Opinion piece

Author for correspondence:
Christian Keysers
e-mail: c.keysers@nin.knaw.nl

Royal Society Publishing

Hebbian learning and predictive mirror neurons for actions, sensations and emotions

Christian Keysers[1,2] and Valeria Gazzola[1,2]

[1]Netherlands Institute for Neuroscience, KNAW, Meibergdreef 47, 1105BA Amsterdam, The Netherlands
[2]Department of Neuroscience, University of Groningen, University Medical Center Groningen, Postbus 30.001, 9700 RB Groningen, The Netherlands

Spike-timing-dependent plasticity is considered the neurophysiological basis of Hebbian learning and has been shown to be sensitive to both contingency and contiguity between pre- and postsynaptic activity. Here, we will examine how applying this Hebbian learning rule to a system of interconnected neurons in the presence of direct or indirect re-afference (e.g. seeing/hearing one's own actions) predicts the emergence of mirror neurons with predictive properties. In this framework, we analyse how mirror neurons become a dynamic system that performs active inferences about the actions of others and allows joint actions despite sensorimotor delays. We explore how this system performs a projection of the self onto others, with egocentric biases to contribute to mind-reading. Finally, we argue that Hebbian learning predicts mirror-like neurons for sensations and emotions and review evidence for the presence of such vicarious activations outside the motor system.

1. Introduction

The discovery of mirror neurons provides neuroscientific evidence for what we call vicarious activations: the neural substrates of our own actions are vicariously activated while witnessing the actions of others through vision [1–4] or sound [3,4]. Twenty years after their discovery, the function of mirror neurons is still heatedly debated [5–9]. Here, we do not address the question of their function, but rather explore how they could develop. Monkeys have mirror neurons that respond to the sound and vision of crumpling a plastic bag [3,4] and human premotor cortices respond to sounds like the hiss of opening a Coca-Cola can [10]. Such selectivity is unlikely to be genetically preprogrammed. Here, we explore a mechanistic perspective of how such mirror neurons could emerge during development. We define what modern neuroscience understands by Hebbian learning based on spike-timing-dependent plasticity (STDP). We explore how this refined understanding of Hebbian learning helps us understand how mirror neurons emerge and suggests how mirror neurons become a form of active predictive mind reading. Finally, we argue that vicarious activations also occur in somatosensory and emotional cortices and that the same Hebbian learning rules could explain the emergence of mirror-like neurons in these brain regions.

2. What is meant by Hebbian learning

(a) Historically

The term Hebbian learning derives from the work of Donald Hebb [11], who proposed a neurophysiological account of learning and memory based on a simple principle: 'When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased' (p. 62). A careful reading of Hebb's principle reveals his understanding of the importance of causality and consistency. He writes not

that two neurons need to fire together to increase the efficiency of their connection but that one neuron needs to repeatedly (consistency) take part in firing (causality) the other. Carla Shatz (but not Hebb himself) has paraphrased his principle in a rhyme: 'what fires together, wires together' [12, p. 64]. While mnemonic, this summary bares the risk of obscuring the importance of causation in Hebb's actual work: if two neurons literally fire together, i.e. at the same time, the firing of one cannot cause that of the other. Temporal precedence, rather than simultaneity, is the signature of causality [13] and would indicate that 'one took part in firing the other'. This paraphrase should thus be read with a pinch of salt.

(b) Neurophysiological understanding

In the 1990s, neurophysiologists laid the foundation for our modern, neurophysiological understanding of Hebbian learning based on STDP [14–16]. Experiments in which two connected neurons were stimulated with various stimulus onset asynchronies evidenced an asymmetric window of STDP (figure 1). When an excitatory synapse connects onto an excitatory neuron, if the presynaptic neuron is stimulated 40 ms or less prior to the postsynaptic neuron, the synapse is potentiated. By contrast, if the presynaptic neuron is stimulated just after the postsynaptic neuron the synapse is depressed. If the two neurons simply fire together, the inevitable temporal jitter would make the presynaptic neuron sometimes fire just before and sometimes just after the postsynaptic neuron, and potentiation and depression would annul each other over time, leading to no substantial net STDP. As Hebb had predicted, causation is thus the key to synaptic plasticity.

Other experiments have refined our understanding of the consistency required for synaptic plasticity to take place. Bauer et al. [17] used a standard STDP protocol, with the presynaptic neuron stimulated 5–10 ms prior to the postsynaptic neuron. Applying 10 of these paired stimulations, they found strong potentiation of the synapse (figure 2a). Repeating the protocol, but intermixing unpaired stimulations in which only the postsynaptic neuron was stimulated cancelled the potentiation despite having applied the exact same 10 paired trials (figure 2b). This indicates that contingency is critical for STDP: in figure 2a, the presynaptic activity predicts the postsynaptic activity ($p(\text{post}|\text{pre}) = 1$, $p(\text{post}|\text{no pre}) = 0$), in figure 2b the presynaptic firing is not informative ($p(\text{post}|\text{pre}) = p(\text{post}|\text{no pre}) = 0.5$). This fleshes out what Hebb intuitively described as 'repeatedly and persistently takes part' and echoes the laws of associative learning [18]. Bauer et al. [17] then shifted the 10 unpaired stimulations to after the 10 paired ones and still found no potentiation (figure 2c). Delivering the 10 unpaired events 15 or 50 min after the paired events, however, no longer cancelled the STDP (figure 2d). Hence, the unpaired stimulations were integrated with the paired stimulations if they occurred in the 7 min window it took to apply 10 paired and 10 unpaired trials, but not if they occurred much later. STDP thus depends on both contiguity and contingency and uses a very narrow time window of approximately 40 ms to determine whether the presynaptic neuron took part in causing a particular postsynaptic action potential (contiguity) and a much longer, approximately 10 min, window to determine whether the presynaptic activity is informative about the postsynaptic activity (contingency). Whether the details of this contingency integration apply to all neurons or might be specific
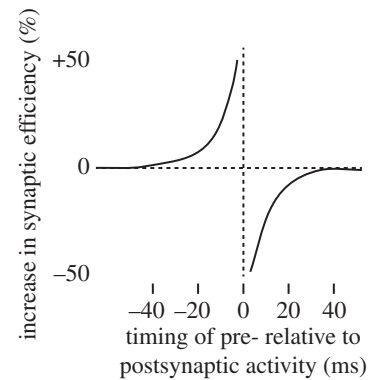

Figure 1. Temporal asymmetry of spike-timing-dependent plasticity.

to the lateral nucleus of the amygdala however remains to be investigated.

In the light of these findings, 'Hebbian learning' in contemporary neurophysiology refers to the rapidly expanding understanding of STDP [15,16] inspired by Hebb's work and emphasizes the sensitivity of STDP for tight temporal precedence (causality) and contingency over minutes. The perseverance of the term 'Hebbian learning' to refer to STDP honours the memory of a man who predicted how much of learning could be explained by such spike-timing-dependent plasticity. Here, we adhere to this use of Hebbian learning. The computational sciences use a similarly refined understanding of Hebbian learning, which also depends on contingency (http://lcn.epfl.ch/~gerstner/SPNM/node70.html).

(c) Alternative definitions

By contrast, in the psychological literature, some authors still equate Hebbian learning to the mnemonic approximation 'what fires together wires togethers'. We explore in particular this alternative definition used by Cooper et al. [19], as an example, because that paper tries to argue against Hebbian learning in the mirror neuron system and understanding the origin of the misunderstanding is important. They write: 'Hebb famously said that "Cells that fire together, wire together" and, more formally, "any two cells or systems of cells that are repeatedly active at the same time will tend to become 'associated,' so that activity in one facilitates activity in the other". Thus, Keysers and Perrett's Hebbian perspective implies that contiguity is sufficient for MNS development; that it does not also depend on contingency'.

