The time course of eye movements during action observation reflects sequence learning

Clare M. Press[a] and James M. Kilner[b]

When we observe object-directed actions such as grasping, we make predictive eye movements. However, eye movements are reactive when observing similar actions without objects. This reactivity may reflect a lack of attribution of intention to observed actors when they perform actions without 'goals'. Alternatively, it may simply signal that there is no cue present that has been predictive of the subsequent trajectory in the observer's experience. To test this hypothesis, the present study investigated how the time course of eye movements changes as a function of visual experience of predictable, but arbitrary, actions without objects. Participants observed a point-light display of a model performing sequential finger actions in a serial reaction time task. Eye movements became less reactive across blocks. In addition, participants who exhibited more predictive eye movements subsequently demonstrated greater learning when required either to execute, or to recognize, the sequence. No measures were influenced by whether participants had been instructed that the observed movements were human or lever generated. The present data indicate that eye movements when observing actions without objects reflect the extent to which the trajectory can be predicted through experience. The findings are discussed with reference to the implications for the mechanisms supporting perception of actions both with and without objects as well as those mediating inanimate object processing. ***NeuroReport*** **24:822–826 © 2013 Wolters Kluwer Health | Lippincott Williams & Wilkins.**

NeuroReport 2013, **24**:822–826

Keywords: action observation, eye movements, learning, point light, prediction, serial reaction time task

[a]Department of Psychological Sciences, Birkbeck University of London and [b]Sobell Department of Motor Neuroscience and Movement Disorders, Institute of Neurology, University College London, London, UK

Correspondence to Clare M. Press, PhD, Department of Psychological Sciences, Birkbeck University of London, Malet Street, London WC1E 7HX, UK
Tel: +44 20 3073 8007; fax: +44 20 7631 6312; e-mail: c.press@bbk.ac.uk

Received 11 May 2013 **accepted** 3 July 2013

Introduction

When both adults and children observe a model grasping an object, they look to the object before contact is made [1,2]. However, eye movements have been found to be reactive when a closed hand moves between the same locations [2]. It has been proposed that eye movements are predictive when observing object-directed actions because the observer attributes to the actor an intention to grasp an object [1,2]. In contrast, eye movements may be intrinsically reactive when observing actions that are not directed towards objects because of no attributed intentionality to actors when movements lack a 'goal' [2].

Alternatively, the reactivity may simply signal that there is no cue present that has been predictive of the subsequent trajectory in the observer's experience. In the course of everyday life, we have much opportunity to learn about likely trajectories, or sequences, of action when objects are present. For example, we can learn that sight of an open hand in the presence of a ball will be followed by movement of the hand towards the ball. According to the 'learned trajectory' hypothesis, when there is no object towards which an action is directed, other cues can signal the action instead, like another action. For example, we can learn that observation of one arm going up in a Mexican wave fashion will be followed by observation of the other arm doing a similar thing. Such prediction may operate solely within the visual system [3] or may be influenced additionally by motor encoding of observed actions due to mirror mechanisms [4]. Therefore, according to the 'learned trajectory' hypothesis, greater learning about the trajectory of actions without objects may result in more predictive eye movements.

To test this hypothesis, the present study required participants to observe repetitive, but arbitrary, sequences of finger actions without objects in a serial reaction time (SRT) paradigm [5], while their eye movements were recorded. Under the 'learned trajectory' hypothesis, eye movements should become less reactive across time. To provide a second test of our hypothesis, we removed all action information other than points of light tracking movement at the fingertips [6]. Presenting stimuli in this manner allowed us to instruct half of the participants that the observed movements were generated by human finger actions and the other half that they were the product of lever movements. According to our hypothesis, the time course of eye movements should proceed similarly in the two instruction groups.

The learned trajectory hypothesis also predicts that the time course of eye movements will be correlated with behavioural measures of sequence learning. Therefore, following the observation phase, we took two measures of sequence learning. First, participants were required to

0959-4965 © 2013 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/WNR.0b013e328364d5fd

execute the observed sequence, performing an SRT task themselves [5]. Second, we presented the sequences and required participants to rate explicitly how likely it was that they had seen them in the observation phase [7].

