong evidence to support the authors' claim, as the neural correlates of these effects were not measured in fine detail. We believe that MVPA enables such detailed measurement and can be used to characterize where and when such changes occur at a neural population level.

To illustrate this, we provide a hypothetical example based on associative account predictions brought forward by Cook et al. of how counter-mirror learning can be adapted to study the changes in neural representations using MVPA (cf. Catmur et al. 2007; Press et al. 2007). Patterns of responses can be measured in the brain when participants observe or execute two different manual actions (Fig. 1a). Before training, observing and executing actions congruent across the visual and motor domain are represented similarly (Fig. 1b). After counter-mirror training, where actions incongruent across the visual and motor domain are associated with each other, this situation is reversed: manual movements incongruent across the visual and motor domain might now be represented similarly (Fig. 1c).

The application of MVPA provides other advantages. First, unlike TMS studies, MVPA does not require defining regions of interest a priori through the use of "searchlight" analyses (Kriegeskorte et al. 2006; Oosterhof et al. 2010; 2011). Second, MVPA enables the study of temporal dynamics of learning and unlearning new visuomotor associations of specific actions across the brain (Fig. 1d). Third, MVPA can be used to test generalization to other experimental factors such as viewpoint and different grasps. Fourth, MVPA allows for comparisons of neural representations across species (macaques and humans) and brain measurement methods (fMRI and neurophysiology), allowing for more detailed comparisons of (dis)similarities across species (Kriegeskorte 2009).

In conclusion, we agree that the existing evidence of a human mirror system is compatible with an associative learning account. However, we argue that the current evidence is not strong enough to fully support all the claims made by Cook et al. in the target article. We believe that the application of MVPA, in particular RSA and information mapping techniques, offers a promising avenue to more fully characterize the human mirror system, and thus provide evidence to support or falsify the associative learning hypothesis. We predict that these methods will be crucial for future fMRI studies if they are to advance our understanding of the human mirror system.

The mirror system in human and nonhuman primates

doi:10.1017/S0140525X13002446

Guy A. Orban

Department of Neuroscience, Parma University, 43125 Parma, Italy.

guy.orban@med.kuleuven.be

http://neuroscienze.unipr.it

Abstract: The description of the mirror neuron system provided by Cook et al. is incomplete for the macaque, and incorrect for humans. This is relevant to exaptation versus associative learning as the underlying mechanism generating mirror neurons, and to the sensorimotor learning as evidence for the authors' viewpoint. The proposed additional testing of the mirror system in rodents is unrealistic.

Cook et al., in the section on mirror neuron basics (sect. 2), provide an incomplete description of the mirror system in the macaque, ignoring some recent developments. Using monkey fMRI, Nelissen et al. (2005) visualized the frontal regions involved in action observation. This study is extremely important as it makes a point which has escaped most researchers in the human imaging field: In order to activate F5c, where mirror neurons (MNs) are housed, in fMRI, the actor has to be in view in the videos. Simply showing an isolated hand preforming an action is insufficient. Cook et al. also ignore the subsequent study (Nelissen et al. 2011), combining fMRI experiments with anatomical tracer studies. This later study revealed that the visual signals conveying action observation travel from the superior temporal sulcus (STS) to F5c over two parietal way stations: cytoarchitectonic PFG and the anterior intraparietal (AIP) area. Thus, this report indicates that, contrary to the assertion in section 3.2 (para. 2), STS neurons need not be linked to premotor neurons for the MNs to be generated, but rather, that visual and motor signals have to be associated in only a few parietal areas: PFG and AIP. This latter area, while not a classical mirror area, also houses MNs, according to fMRI (Nelissen et al. 2011) and unpublished single-cell results from studies conducted in Parma and Japan. Although this restricted association in parietal cortex is compatible with a genetic/evolutionary as well as an associative learning origin, it suggests that a domain-general mechanism such as associative learning may not be needed and that a more specialized hybrid mechanism, such as exaptation (sect. 8.2), might be more relevant than Cook et al. indicate. The sentence "there is no evidence that the sensorimotor learning involved in MN development is modified or constrained relative to the associative learning that occurs in standard conditioning experiments" (sect. 8.2, para. 1) may thus require serious revision.

The other development regarding monkeys concern two recent studies from the Lemon group showing how the mirror signal is "extinguished" by the addition of suppressed MNs among cortico-spinal neurons in F5 or M1 (Kraskov et al. 2009; Vigneswaran et al. 2013). Any account of how mirror neurons acquire their intriguing properties should take into account this transformation from purely excitatory to mixed excitatory/suppressed populations of MNs along the motor hierarchy. This point is again unaddressed, and it may well be that a genetic or hybrid mechanism can account for this range of responses more easily than an associative learning process. The data from monkeys concerning suppressed MNs also bear upon the interpretation of the Mukamel et al. (2010) data, suggesting that these human recordings may have been made in areas situated at a level of the motor hierarchy other than the planning level of the classical mirror areas (PFG/AIP and F5c).

In the section discussing mirror neuron basics, the authors also claim that the human mirror system is known. Ironically, the only mirror area for which the homology is known is the least documented in monkeys: AIP. There is indeed excellent evidence for the existence of this homologue, a combination of dorsal intraparietal sulcus anterior (DIPSA) and the so-called putative human AIP (phAIP) in anterior IPS (Durand et al. 2009). In contrast, most evidence for human areas cited by Cook et al. is weak or inexistent. Overlapping activations for action observation and execution are by no means proof for the existence of MNs, as the voxels contain thousands of neurons (Dinstein et al. 2008). Similarly, the repetition suppression studies have yielded contradictory results, unsurprisingly so, given that repetition suppression overestimates selectivity (Sawamura et al. 2006) and visual responses of premotor neurons do not adapt (Caggiano et al. 2013). That imitation constitutes an argument supporting the