We think there are a number of misunderstandings in this statement. First, Hebb himself never wrote 'Cells that fire together, wire together'. This mnemonic phrase was first introduced by Carla Shatz [12] in an article for the Scientific American aimed at lay public. Second, what is quoted as Hebb's formal postulate 'any two cells . . .', is not. Hebb used this sentence to summarize old ideas: he wrote 'The general idea is an old one, that any two cells . . .' [p. 70]. Both the mnemonic phrase misattributed to Hebb and Hebb's summary of old ideas occlude the causal element of Hebb's true postulate 'When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased' ([11], p. 62). As a result of these misunderstandings, Cooper et al.'s notion of Hebbian learning diverges from ours: theirs boils down to contiguity, while ours includes temporal
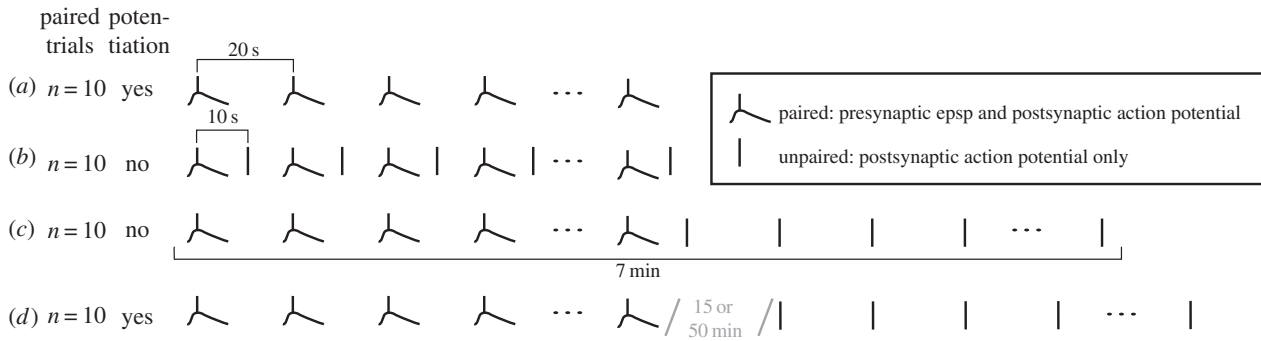
Figure 2. (*a*) Applying 10 paired pre- and postsynaptic stimulations leads to significant potentiation of the synapse. (*b*) Intermixing 10 unpaired, postsynaptic stimulations only cancels the potentiation. (*c*) Applying 10 unpaired stimulations after 10 paired also cancels potentiation. (*d*) Delaying the unpaired stimulations by 15 or 50 min preserves the potentiation of the 10 paired trials. The presynaptic stimulation is shown as a curve to represent the excitatory postsynaptic potential that arrives in the postsynaptic neuron, the postsynaptic stimulation as a vertical bar to represent an action potential. Adapted from Bauer *et al.* [17]; epsp, excitatory postsynaptic potential.

precedence, causality and contingency. Hebb himself might be the only one to exactly know whether he would have preferred our definition to that of Cooper *et al.*, but it is important to understand this divergence of definition to prevent doing what we believe Cooper *et al.* have done: use their own, contiguity-based definition, and apply it to *our* theory of the emergence of mirror neurons that is based on a different notion of Hebbian learning. Doing that, leads to a misunderstanding of our theory, and in this case to claims against our theory that are unwarranted.

3. Hebbian learning and mirror neurons: a macro-temporal perspective

Mirror neurons exist at least in the monkey's ventral premotor (PM; area F5, [2–4,20]) and inferior posterior parietal (area PF/PFG, [21]) cortex. Neurons in these two regions are *reciprocally* connected [22]: PF/PFG sends information to PM and PM back to PF/PFG. Neurons in area PF/PFG are also reciprocally connected with those in the superior temporal sulcus (STS [22,23]), a region known to respond to the sight of body movements, faces and the sound of actions [24]. Other brain regions contain mirror neurons as well [25–27] but to illustrate how the Hebbian learning account of the emergence of mirror neurons *could in principle* explain the emergence of mirror neurons a simple system encompassing only two brain regions, STS and PM, together with reciprocal connections from STS to PM and from PM to STS suffices. In this section, we will adopt a relatively coarse temporal resolution of about 1 s for the first approximation of the Hebbian account of how mirror neurons could arise. At this level of description, Hebbian learning makes predictions at the neural level that are similar to those that associative sequence learning—a cognitive model initially developed to describe the emergence of imitation [28]—makes at the functional level. The original papers explaining Hebbian learning at this temporal resolution are those of Keysers & Perrett and Del Giudice *et al.* [24,29], those describing associative sequence learning include Heyes, Brass & Heyes and Cook *et al.* [28,30,31]. In §4, we then look at a finer time-scale to reveal how mirror neurons could organize into a dynamic system that generates active inferences.

(a) Re-afference as a training signal

In the newborn human and monkey babies, we know little about the selectivity of the relevant STS and PM neurons and their connections. Accordingly, we will assume relatively random bidirectional connections between neurons in the STS that respond to the vision and sound of different actions and neurons in PM that code for the execution of similar actions. These connections go via the posterior parietal lobe (in particular PF/PFG), but for simplicity's sake, we do not explicitly mention this mediating step.

When an individual performs a new hand action, he sees and hears himself perform this action. This sensory input resulting from one's own action is called 're-afference'. The universal tendency of typically developing babies to stare at their own hands ensures that such re-afference will occur often when baby performs new movements [32]. As a result, activity in PM neurons triggering a specific action, and activity in neurons responding to the sound and vision of this specific action in the STS would, to the first approximation (but see §4), consistently and repeatedly overlap in time. For instance, a grasping neuron in STS will have firing that will consistently overlap in time with the activity of PM grasping neurons while the individual observes himself grasp. Throwing STS neurons, on the other hand, will have firing that consistently overlaps in time with that of throwing PM neurons while the individual observes himself throw. By contrast, the firing of STS grasping neurons will not systematically overlap in time with that of PM throwing neurons and vice versa. Accordingly, re-afference will create a situation in which the firing of STS and PM neurons for the same action will overlap more systematically than those for two different actions. There is a rough contiguity (firing at about the same time) and contingency (e.g. p(sight of grasping|grasping execution) > p(sight of throwing|grasping execution)). At this macroscopic time-scale, the synapses connecting STS and PM representations of the same action should be potentiated based on the understanding of Hebbian learning outlined above, while those that represent different actions should be weakened.

(b) Re-afference should favour matching connections

We hypothesize that after repeated re-afference and the Hebbian learning that it will cause the prevalent STS-PM connections should be matching (i.e. connect representations of similar actions). This is based on the largely untested assumption that over a person's life the statistical relationship between a person's actions and the sensory input are such that the criteria of Hebbian learning should primarily create matching synaptic connections. For the case of direct auditory or visual re-afference, this is trivial, as the sound and vision of our own actions always match our actions.