Materials and methods

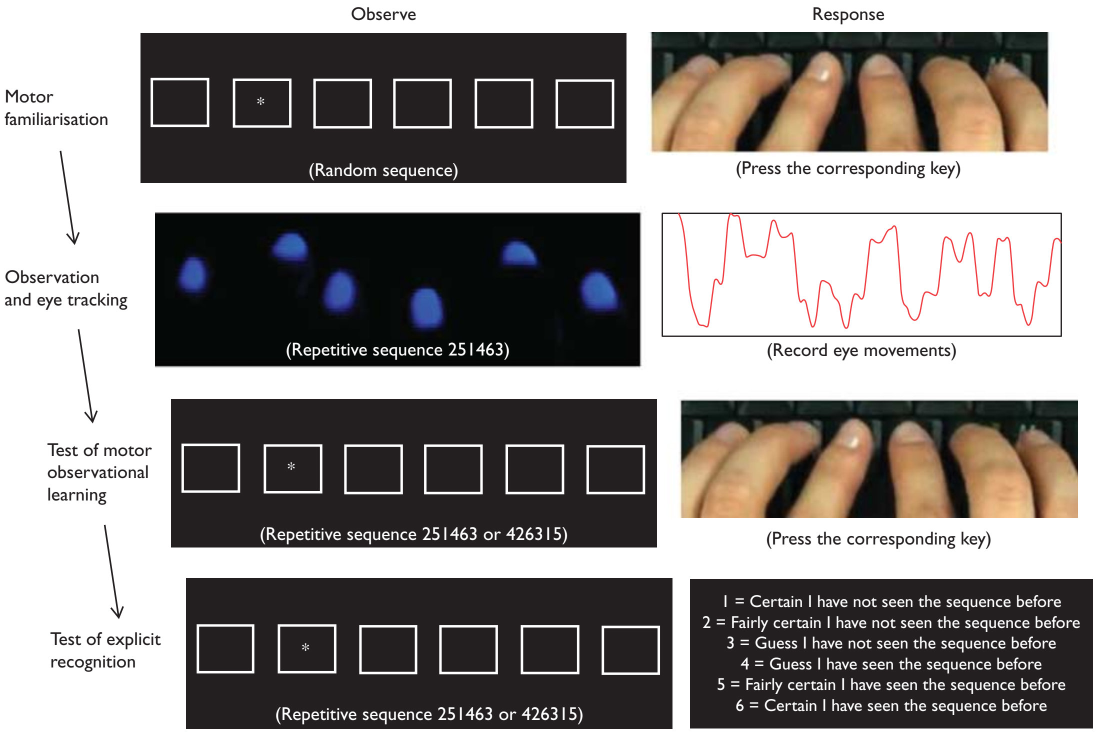
Twenty individuals (12 women) participated in this study, 10 in the human-instruction group and 10 in the lever-instruction group (mean age: 23.6 years, range = 20–29). The experiment was conducted with the approval of the local ethics committee and in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Participants first familiarized themselves with the finger actions, but not sequences, used in the present study [5]. On the monitor were six empty boxes arranged in a horizontal line. Asterisks appeared in these boxes in a random sequence. Participants were told that when an asterisk appeared in a box, they were to press the spatially corresponding key as quickly as possible (with their ring to index fingers from both hands arranged on the ‘X’ through to ‘M’ keys; Fig. 1). After a correct keypress, there was an intertrial interval of 200 ms before the next asterisk appeared. There were 200 trials in this phase.

In the subsequent observation phase, participants watched point-light videos of a human model making a six-item repetitive sequence of keypresses (251463, where 1 = left ring finger and 6 = right ring finger). Videos were generated by painting ultraviolet nail polish onto the fingernails of the six fingers involved in performing the actions and then filming in the dark with an ultraviolet light. All remaining perceptual information other than the six moving points was removed using Adobe after effects software (Adobe Systems Inc., San Jose, California, USA). The sequence was filmed with a metronome to establish a rate of one keypress every 750 ms, selected as a likely rate at which participants would press the buttons in the test phase [5]. The model made no errors. The human-instruction group was told that these videos had been made by tracking a model’s fingertips. The lever-instruction group was informed that they were generated by tracking the ends of moving levers. Both groups were told that it has been shown that the more closely they attend, the better they will perform in later stages of the experiment [5], but they were not informed that a sequence was present or of the nature of subsequent tasks. Participants’ eye position in the vertical and horizontal dimensions was tracked at all times using an Eyelink 1000 eyetracker (SR Research Ltd, Mississauga, Ontario, Canada) at a sampling rate of 1000 Hz. Participants’ heads were kept in position with a chin and forehead rest. There were six blocks of 100 trials.

