

2012-11

Doing, seeing, or both: Effects of learning condition on subsequent action perception

Wiggett, AJ

<http://hdl.handle.net/10026.1/17634>

10.1080/17470919.2012.686926

Social Neuroscience

Informa UK Limited

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.



Social Neuroscience

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/psns20>

Doing, seeing, or both: Effects of learning condition on subsequent action perception

Alison J. Wiggett^a, Matthew Hudson^b, Angela Clifford^c, Steven P. Tipper^a & Paul E. Downing^a

^a School of Psychology, Bangor University, Wales, UK

^b Universidad del Norte, Barranquilla, Colombia

^c School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK

Available online: 17 May 2012

To cite this article: Alison J. Wiggett, Matthew Hudson, Angela Clifford, Steven P. Tipper & Paul E. Downing (2012): Doing, seeing, or both: Effects of learning condition on subsequent action perception, *Social Neuroscience*, DOI:10.1080/17470919.2012.686926

To link to this article: <http://dx.doi.org/10.1080/17470919.2012.686926>



PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Doing, seeing, or both: Effects of learning condition on subsequent action perception

Alison J. Wiggett¹, Matthew Hudson², Angela Clifford³, Steven P. Tipper¹,
and Paul E. Downing¹

¹School of Psychology, Bangor University, Wales, UK

²Universidad del Norte, Barranquilla, Colombia

³School of Sport, Exercise and Health Sciences, Loughborough University, Loughborough, UK

It has been proposed that common codes for vision and action emerge from associations between an individual's production and simultaneous observation of actions. This typically first-person view of one's own action subsequently transfers to the third-person view when observing another individual. We tested vision–action associations and the transfer from first-person to third-person perspective by comparing novel hand-action sequences that were learned under three conditions: first, by being performed and simultaneously viewed from a first-person perspective; second, by being performed but not seen; and third, by being seen from a first-person view without being executed. We then used functional magnetic resonance imaging (fMRI) to compare the response to these three types of learned action sequences when they were presented from a third-person perspective. Visuomotor areas responded most strongly to sequences that were learned by simultaneously producing *and* observing the action sequences. We also note an important asymmetry between vision and action: Action sequences learned by performance alone, in the absence of vision, facilitated the emergence of visuomotor responses, whereas action sequences learned by viewing alone had comparably little effect. This dominance of action over vision supports the notion of forward/predictive models of visuomotor systems.

Keywords: Mirror systems; Action observation; Learning.

The learning of new actions and skills takes place throughout the entire life span: from babies learning to reach for and grasp objects, to children learning to write or play an instrument, to adults learning to drive a car. Moreover, as actions often take place in a social context, we must also be able to understand what other people are doing and whether their actions are the same as or different from our own actions: for example, when trying to learn a new dance sequence by watching somebody else perform it. This ability requires a system that is able to abstract across different viewpoints of actions—from first-person views of our own actions to third-person views of other people.

This study investigated the necessary conditions for building action representations that abstract across views in this way, and the brain regions that support these representations.

Behavioral and neuroimaging studies have suggested a strong link between the observation of others' and the execution of our own actions. For example, observing another person perform an action facilitates our execution of that action (Bach, Peatfield, & Tipper, 2007; Bach & Tipper, 2007; Brass, Bekkering, Wohlschläger, & Prinz, 2000). This priming effect is thought to reflect the activation of specific action representations by observation—the same representations

Correspondence should be addressed to: Alison J. Wiggett, School of Psychology, Adeilad Brigantia, Bangor University, Bangor, Gwynedd LL57 2AS, Wales, UK. E-mail: a.wiggett@bangor.ac.uk

We thank P. Bach and E. Cross for helpful comments on an earlier draft of this manuscript, and P. Mullins and S. Johnston for technical support. This work was supported by ESRC RES-062-23-0980 to PED and SPT.

that are activated when a person actually performs the action. This idea is supported by studies using transcranial magnetic stimulation (TMS) that have shown specific facilitation of relevant motor pathways during action observation (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Romani, Cesari, Urgesi, Facchini, & Aglioti, 2005; Strafella & Paus, 2000). Furthermore, functional magnetic resonance imaging (fMRI) studies have found that areas of the parietal and premotor cortices are activated by both the perception and the execution of actions (e.g., Buccino, Vogt, et al., 2004; Dinstein, Hasson, Rubin, & Heeger, 2007; Etzel, Gazzola, & Keysers, 2008). Although these areas are commonly thought to be the homologue of “mirror neurons” found in macaques (neurons that fire both when the animal executes a specific action and when it perceives the experimenter performing the same action [Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992]), there is still debate about the neuroimaging evidence for a common code of vision and action in the human brain (Dinstein, Gardner, Jazayeri, & Heeger, 2008; Hickock, 2009; Kilner, Neal, Weiskopf, Friston, & Frith, 2009; Lingnau, Gesierich, & Caramazza, 2009; Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010; Oosterhof, Wiggett, Diedrichsen, Tipper, & Downing, 2010). Below, we review some of the behavioral and neuroscientific literature demonstrating strong links between the observation and the execution of actions, with a particular focus on the role of learning and experience.