Some actions are, however, perceptually opaque. A classic example is the case of facial expressions. One might argue that we are actually born with mirror neurons for facial expressions, based on evidence that newborn babies are more likely to produce certain facial expressions when they see others do so, before learning can have created that phenomenon [33]. The exact extent to which newborns can imitate facial expressions is a matter of debate. There is robust evidence that at least tongue protrusion is imitated by newborns [34,35], but there is less evidence that any other facial expressions are robustly imitated [31,36,37]. We [6,24,29] and others [31] have argued that indirect re-afference might thus provide the kind of contiguous and contingent signals necessary to train matching connections between STS neurons responding to the sight of facial expressions and the motor programmes for performing them. Parents imitate the facial expressions of their babies, and babies experience numerous instances of imitation in their face-to-face interactions with their parents [37]. A baby would thus often experience the indirect re-afference of seeing/hearing his facial expressions being imitated, causing matching Hebbian associations. In this sense, we propose that our genetic make-up might facilitate the development of mirror neurons for facial expressions, but not (or at least not generally) by pre-wiring the STS-PM connections to have newborns equipped with mature mirror-neurons for facial expressions, but by equipping babies with a tendency to stare at the face of their parents, and parents with a tendency to imitate their babies facial expressions [29]. Ultimately, the development of mirror neurons for facial expressions then still depends on learning during the lifetime, but this would be canalized by these behavioural predispositions. The effect of being imitated on Hebbian learning would probably be less rapid than direct re-afference, given that people's imitation of our facial expressions will be more variable in time and visual properties. In our modern world, physical mirrors could also contribute to creating re-afference in the case of facial expressions, and this re-afference would be particularly suitable for Hebbian learning, but it is unclear how much of a role these physical mirrors play in typical development.

Re-afference need however not be visual. Because babies hear themselves cry and laugh, auditory mirror neurons for these emotional sounds could emerge robustly even when deprived of visual parental imitation. During babbling, baby also creates contingencies in the firing of premotor neurons triggering the pseudo-speech and neurons in the temporal lobe responding to such speech. Once the synaptic connections have been trained by its own babbling, hearing a parent speak could trigger the motor programmes to replicate the words [6]. This process would be assisted by the fact that parents change the tone of their own speech to be more similar to that of the baby (motherese [38]). Here, the cross-cultural tendency of parents to motherese and the tendency of babies to babble would canalize the emergence of appropriate articulatory mirror neurons.

By contrast, many other stimuli that do not match our motor programmes occasionally occur while we perform an action (imagine a baby grasping at a daycare full of other babies crawling around and throwing things), but these sensory inputs will not have the same contingency or tight temporal precedence to the activation of specific motor programmes, and should hence average out like noise. Certain special cases, however, could create close temporal precedence and non-matching contingencies. For instance, each time a person gives something to baby, the sight of the placing hand will just precede the execution of grasping and could lead to some degree of association between STS neurons for placing and PM neurons for grasping. Indeed, so-called 'logically related mirror neurons' seem to exist [2], and laboratory experiments suggest that repeatedly experiencing non-matching contingencies can temporarily link motor programmes to non-matching action observations (see section 4 of [39] for a review). Additionally, an object that can be grasped in a particular way will always be systematically present when baby grasps that object in that way, predicting Hebbian connections between shape neurons in the visual system and PM neurons that code the affordances of this object. Indeed, such connections seem to exist and can be observed in so-called canonical neurons [40].

Unfortunately, there is very little work that empirically tests our assumption that the statistical relationships (contingency and occurrence within the temporal window of Hebbian learning) between what we do and what we sense (hear and see) on average is such that matching relationships in the mirror neuron system would prevail. A small number of studies have analysed movies of babies and their parents and found that parents often imitate the vocalizations and facial expressions of their babies, and babies are known to spend much of their time looking at their own hands and the facial expressions (which are often imitative) of their carers (see [32] and section 5 of [37] for a review). A powerful way to test our hypothesis would be to record the sensory input to the baby, and the baby's own actions over substantial amounts of time to examine the statistical relationship between motor output and auditory-visual input. Currently, what makes such a project unlikely is the manual labour required to analyse days of such recordings. However, with head-mounted devices (e.g. Google glasses) and three-dimensional motion tracking (e.g. Microsoft's Kinect) becoming mainstream, we might soon be able to systematically quantify the sensorimotor contingencies experienced by real babies. Until then, the rest of our argument is based on the mere assumption that sensorimotor contingencies would favour a significant proportion of matching Hebbian connections between STS and PM.

(c) From re-afference to mirror properties

If such matching connections have been trained and the individual hears someone perform a similar action, the sound of the action, by resemblance to the re-afferent sounds that were associated with the listener's past actions, would activate STS neurons, which would trigger, through the potentiated synapses, PM neurons triggering the execution of actions generating similar sounds. The PM neurons would become mirror neurons. The activity of the PM neurons while listening to the actions of others would essentially be a recollection of past procedural memories of what motor state occurred together with these sensory events, but a recollection that is activated through an external social stimulus. This places mirror neurons in a wider family of reactivation phenomena, also including memory and imagination.

The case of vision is more complex: one's own actions are seen from an egocentric perspective, those of others from a different, allocentric perspective. So how would the sight of the actions of others trigger STS neurons responding to the sight of our own actions? First, some STS neurons respond to the sight of an action seen from a number of different perspectives [24]. How these neurons acquire this property is not entirely clear, but in monkeys such viewpoint invariance can

emerge after experiencing different perspectives of the same three-dimensional object [41]. Accordingly, it might be the opportunity to see the actions of others from a number of perspectives that endows STS neurons with the capacity to respond to the sight of actions across perspectives, and thus agents. Second, neurons might represent certain viewpoint invariant properties of an action (e.g. rhythmicity, temporal frequency, etc.), that can be matched to actions with similar properties [42], without needing to rotate the action in the mind's eye. Third, instances of imitation or physical mirrors would allow humans to experience the kind of contingencies that would favour Hebbian learning also between the third-person perspective of seeing the actions of others and performing their own actions. Finally, for actions that have a characteristic sound, individuals might first experience the contingencies between seeing and hearing other people perform these actions (e.g. hearing speech while seeing lip movements). This could lead to multimodal neurons in the STS [43]. The sight of the action could then trigger matching motor actions because it triggers activity in the same audio-visual speech neurons that have been linked to the viewer's motor programme during auditory re-afference. Which (combination) of these phenomena account for the emergence of visuo-motor mirror neurons that can cope with the difference in perspective remains for experiments to investigate.

(d) Alternative accounts

The mainstream of papers on the mirror-neuron system do not directly address the question of how mirror neurons emerge during development but suggest that these neurons could serve social cognition, and thereby promote survival [2,6,9,25,44–47]. Some (e.g. Cecilia Heyes) have read such functional claims as indicating that 'The standard view of MNs, which we will call the "genetic account", alloys a claim about the origin of MNs with a claim about their function. It suggests that the mirrorness of MNs is due primarily to heritable genetic factors, and that the genetic predisposition to develop MNs evolved because MNs facilitate action understanding' [31]. We believe that this is an inaccurate reading of the neuroscientific work on mirror neurons: for a neuroscientist, stating that mirror neurons could contribute to social cognition and thus endow animals with fitness advantages does not automatically translate into suggesting that humans and monkeys are hardwired to have mirror neurons and that learning must only have a minimal impact on mirror neurons. What neuroscientists mean is that if one were to disturb the function of mirror neurons, this would lead to impairments in social cognition and a growing body of evidence now exists to support this claim [48]. Such a claim is compatible with the genome undergoing selective pressure to facilitate mirror neurons, but it does not imply that this selection has already generated a strong genetic encoding or that the genetic influence takes the form of pre-wiring at birth. As described above, the genome could canalize Hebbian learning of mirror neurons instead of predisposing individuals to generate the right kind of learning opportunities [29]. In short, the 'standard view' can be criticized for neglecting the ontogenesis of mirror neurons but does not hold that mirror neurons in all their complexity are genetically encoded and immune to learning.