Following the observation phase, participants were required to perform the same task as in the familiarization phase. In blocks 1 and 3, the asterisks appeared in

Fig. 1



1 = Certain I have not seen the sequence before
2 = Fairly certain I have not seen the sequence before
3 = Guess I have not seen the sequence before
4 = Guess I have seen the sequence before
5 = Fairly certain I have seen the sequence before
6 = Certain I have seen the sequence before

Schematic of the experimental phases.

the order that had been observed (251463), and in blocks 2 and 4, they appeared in a novel six-item sequence (426315). These sequences mirrored those used in the study of Bird and Heyes [5]. There were 100 trials in each block. Finally, participants were presented with groups of six trials, in the same manner as the previous blocks. At the end of each group of six, they were asked to rate whether they had seen the sequence earlier. The responses ranged from '1' if they were certain that they had not seen the sequence before to '6' if they were certain that they had seen it. There were six groups of trials presented in the observed sequence and six groups in the novel sequence, and these were presented in a random order.

Results

Greenhouse–Geisser corrections were applied where appropriate. *t*-tests and correlations were one-tailed because of clear predictions concerning the direction of effects. There were no effects of instruction on stimulus identity in any of the analyses (all Fs < 1.9, all Ps > 0.19; Fig. 2).

Eye tracking

For each block, a vector of the position of the eyes in the horizontal dimension was low pass filtered at 5 Hz, and 1000 datapoints were subsequently removed from either side of a missing datapoint. The resulting vector was cross-correlated with one of the finger positions (1–6), with the finger position vector aligned between 2000 ms before and 2000 ms after the eye position vector. We determined the time of peak correlation between these vectors. Positive values indicated that the peak correlation was obtained when the eye vector followed the finger vector by that amount (i.e. eye movements were reactive) and negative values reflected a peak correlation when the finger vector followed the eye vector (i.e. eye movements were predictive). One participant had no data from block 6; the missing value was replaced with the series mean.

First, a one-sample *t*-test on the overall time of peak correlation (compared against 0) demonstrated that eye movements were reactive on average [$t(19) = -5.0$, $P < 0.001$; Fig. 2a], with a mean reactivity of 365.5 ms (SEM = 73.3 ms). Second, a one-way analysis of variance with time point in the observation phase as a within-participant factor (first-third, middle-third or final-third), instruction as a between-participant factor (human, lever) and mean peak correlation as a covariate indicated a significant linear trend across time [$F(1,17) = 10.9$,

Fig. 2

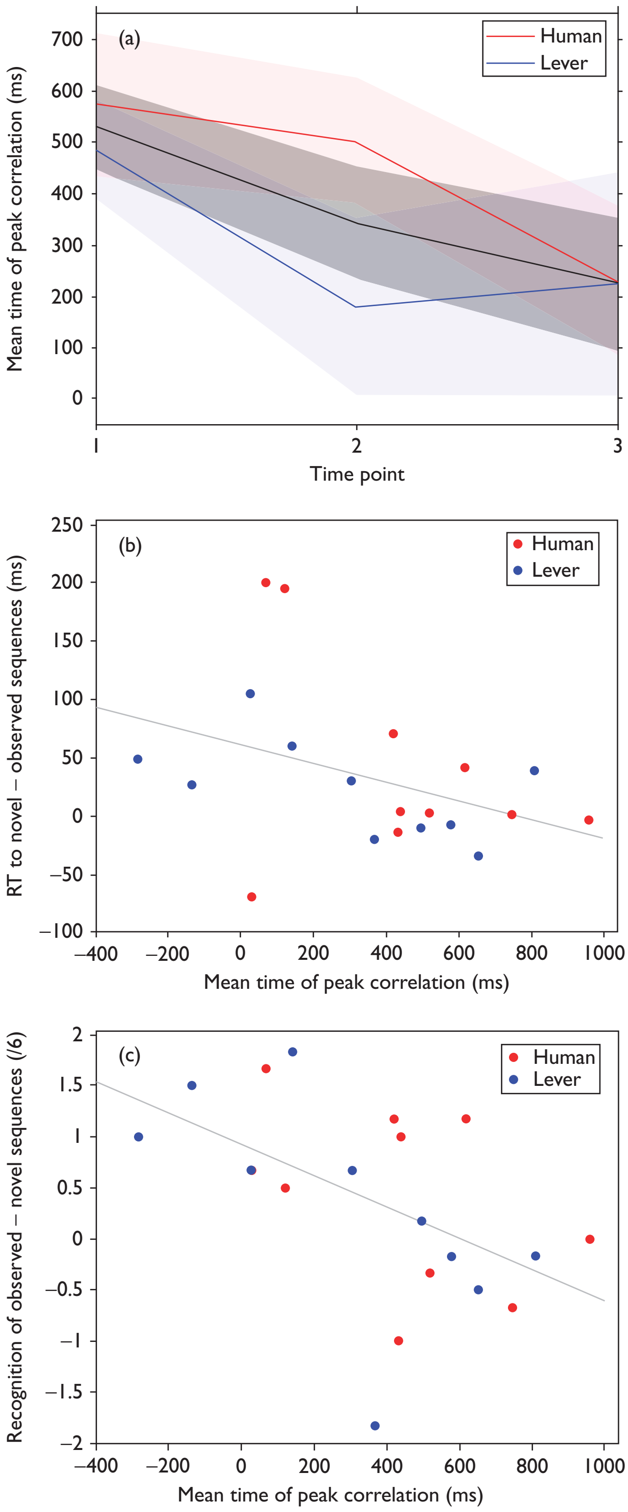


Fig. 2

Eyetracking data during the observation phase. (a) Time of peak correlation according to time point, across all participants (black), and separately for those who were instructed that stimuli were human (red) or lever (blue) generated. Positive values indicate reactive eye movements and negative values reflect predictive eye movements. Error bars indicate SEM. (b) and (c) Individual datapoints and regression lines representing the relationship between the mean time of peak correlation and observational sequence learning effects measured by the serial reaction time execution (b) and explicit recognition (c) tasks. Note that in (c), one of the datapoints for the human-instruction condition is partly masked by a datapoint for the lever-instruction condition (third from the left). RT, reaction time.

$P < 0.005$] such that eye movements became less reactive from the first through to the final third (Fig. 2a).

Execution serial reaction time task

Reaction times (RTs) over 1000 ms (1.5%) and errors (5.3%) were excluded from the analysis of RTs, and means were calculated for each participant for observed and novel sequences. Analysis of variance with sequence type as a within-participant factor and instruction as a between-participant factor showed that participants were faster to execute the observed sequences (mean = 404.1 ms, SEM = 16.9 ms) relative to the novel sequences [mean = 437.5 ms, SEM = 20.7 ms; $F(1,18) = 4.6$, $P < 0.05$], and made fewer errors [observed mean = 4.6%, SEM = 0.9%; novel mean = 5.9%, SEM = 0.9%; $F(1,18) = 5.8$, $P < 0.05$]. There was no difference in the number of RTs over 1000 ms according to sequence type [$F(1,18) = 1.4$, $P = 0.2$].

Explicit recognition task

Mean ratings for observed and novel sequences were calculated for each participant. There was a trend for participants to rate that they were more certain that they had previously seen the observed (mean = 3.9, SEM = 0.2) than the novel sequences [mean = 3.6, SEM = 0.1; $F(1,18) = 2.8$, $P = 0.1$].

Correlations between eye movements and sequence learning

The mean time of peak correlation correlated negatively with the extent of sequence learning, both as measured by the execution SRT task (mean RT on novel sequence blocks – mean RT on observed sequence blocks; Pearson's $r = -0.38$, $P < 0.05$, $N = 20$; Fig. 2b) and the recognition task (mean rating for observed sequences – mean rating for novel sequences; Pearson's $r = -0.52$, $P < 0.01$, $N = 20$; Fig. 2c). These findings indicate that those participants who made eye movements that were more predictive demonstrated greater sequence learning according to both measures.