Of the theories that address the ontogenesis of mirror neurons, all seem to give experience a very significant role. This is true for our Hebbian learning model, associative sequence learning [28,30,31], the model of Casile et al. [34], the epigenetic model of Ferrari et al. [49], the Bayesian model of Kilner et al. [50] and for the vast majority of computational models of the mirror-neuron system [51]. All of the models also incorporate an important role for genetic predisposition, at least in connecting sensory and motor regions with synaptic connections and implementing some basic learning rules into the system. The main difference between the models is probably the level of description they most directly target. Our Hebbian learning model is a neuroscientific bottom-up approach, which starts with the small building blocks of the system—the spike-time-dependent plasticity that occurs at the synapses and the anatomical details of the connections (see also §4)—and examines whether mirror neurons would emerge bottom-up from the interaction of these building blocks. Associative sequence learning is not a neural but a cognitive model and emphasizes the system-level variables that behavioural experiments have shown to be critical for associative learning, but does not address how the learning is implemented in the biology of synapses [28,31]. Casile et al. [34] alert us to the possibility that mirror neurons for different actions might emerge in different ways: genetic pre-wiring might be more important for facial expressions, while Hebbian learning might be more important for hand actions [34]. The epigenetic model adds that experience could act not only by triggering Hebbian learning, but also by epigenetically modifying what part of the genes can be expressed [49]. Computational models emphasize the overall architecture of the system in terms of information content but often use error back-propagation algorithms with no specific hypotheses about the biological implementation of these learning rules [51]. Finally, Kilner et al.'s predictive coding account [50] describes mirror neurons at the systems level using Bayesian statistics. How these statistics are computed in the biology of the synapses is not in the scope of Kilner et al.'s theory.

We therefore feel that there is basic consensus on the importance of learning in mirror neuron ontogenesis. Depicting the field as made of two camps, with one supposedly claiming that it is all genetics and the other generating experiments to disprove the genetic hypothesis seems a distortion. Instead, existing theories seem not so much competing alternatives, but parallel attempts to explore how experience can forge a very complex phenomenon starting from different levels of focus, with some focusing on the lowest, synaptic level, others on the interaction between brain regions and others still at the level of associations of cognitive entities. Over the next decades, the major challenge will be to unify these somewhat 'local' attempts into a unified model that accounts for all levels. In the meantime, it seems of little use to debate which approach is the best. Letting the different models develop further will shed light onto the levels each model explores most directly. It is in this spirit that we now switch to explore a finer temporal dimension of Hebbian learning to show how synaptic bottom-up predictions dove-tail with more top-down predictions made by Kilner et al. [50].

4. The micro-temporal Hebbian perspective and predictions

A key feature of our modern understanding of Hebbian learning is its exquisite sensitivity to the fine temporal relations of pre- and postsynaptic activity. Here we therefore examine

the core idea of our Hebbian learning account—direct or indirect re-afference—at this millisecond time-scale, expanding the brief account presented in Keysers and Keysers et al. [6,52].

(a) Predictive forward connections

If you think of reaching for a cookie, grasping it, and then bringing it to the mouth, in the outside world, the timing of each subcomponent of the action and their sensory consequences coincide exactly in time (figure 3a). However, it takes approximately 100 ms for premotor activity to trigger complex overt actions like reaching and grasping [53]. It then takes another 100 ms for the sound/vision of that action to trigger activity in the STS [54]. This will therefore shift the spiking of the STS neurons representing the vision and sound of an action by approximately 200 ms relative to that of the PM neurons that triggered the action (figure 3b). Hence, the macro-temporal notion that activity in the STS neurons for an action overlaps in time with that of the PM neurons that trigger the action is actually an oversimplification. This has consequences for Hebbian learning, because STS responses to the sight of reaching no longer occur just before activity in PM neurons for reaching, as the 40 ms window of spike-time-dependent plasticity (figure 1) would require. Instead, the firing of neurons in STS responding to a particular phase of the action (e.g. reaching) precedes PM neural activity triggering the next phase (e.g. grasping), and Hebbian learning should primarily reinforce the connections between STS reaching and PM grasping neurons. The dominant learning result should thus be a connection with predictive properties. Some Hebbian learning might still occur within a given action phase, because early spikes of the STS reaching neurons occur just before late spikes of the PM reaching neurons.

How much would this system predict? If we have a temporal delay of approximately 200 ms between PM neuron activity and the firing of STS neurons that represent the re-afference, the sight of an action component occurring in the outside world at time t would trigger activity (through the synapses that were shaped by Hebbian learning) in PM neurons that represent the action component that normally occurs in the outside world at $t + 200$ ms. The motor and sensory delays therefore directly determine the predictive horizon of the sensorimotor connectivity. Hence, Hebbian learning would train a predictive system simply owing to the temporal asymmetry of STDP (figure 1) and the known latencies in the sensory and motor system (figure 3b).

In the real world, action components can organize in many different action sequences like letters in words, and these predictive STS → PM connections would be likely to reflect the transition probability distribution of our actions: if during our past motor history, action A was never followed by action x_1 ($p(x_1|A) = 0$), sometimes by x_2 ($p(x_2|A) = 0.2$), and often by x_3 ($p(x_3|A) = 0.8$), Hebbian learning would expect an STS neuron responding to A to have a quasi-zero connection weight with PM neurons triggering $\times 1$, a 0.2 weight with those triggering $\times 2$ and a 0.8 weight with those triggering $\times 3$. Hence, the PM neurons for these three actions should have activity states of 0, 0.2 and 0.8 following the representation of action A in STS. The activity pattern in PM is then a probability distribution of upcoming actions that reflect the past motor contingencies of the observer and could act as a prior (in the Bayesian sense) for the action that is likely to be seen next.
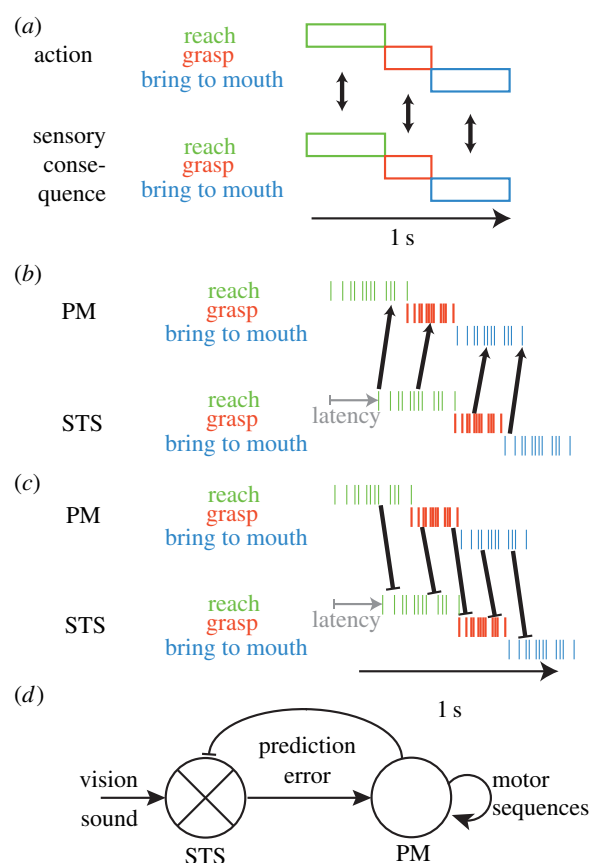

Figure 3. (a) In the real world, the execution of an action and the sight and sound of each phase of an action occur at the same time, and one might therefore predict that corresponding phases in the sensory and motor domain would become associated. (b) Instead, latencies shift the responses in the STS relative to the premotor (PM) neurons, and Hebbian learning at a fine temporal scale predicts associations between subsequent phases, i.e. predictions. (c) Inhibitory feedback from PM to STS is also subjected to Hebbian learning and generates prediction errors in the emerging dynamic system (d). (Online version in colour.)