Discussion

The present data indicate that when participants observed predictable, but arbitrary, sequences of point-light actions without objects, their eye movements became less reactive across time. Participants were subsequently faster to execute this sequence relative to a novel one, and there was a trend for greater explicit recognition of the observed sequences (for debates concerning the possibility that sequence learning can be implicit [8,9]). In addition, those participants exhibiting more predictive eye movements demonstrated greater learning of the sequence, reflected both in execution SRT and in explicit recognition measures. No measures were influenced by whether participants had been instructed that the observed movements were human or lever generated. These data provide support for the 'learned trajectory' hypothesis; that the time course of our eye movements when observing actions without objects is determined by the extent to which we have learned to anticipate the trajectory. Furthermore, these data indicate that eye tracking can provide useful information about learning during observation in observational learning paradigms, overcoming a difficult problem in such procedures.

Differential opportunities for learning the trajectories of object-directed and non-object-directed actions may explain the disparity in the time course of eye movements towards them [2]. Under this hypothesis, we have had much opportunity to learn that sight of a grasping hand configuration in the presence of an object will be followed by movement towards the object. Given appropriate experience, we can learn to anticipate action trajectories for non-object-directed actions in a similar way. If perception of the two action types is mediated by common mechanisms [10], there may be no intrinsic differences in eye movements between action types. It should be noted that the actions observed in the present study were object directed, but with the objects removed from the videos. The kinematics of object-directed and non-object-directed actions are likely to differ, but the reactive pattern of eye movements at the outset mirrors effects found with intransitive actions [2]. Therefore, the crucial feature that we propose was responsible for the pattern of eye movements in the present study is that participants initially had no cue upon which to determine the action that would come next and they subsequently learnt to make predictions throughout the observation phase. This logic would hold irrespective of the precise kinematics of action and also irrespective of whether objects are present or not.

The nature of the learning that generates less reactivity in eye movements could be purely perceptual (for the role of expectation and learning in perceptual processing [3,11,12]); participants learn that perception of one action element is followed by perception of another. Infants can learn across trials to make appropriate predictive eye movements when observing inanimate objects move [13–15], and evidence of this learning has been found before they are able to perform object-directed reaches [14,16]. Alternatively, and in line with assumptions about mechanisms underlying predictive eye movements towards object-directed actions [1,2,17], the mirror system may play a role. Much research shows that corresponding motor codes are activated when we observe actions [4,18], and predictive eye movements are demonstrated when we execute actions [19]. Consistent with this hypothesis, participants showed superior execution of the observed sequence relative to one which they had not observed, and such learning effects have been found to rely on motor encoding of sequences during observation. For example, observation of sequential manual actions leads to superior performance of the sequence only when participants are required to execute the same sequence of finger movements, rather than responding at the same sequence of spatial locations [5]. Furthermore, greater premotor activation has been found when observing a previously observed sequence [20], and repetitive transcranial magnetic stimulation over the primary motor cortex after the observation phase can abolish learning effects [21]. Under this hypothesis, as participants learn perceptually that one action element will be followed by another (within-modality learning [11,22]) the motor code of the subsequent element is activated in advance as well as the perceptual code [23], and the anticipatory activation of the motor codes underlies the pattern of eye movements [1].

Whether attribution of intention [2] is the optimal way of characterizing predictive eye movements during action observation is unclear. Clarification requires better characterizing the processes hypothesized to be implicated in such a function. There are at least two possibilities. First, intention ascription may only mean anticipation of what is going to happen next in the context of action observation and this prediction may even operate via similar mechanisms as that with inanimate objects [11]. Our results are consistent with this hypothesis. Second, and more in line with classic definitions, intention ascription may require an expectation that the action is executed by an agent to achieve certain higher level aims (e.g. 'to drink'; [24]). Our data are less consistent with this hypothesis. Eye movements towards arbitrary sequences change their time course solely with perceptual experience of what is coming next, and there were no differences in the pattern of eye movements in the present study between the groups instructed that the stimuli constituted finger or lever movements. To further assess this second version of the intention attribution hypothesis, future research must investigate whether there are differences in the time course of eye movements during and after learning when observing actions likely to achieve higher level aims, actions that appear arbitrary or 'meaningless', and equally predictable inanimate movements. Despite the many processes that will differ when observing these three stimulus categories, the mechanisms mediating the time course of eye movements may be similar.

Conclusion

The present data indicate that eye movements during observation of actions without objects reflect the extent to which we have learned to anticipate the perceptual trajectory, or sequence of events. Future research must establish whether similar mechanisms underlie the time course of eye movements when observing object-directed actions and inanimate object movement.

Acknowledgements

The authors are grateful to Chris Freemantle and Derek Tutssel for assistance in video generation, to Tim Smith for useful discussions about analysis and to Cecilia Heyes and Geoff Bird for comments on an earlier version of the manuscript. C.M.P. was funded by an MRC-ESRC postdoctoral fellowship and J.M.K. by a Wellcome Trust Research Career Development Fellowship.

Conflicts of interest

There are no conflicts of interest.

References

1 Flanagan JR, Johansson RS. Action plans used in action observation. *Nature* 2003; **424**:769–771.
2 Gredebäck G, Stasiewicz D, Falck-Ytter T, Rosander K, Von Hofsten C. Action type and goal type modulate goal-directed gaze shifts in 14-month-old infants. *Dev Psychol* 2009; **45**:1190–1194.
3 Summerfield C, Egner T. Expectation (and attention) in visual cognition. *Trends Cogn Sci* 2009; **13**:403–409.
4 Rizzolatti G, Craighero L. The mirror-neuron system. *Annu Rev Neurosci* 2004; **27**:169–192.
5 Bird G, Heyes C. Effector-dependent learning by observation of a finger movement sequence. *J Exp Psychol Hum Percept Perform* 2005; **31**: 262–275.
6 Johansson G. Visual perception of biological motion and a model for its analysis. *Percept Psychophys* 1973; **14**:201–211.
7 Shanks DR, Perruchet P. Dissociation between priming and recognition in the expression of sequential knowledge. *Psychon Bull Rev* 2002; **9**:362–367.
8 Destrebecqz A, Cleeremans A. Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychon Bull Rev* 2001; **8**:343–350.
9 Shanks DR, Wilkinson L. Intentional control and implicit sequence learning. *J Exp Psychol Learn Mem Cogn* 2004; **30**:354–369.
10 Fleischer F, Giese MA. Computational mechanisms of the visual processing of action stimuli. In: Johnson K, Shiffrar M, editors. *People watching: social, perceptual, and neurophysiological studies on body perception*. New York: Oxford University Press; 2013. pp. 388–415.
11 Kirkham NZ, Slemmer JA, Johnson SP. Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* 2002; **83**:B35–B42.
12 Kok P, Jehee JFM, De Lange FP. Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* 2012; **75**:265–270.
13 Johnson SP, Amso D, Slemmer JA. Development of object concepts in infancy: evidence for early learning in an eye-tracking paradigm. *Proc Natl Acad Sci USA* 2003; **100**:10568–10573.
14 Johnson SP, Shuwairi SM. Learning and memory facilitate predictive tracking in 4-month-olds. *J Exp Child Psychol* 2009; **102**:122–130.
15 Kochukhova O, Gredebäck G. Learning about occlusion: initial assumptions and rapid adjustments. *Cognition* 2007; **105**:26–46.
16 Konczak J, Borutta M, Topka H, Dichgans J. The development of goal-directed reaching in infants: hand trajectory formation and joint torque control. *Exp Brain Res* 1995; **106**:156–168.
17 Hunnius S, Bekkering H. The early development of object knowledge: a study of infants' visual anticipations during action observation. *Dev Psychol* 2010; **46**:446–454.
18 Kilner JM, Neal A, Weiskopf N, Friston KJ, Frith CD. Evidence of mirror neurons in human inferior frontal gyrus. *J Neurosci* 2009; **29**:10153–10159.
19 Johansson RS, Westling G, Bäckström A, Flanagan JR. Eye-hand coordination in object manipulation. *J Neurosci* 2001; **21**:6917–6932.
20 Cross ES, Kraemer DJM, Hamilton AFC, Kelley WM, Grafton ST. Sensitivity of the action observation network to physical and observational learning. *Cereb Cortex* 2009; **19**:315–326.
21 Brown LE, Wilson ET, Gribble PL. Repetitive transcranial magnetic stimulation to the primary motor cortex interferes with motor learning by observing. *J Cogn Neurosci* 2008; **21**:1013–1022.
22 Heyes C. Causes and consequences of imitation. *Trends Cogn Sci* 2001; **5**:253–261.
23 Press C, Cook J, Blakemore SJ, Kilner J. Dynamic modulation of human motor activity when observing actions. *J Neurosci* 2011; **31**:2792–2800.
24 Rizzolatti G, Sinigaglia C. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat Rev Neurosci* 2010; **11**:264–274.