(b) Inhibitory backward connections and prediction errors

An often-ignored element of the anatomy of the mirror neuron system is the presence of backward connections from PM to STS, which seem to have a net inhibitory influence [55,56]. From a Hebbian point of view, for these connections the situation is a little different, as the PM neurons indeed fire prior to the STS neurons, as Hebbian learning requires, albeit 200 ms instead of the 40 ms prior that are optimal for Hebbian learning. Hence, for these inhibitory feedback connections, inhibitory projections from PM neurons encoding a particular phase of the action should be strengthened with STS representations of the same action and that occurring just before (figure 3c).

Once we consider both the forward and backwards information flow, the mirror neuron system no longer seems a simple associative system in which the sight of a given action triggers the motor representation of that action. Instead, it becomes a dynamic system (figure 3d). The sight and sound of an action triggers activity in STS neurons. This leads to a pattern of predictive activation of PM neurons encoding the action that occurs 200 ms after what the STS neurons represent, with their respective activation levels representing the

likelihood of their occurrence based on past sensorimotor contingencies. However, the system would not stop at that point. This prediction in PM neurons is sent backwards as an inhibitory signal to STS neurons. Because the feedback should be onto neurons representing the previous and current actions represented in PM, it should have two consequences. It would terminate the sensory representation of past actions, which could contribute to what is often termed backward masking in the visual literature [57]. Second, by cancelling representations associated with x_1, x_2 and x_3 with their respective probabilities, it will essentially inhibit those STS neurons that represent the expected sensory consequences of the action that the PM neurons predict to occur. At a more conceptual level, it would inhibit the hypothesis that PM neurons entertain about the next action to be perceived. As the brain then sees and hears what action actually comes next, if this input matches the hypothesis, the sensory consequences of that action would be optimally inhibited, and little information would be sent from STS → PM. Because PM neurons (and the posterior parietal neurons [58]) are organized in action chains within the premotor cortex, the representation of action x_3 would then trigger activation of those actions that normally follow action x_3 during execution, actively generating a whole stream of action representations of PM neurons without the need for any further sensory drive, and these further predictions would keep inhibiting future STS input. If action x_2 were to follow action A, the inhibition would be weaker and more of the sensory representation of x_2 would leak through to PM. This would represent a 'prediction error', which will change the pattern of PM activity to better match the input, away from the prior expectations. If action x_1 were to follow action A, no cancellation would be in place in the STS, and the strongest activity would be sent from STS → PM, rerouting PM activity onto a stream of actions that normally follows x_1, rather than x_3, as initially hypothesized.

At this temporal resolution, during action observation/listening, the pattern of activity across nodes in PM is no longer a simple mirror of what happens in STS, but an actively predicted probability distribution for what the observer should perceive the observed individual to do next. By virtue of Hebbian learning, the entire STS-PM loop becomes a dynamic system that performs predictive coding. When the observed action unfolds entirely as expected, activity in the PM would actually be generated using the sequences of normal motor control rather than by visual input.

(c) Regulating learning and contact points with other models

An important consequence of the feedback inhibition is also that during re-afference, once the system has learned, the execution of an action would trigger inhibitory feedback to STS neurons that will ensure that STS neurons actually limit their input to the PM neurons that caused the action, and Hebbian learning would be self-limiting. If the contingencies change, e.g. a person learns a new skill like playing the piano, PM neurons fail to predict the auditory re-afference, and new learning occurs because new input from STS is sent to PM neurons with the potential for Hebbian learning.

This calculation of prediction errors within the Hebbian learned system creates an important contact point with other models of the mirror neuron system. The predictive coding model of Kilner *et al.* [50] does not indicate how the brain performs Bayesian predictions within its synapses, but proposes that PM activity represents a Bayesian estimate of future actions, which allows the observer to deduce the motor intentions of the observed individual. This model sees STS → PM information flow as a mere updating signal for the Bayesian probabilities of premotor states. Our model arrives at very similar interpretations from a bottom-up perspective. Our Hebbian learning model can thus complement the Bayesian prediction model with a plausible biological bottom-up implementation. In turn, the Bayesian prediction model helps interpret the information processing we describe in the light of what could be called a Bayesian or predictive coding revolution in brain science. Indeed many domains of brain science now stop to consider perception as a hierarchical process in which sensory information is passively sent forward from lower to higher brain regions. Instead, perception is increasingly seen as a more active process, in which the brain makes predictions based on past experience (the equivalent of prior probabilities in Bayesian terms), that are sent from higher to lower brain regions in the hierarchy and are subtracted from the actual sensory input. The sensory input that is sent from lower to higher regions after the subtraction of predictions is then a prediction error that serves to update predictions, rather than directly driving perception. This very general framework has been very successfully used to understand neural activity in the early stages of the visual cortex [59,60] but has also recently been used to conceptualize the mirror neuron system [50] and even mentalizing [61].

Evidence for predictive coding in the mirror neuron system is still rare but starts to emerge. The predictive nature of the PM response is evident from the fact that images of reaching increase the excitability of muscles involved in the most likely following action phase, grasping [62]. The possibility that PM activity can be driven by internal predictions in the absence of explicit visual input comes from the observation that mirror neurons that respond during the execution of grasping respond to the sight of reaching behind an opaque screen [20] and that auditory mirror neurons that respond to the cracking sound of a peanut being shelled start firing ahead of this phase when viewing the hands grasping the peanut [3]. Evidence that predictions from PM → STS cancel out predicted actions and thereby silence the STS → PM information flow if but only if the actions are predictable, stems from the fact that the predominant direction of information flow is from the PM → STS when observing predictable actions, but STS → PM when observing the unpredictable beginning of an action [63].

(d) Hebbian learning and joint actions

Humans can act together with surprising temporal precision. Pianists in a duet can synchronize their actions within 30 ms of a leader [64]. Given approximately 200 ms of sensorimotor delays we mentioned above, how is this possible? Should not it take 200 ms for a musician to hear what the leader played and respond to it? One of the powerful implications of a fine-grained analysis of Hebbian learning is that because the synaptic connections are trained by re-afference that includes typical human sensorimotor delays, they train the connections from STS → PM to perform predictions into the future with a time-shift that will offset the sensorimotor delays that are encountered when acting with another individual subject to similar delays. This is because it will take
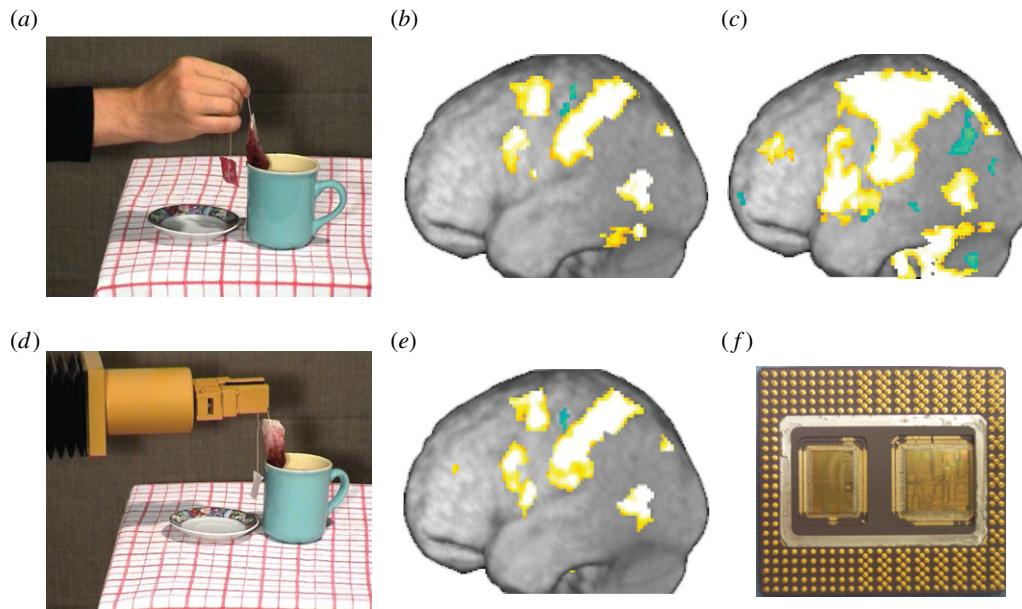
Figure 4. Seeing a human perform an action (*a*) leads to activations in the mirror system (*b*) that resembles the activity during the execution of similar actions by a human (*c*). Seeing a robot perform similar actions (*d*) generates a pattern of activity in the mirror system (*e*) that is very different from the pattern of activity that caused the robot to act (*f*), but resembles that which the human viewer would use to perform a similar action (*c*). Panels (*a*−*e*) adapted from Gazzola *et al.* [65]. (Online version in colour.)

approximately the same time (approx. 200 ms) for your motor programme to activate your STS neurons, as it would take for your motor programme to activate my STS neurons while I am listening/watching you. Hence, Hebbian learning by re-afference trains sensorimotor predictions that permit accurate joint actions despite long sensorimotor delays.

(e) Hebbian learning and projection
An important consequence of the notion that the mirror neuron system is wired up based on re-afference is that the brain associates the internal states that were present when we produced a certain action with the sound and vision of that action. Accordingly, when we witness the actions of others, the pattern of motor activity that would be predictively activated in the witness is not so much a reflection of what happens in the brain of the actor, but rather a projection of what happened in our own brain when we performed such actions. Because humans share approximately 99% of their genes with other humans, and probably over 90% of genes with macaques, assuming that hidden motor states that occurred during our own actions are a decent model for those that happen in the brain of another human or monkey is not unreasonable. It constitutes an informative 'prior' that can be updated by contrary evidence if available. However, the more different the observer is from the observed agent, the more the projective nature of this process should become evident.

To test this prediction, we measured brain activity using functional magnetic resonance imaging (fMRI) in three conditions [65]. Participants performed hand actions (e.g. swirling a wine glass). They saw another human perform similar actions. Finally, they saw an industrial robot perform similar actions. Seeing the human perform the action (figure 4*a*) activated a network of somatosensory, premotor and parietal brain activity (figure 4*b*) that was similar to that used by participants to perform similar action (figure 4*c*). Comparing the activity pattern of observers and executers (*b*−*c*) reveals a significant similarity ($r(b,c) = 0.5$)—the brain succeeded in simulating

the brain activity of the agent accurately. However, when participants viewed the robot perform the action (figure 4*d*), they generated a pattern of brain activity (figure 4*e*) unlike the activity of the processor that caused the robot to move (figure 4*f*). Instead, the pattern continued to resemble that which the participant would have used to perform this action (figure 4*c*). This illustrates the projective nature of mind-reading through the mirror neuron system.

5. Beyond the motor system
Because mirror neurons were first found in PM [1–4,20] and in the posterior parietal regions [21,58], which control actions, motor aspects of mind-reading were in the limelight. But evidence from a number of sources now suggests that the highest levels of the primary somatosensory cortex are also vicariously activated when we see the actions of others and the secondary somatosensory cortex when we see others be touched [45]. In addition, regions involved in experiencing emotions also become vicariously activated when we witness others experience similar emotions [6,66], including the insula for disgust, pain and pleasure [67–69], the rostral cingulate for pain [68] and the striatum for reward [70].

We still lack single-cell recordings that prove that vicarious somatosensory and emotional activations in fMRI are caused by single cells responding to both the experience and observation of somatosensation and emotions (but see [71]). However, from a Hebbian learning perspective, mirror-like neurons for somatosensation or emotions are not surprising. Whenever something touches our skin, we see our body touched, and we feel the somatosensory stimulation. Unlike in the sensorimotor system, in which the motor activity precedes the visual/auditory re-afference, in the case of feeling touched, both the tactile and visual/auditory signal would be affected by similar latencies relative to the outside event. Spikes from visual/auditory and somatosensory neurons would therefore naturally fall within the narrow temporal windows of Hebbian learning and would reinforce the connection between neurons

encoding our inner sense of touch in S2 with those encoding what touch looks and sounds like in regions like the STS. When viewing/hearing others be touched these connections might then trigger mirror-like activity in S2 and project our own feeling of touch onto the person we see. Because anticipation in the STS → PM system is owing to differences in latencies between these neurons, which would be small between STS and S2, we would expect the STS-S2 connections to show little predictive coding. If the tactile sensation would result from actions that can be predicted by the STS → PM system, however, such anticipations could be computed.

Similarly, when we explore objects actively with our hand, activity in PM neurons controlling the action would precede activity not only in STS neurons viewing and hearing the action, but also that of neurons in BA2, which encode the haptic sensations experienced during the touch. We would thus expect the emergence of a dynamic system akin to that in figure 1, not only including STS and PM, but also BA2. In this system, Hebbian learning could then also explain how people learn to suppress tactile sensations that are self-caused, to generate the haptic prediction errors so central to motor control [72], and thus why it is impossible to tickle yourself [73].

Finally, for emotions, many neurons would also become Hebbianly connected. If we feel pain, because our bigger sister inadvertently hit us with her toy, we see the toy hit us, we feel the pain, we make a facial expression, cry and our parents will mirror that facial expression. The vision of the hit precedes our pain, which precedes the facial expression and cries we make, which precedes the facial mimicry of our parents. This could, if our theory is correct, lead to a chain of Hebbian associations across the neurons representing these states. When we then later see or hear someone wince in pain, the sound and vision will trigger our matching facial motor programmes, which will in turn activate our inner feelings [74]. If we see someone get hit,

we will vicariously recruit our somatosensory and emotional cortices. And all of these vicarious activations would be the result of synaptic plasticity during our own experiences. They will associate observable events with what we had felt and done in these situations. When applying them to others, we would project our states, with all the inevitable egocentric biases this would predict.

6. Overall conclusion

When two decades ago, the mirror neurons were first reported, they generated a vision in which the motor systems play a privileged role in reading the mind of others through embodied cognition [75]. Here, we propose that what we know about spike-timing-dependent synaptic plasticity shapes our modern understanding of Hebbian learning and provides a framework to explain not only how mirror neurons could emerge, but also how they become endowed with predictive properties that would enable quasi-synchronous joint actions. We show that this could create a system that can provide an approximate solution to the inverse problem of inferring hidden internal states of others from observable changes in the world, but that this solution is a projection plagued by egocentric biases. We also show that mirror neurons are probably a special case of vicarious activations that Hebbian learning and fMRI data suggest to also apply to how we share the emotions and sensations of others.

References

1. di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G. 1992 Understanding motor events: a neurophysiological study. *Exp. Brain Res.* **91**, 176–180. (doi:10.1007/BF00230027)

2. Gallese V, Fadiga L, Fogassi L, Rizzolatti G. 1996 Action recognition in the premotor cortex. *Brain* **119**, 593–609. (doi:10.1093/brain/119.2.593)

3. Keysers C, Kohler E, Umilta MA, Nanetti L, Fogassi L, Gallese V. 2003 Audiovisual mirror neurons and action recognition. *Exp. Brain Res.* **153**, 628–636. (doi:10.1007/s00221-003-1603-5)

4. Kohler E, Keysers C, Umilta MA, Fogassi L, Gallese V, Rizzolatti G. 2002 Hearing sounds, understanding actions: action representation in mirror neurons. *Science* **297**, 846–848. (doi:10.1126/science.1070311)

5. Hickok G. 2013 Do mirror neurons subserve action understanding? *Neurosci. Lett.* **540**, 56–58. (doi:10.1016/j.neulet.2012.11.001)

6. Keysers C. 2011 *The empathic brain*. Amsterdam, The Netherlands: Social Brain Press.

7. Keysers C, Gazzola V. 2006 Towards a unifying neural theory of social cognition. *Prog. Brain Res.*

156, 383–406. (doi:10.1016/S0079-6123(06)56021-2)

8. Rizzolatti G, Fabbri-Destro M, Cattaneo L. 2009 Mirror neurons and their clinical relevance. *Nat. Clin. Pract. Neurol.* **5**, 24–34. (doi:10.1038/ncpneuro0990)

9. Rizzolatti G, Sinigaglia C. 2010 The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat. Rev. Neurosci.* **11**, 264–274. (doi:10.1038/nrn2805)

10. Gazzola V, Aziz-Zadeh L, Keysers C. 2006 Empathy and the somatotopic auditory mirror system in human. *Curr. Biol.* **16**, 1824–1829. (doi:10.1016/j.cub.2006.07.072)

11. Hebb D. 1949 The organisation of behaviour. New York, NY: John Wiley and Sons.

12. Shatz CJ. 1992 The developing brain. *Sci. Am.* **267**, 60–67. (doi:10.1038/scientificamerican0992-60)

13. Granger CWJ. 1969 Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 414.

14. Markram H, Lubke J, Frotscher M, Sakmann B. 1997 Regulation of synaptic efficacy by coincidence of

postsynaptic APs and EPSPs. *Science* **275**, 213–215. (doi:10.1126/science.275.5297.213)

15. Bi G, Poo M. 2001 Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.* **24**, 139–166. (doi:10.1146/annurev.neuro.24.1.139)

16. Caporale N, Dan Y. 2008 Spike timing-dependent plasticity: a Hebbian learning rule. *Annu. Rev. Neurosci.* **31**, 25–46. (doi:10.1146/annurev.neuro.31.060407.125639)

17. Bauer EP, LeDoux JE, Nader K. 2001 Fear conditioning and LTP in the lateral amygdala are sensitive to the same stimulus contingencies. *Nat. Neurosci.* **4**, 687–688. (doi:10.1038/89465)

18. Rescorla RA. 1967 Pavlovian conditioning and its proper control procedures. *Psychol. Rev.* **74**, 71–80. (doi:10.1037/h0024109)

19. Cooper RP, Cook R, Dickinson A, Heyes CM. 2013 Associative (not Hebbian) learning and the mirror neuron system. *Neurosci. Lett.* **540**, 28–36. (doi:10.1016/j.neulet.2012.10.002)

20. Umilta MA, Kohler E, Gallese V, Fogassi L, Fadiga L, Keysers C, Rizzolatti G. 2001 I know what you are

doing: a neurophysiological study. *Neuron* **31**, 155–165. (doi:10.1016/S0896-6273(01)00337-3)

21. Rozzi S, Ferrari PF, Bonini L, Rizzolatti G, Fogassi L. 2008 Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. *Eur. J. Neurosci.* **28**, 1569–1588. (doi:10.1111/j.1460-9568.2008.06395.x)

22. Rozzi S, Calzavara R, Belmalih A, Borra E, Gregoriou GG, Matelli M, Luppino G 2006 Cortical connections of the inferior parietal cortical convexity of the macaque monkey. *Cereb. Cortex* **16**, 1389–1417. (doi:10.1093/cercor/bhj076)

23. Nelissen K, Borra E, Gerbella M, Rozzi S, Luppino G, Vanduffel W, Rizzolatti G, Orban GA 2011 Action observation circuits in the macaque monkey cortex. *J. Neurosci.* **31**, 3743–3756. (doi:10.1523/JNEUROSCI.4803-10.2011)

24. Keysers C, Perrett DI. 2004 Demystifying social cognition: a Hebbian perspective. *Trends Cogn. Sci.* **8**, 501–507. (doi:10.1016/j.tics.2004.09.005)

25. Keysers C, Gazzola V. 2009 Expanding the mirror: vicarious activity for actions, emotions, and sensations. *Curr. Opin. Neurobiol.* **19**, 666–671. (doi:10.1016/j.conb.2009.10.006)

26. Mukamel R, Ekstrom AD, Kaplan J, Iacoboni M, Fried I. 2010 Single-neuron responses in humans during execution and observation of actions. *Curr. Biol.* **20**, 750–756. (doi:10.1016/j.cub.2010.02.045)

27. Caspers S, Zilles K, Laird AR, Eickhoff SB. 2010 ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage* **50**, 1148–1167. (doi:10.1016/j.neuroimage.2009.12.112)

28. Heyes C. 2001 Causes and consequences of imitation. *Trends Cogn. Sci.* **5**, 253–261. (doi:10.1016/S1364-6613(00)01661-2)

29. Del Giudice M, Manera V, Keysers C. 2009 Programmed to learn? The ontogeny of mirror neurons. *Dev. Sci.* **12**, 350–363. (doi:10.1111/j.1467-7687.2008.00783.x)

30. Brass M, Heyes C. 2005 Imitation: is cognitive neuroscience solving the correspondence problem? *Trends Cogn. Sci.* **9**, 489–495. (doi:10.1016/j.tics.2005.08.007)

31. Cook R, Bird G, Catmur C, Press C, Heyes C. In press. Mirror neurons: from origin to function. *Behav. Brain Sci.*

32. Rochat P. 1998 Self-perception and action in infancy. *Exp. Brain Res.* **123**, 102–109. (doi:10.1007/s002210050550)

33. Meltzoff AN, Moore MK. 1977 Imitation of facial and manual gestures by human neonates. *Science* **198**, 75–78. (doi:10.1126/science.198.4312.75)

34. Casile A, Caggiano V, Ferrari PF. 2011 The mirror neuron system: a fresh view. *Neuroscientist* **17**, 524–538. (doi:10.1177/1073858410392239)

35. Nagy E, Pilling K, Orvos H, Molnar P. 2013 Imitation of tongue protrusion in human neonates: specificity of the response in a large sample. *Dev. Psychol.* **49**, 1628–1638. (doi:10.1037/a0031127)

36. Anisfeld M. 1991 Neonatal imitation. *Dev. Rev.* **11**, 60–97. (doi:10.1016/0273-2297(91)90003-7)

37. Jones SS. 2009 The development of imitation in infancy. *Phil. Trans. R. Soc. B* **364**, 2325–2335. (doi:10.1098/rstb.2009.0045)

38. Falk D. 2004 Prelinguistic evolution in early hominins: whence motherese? *Behav. Brain Sci.* **27**, 491–503 (discussion 83).

39. Catmur C. 2013 Sensorimotor learning and the ontogeny of the mirror neuron system. *Neurosci. Lett.* **540**, 21–27. (doi:10.1016/j.neulet.2012.10.001)

40. Murata A, Fadiga L, Fogassi L, Gallese V, Raos V, Rizzolatti G. 1997 Object representation in the ventral premotor cortex (area F5) of the monkey. *J. Neurophysiol.* **78**, 2226–2230.

41. Logothetis NK, Pauls J. 1995 Psychophysical and physiological evidence for viewer-centered object representations in the primate. *Cereb. Cortex.* **5**, 270–288. (doi:10.1093/cercor/5.3.270)

42. Cook R, Johnston A, Heyes C. 2012 Self-recognition of avatar motion: how do I know it's me? *Proc. R. Soc. B* **279**, 669–674. (doi:10.1098/rspb.2011.1264)

43. Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI. 2005 Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J. Cogn. Neurosci.* **17**, 377–391. (doi:10.1162/0898929053279586)

44. Gallese V, Keysers C, Rizzolatti G. 2004 A unifying view of the basis of social cognition. *Trends Cogn. Sci.* **8**, 396–403. (doi:10.1016/j.tics.2004.07.002)

45. Keysers C, Kaas JH, Gazzola V. 2010 Somatosensation in social perception. *Nat. Rev. Neurosci.* **11**, 417–428. (doi:10.1038/nrn2833)

46. Rizzolatti G, Craighero L. 2004 The mirror-neuron system. *Annu. Rev. Neurosci.* **27**, 169–192. (doi:10.1146/annurev.neuro.27.070203.144230)

47. Gallese V, Sinigaglia C. 2011 What is so special about embodied simulation? *Trends Cogn. Sci.* **15**, 512–519. (doi:10.1016/j.tics.2011.09.003)

48. Avenanti A, Candidi M, Urgesi C. 2013 Vicarious motor activation during action perception: beyond correlational evidence. *Front. Hum. Neurosci.* **7**, 185. (doi:10.3389/fnhum.2013.00185)

49. Ferrari PF, Tramacere A, Simpson EA, Iriki A. 2013 Mirror neurons through the lens of epigenetics. *Trends Cogn. Sci.* **17**, 450–457. (doi:10.1016/j.tics.2013.07.003)

50. Kilner JM, Friston KJ, Frith CD. 2007 Predictive coding: an account of the mirror neuron system. *Cogn. Process.* **8**, 159–166. (doi:10.1007/s10339-007-0170-2)

51. Oztop E, Kawato M, Arbib MA. 2013 Mirror neurons: functions, mechanisms and models. *Neurosci. Lett.* **540**, 43–55. (doi:10.1016/j.neulet.2012.10.005)

52. Keysers C, Perrett DI, Gazzola V. In press. Hebbian learning is about contingency, not contiguity, and explains the emergence of predictive mirror neurons. *Behav. Brain Res.*

53. Graziano MS, Aflalo TN, Cooke DF. 2005 Arm movements evoked by electrical stimulation in the motor cortex of monkeys. *J. Neurophysiol.* **94**, 4209–4223. (doi:10.1152/jn.01303.2004)

54. Keysers C, Xiao DK, Foldiak P, Perrett DI. 2001 The speed of sight. *J. Cogn. Neurosci.* **13**, 90–101. (doi:10.1162/089892901564199)

55. Hietanen JK, Perrett DI. 1993 Motion sensitive cells in the macaque superior temporal polysensory area. I. Lack of response to the sight of the animal's own limb movement. *Exp. Brain Res.* **93**, 117–128. (doi:10.1007/BF00227786)

56. Hietanen JK, Perrett DI. 1996 Motion sensitive cells in the macaque superior temporal polysensory area: response discrimination between self-generated and externally generated pattern motion. *Behav. Brain Res.* **76**, 155–167. (doi:10.1016/0166-4328(95)00193-X)

57. Keysers C, Xiao DK, Foldiak P, Perrett DI. 2005 Out of sight but not out of mind: the neurophysiology of iconic memory in the superior temporal sulcus. *Cogn. Neuropsychol.* **22**, 316–332. (doi:10.1080/02643290442000103)

58. Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G. 2005 Parietal lobe: from action organization to intention understanding. *Science* **308**, 662–667. (doi:10.1126/science.1106138)

59. Lee TS, Mumford D. 2003 Hierarchical Bayesian inference in the visual cortex. *J. Optic. Soc. Am. A* **20**, 1434–1448. (doi:10.1364/JOSAA.20.001434)

60. Rao RP, Ballard DH. 1992 Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87. (doi:10.1038/4580)

61. Koster-Hale J, Saxe R. 2013 Theory of mind: a neural prediction problem. *Neuron* **79**, 836–848. (doi:10.1016/j.neuron.2013.08.020)

62. Urgesi C, Maieron M, Avenanti A, Tidoni E, Fabbro F, Aglioti SM. 2010 Simulating the future of actions in the human corticospinal system. *Cereb. Cortex* **20**, 2511–2521. (doi:10.1093/cercor/bhp292)

63. Schippers MB, Keysers C. 2011 Mapping the flow of information within the putative mirror neuron system during gesture observation. *Neuroimage* **57**, 37–44. (doi:10.1016/j.neuroimage.2011.02.018)

64. Keller PE, Knoblich G, Repp BH. 2007 Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Conscious. Cogn.* **16**, 102–111. (doi:10.1016/j.concog.2005.12.004)

65. Gazzola V, Rizzolatti G, Wicker B, Keysers C. 2007 The anthropomorphic brain: the mirror neuron system responds to human and robotic actions. *Neuroimage* **35**, 1674–1684. (doi:10.1016/j.neuroimage.2007.02.003)

66. Bastiaansen JA, Thioux M, Keysers C. 2009 Evidence for mirror systems in emotions. *Phil. Trans. R. Soc. B* **364**, 2391–2404. (doi:10.1098/rstb.2009.0058)

67. Jabbi M, Swart M, Keysers C. 2007 Empathy for positive and negative emotions in the gustatory cortex. *Neuroimage* **34**, 1744–1753. (doi:10.1016/j.neuroimage.2006.10.032)

68. Lamm C, Decety J, Singer T. 2011 Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage* **54**, 2492–2502. (doi:10.1016/j.neuroimage.2010.10.014)

69. Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G. 2003 Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron* **40**, 655–664. (doi:10.1016/S0896-6273(03)00679-2)

70. Monfardini E, Gazzola V, Boussaoud D, Brovelli A, Keysers C, Wicker B. 2014 Vicarious neural processing of outcomes during observational learning. *PLoS ONE* **9**, e73879. (doi:10.1371/journal.pone.0073879)

71. Ishida H, Nakajima K, Inase M, Murata A. 2009 Shared mapping of own and others' bodies in visuotactile bimodal area of monkey parietal cortex. *J. Cogn. Neurosci.* **22**, 83–96. (doi:10.1162/jocn.2009.21185)

72. Cui F, Arnstein D, Thomas RM, Maurits NM, Keysers C, Gazzola V. 2014 Functional magnetic resonance imaging connectivity analyses reveal efference-copy to primary somatosensory area, BA2. *PLoS ONE* **9**, e84367. (doi:10.1371/journal.pone.0084367)

73. Blakemore SJ, Wolpert D, Frith C. 2000 Why can't you tickle yourself? *Neuroreport* **11**, R11–R16. (doi:10.1097/00001756-200008030-00002)

74. Jabbi M, Keysers C. 2008 Inferior frontal gyrus activity triggers anterior insula response to emotional facial expressions. *Emotion* **8**, 775–780. (doi:10.1037/a0014194)

75. Gallese V, Goldman A. 1998 Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* **2**, 493–501. (doi:10.1016/S1364-6613(98)01262-5)

rstb.royalsocietypublishing.org   *Phil. Trans. R. Soc. B* **369**: 20130175

11