The ability to perform a specialized set of motor skills results in increased perceptual accuracy when observing and discriminating the corresponding actions. Aglioti, Cesari, Romani, and Urgesi (2008), for example, compared elite basketball players to basketball coaches and sports journalists, and found that the basketball players were faster and more accurate at predicting whether a shot was going to be on target. (Using TMS, the authors also found increased motor excitability in the hand of the basketball players when observing shots that resulted in misses.) Calvo-Merino, Ehrenberg, Leung, and Haggard (2010) looked at the visual discrimination of biological motion using point-light ballet moves. The authors found that expert ballet dancers were significantly better at discriminating between different exemplars of the same movement compared to nonexperts. Similar effects have also been shown for newly learned actions. Casile and Giese (2006) investigated the visual recognition of gait patterns before and after nonvisual motor training. Participants learned a novel coordinated upper-body movement pattern that consisted of swinging the arms 270° out of phase with the legs (as opposed to the usual 180° seen in natural

walking). After training, the participants completed a same/different matching task with point-light walkers presented at different phase shifts. The authors found selective improvement of the visual recognition performance for learned movements. Finally, a study by Serino et al. (2010) found comparable effects in hemiplegics. The patients viewed and named arm gestures presented as point-light displays. Action recognition was significantly better for actions performed by their unaffected arm compared to actions that appeared to have been performed by their hemiplegic arm. These results support the idea of a significant motor contribution to visual action processing.

Further evidence for the action perception link in the context of expertise and familiarity has come from recent neuroimaging studies. These have shown that neural activity in visuomotor areas tends to be stronger for familiar compared to unfamiliar actions. Calvo-Merino, Glaser, Grèzes, Passingham, and Haggard (2005), for example, scanned ballet and capoeira dancers and nondancers while they watched videos of ballet and capoeira dance moves. The authors found greater activation in premotor, inferior parietal lobe (IPL), right superior parietal lobe, and left superior temporal sulcus (SPS) for movements that were in the observer's motor repertoire. In a further study, Calvo-Merino, Grèzes, Glaser, Passingham, and Haggard (2006) compared male and female ballet dancers. As ballet moves are only performed by one gender, the authors were able to test specifically for motor familiarity while controlling for visual familiarity. The authors found greater premotor, parietal, and cerebellar activities when viewing moves from own motor repertoire. A study by Cross, Hamilton, and Grafton (2006) tested expert dancers as they learned new dance moves over a 5-week period. While in the scanner, participants watched videos of the dance moves and were instructed to imagine dancing the presented movements. The results showed that the IPL and the ventral premotor areas were modulated as a function of the dancers' rating of ability and motor experience. Further studies by Cross and colleagues (Cross, Hamilton, Kraemer, Kelley, & Grafton, 2009; Cross, Kraemer, Hamilton, Kelley, & Grafton, 2009) have shown similar effects in nonexpert dancers for newly learned dance moves. Taken together, these studies point to a significant role of “motor resonance” in perceptual effects of expertise.

A key question is how the associations between sensory inputs and motor outputs are learned in the first place. In Keysers and Perrett's (2004) Hebbian model, it is the repeated experience of watching oneself perform an action that plays a critical role in the emergence of mirror systems. Executing and

simultaneously observing an action causes associations to be made between the motor output and the visual input, such that simply viewing an action becomes sufficient to activate motor representations of that action. Furthermore, the system learns to generalize across viewpoints (e.g., when viewing someone else perform an action) such that actions come to be represented visually in both first-person and third-person coordinates. The associative sequence learning (ASL) model by Heyes and colleagues (Heyes, 2001, 2005; Heyes, Bird, Johnson, & Haggard, 2005; Heyes & Ray, 2000) likewise emphasizes the importance of correlated sensorimotor experience. However, a crucial difference between the ASL model and the Hebbian model is that the latter assumes that contiguity is sufficient for the associations between vision and action to be learned (i.e., the execution and observation of an action occurring together in time), whereas the ASL model emphasizes the importance of contingency (i.e., a predictive relationship between the two events) for the development of a “mirror” system (Cook, Press, Dickinson, & Heyes, 2010). Furthermore, the correlated experience of observing and executing the same actions can arise not just from self-observation, but also from mirror exposure or from performing synchronous actions with other people (e.g., being imitated by adults). These different forms of correlated sensorimotor experience are also likely to be critical for learning to recognize actions performed by other people, and to match those actions to our own.

The aim of this study was to test whether the perception of others’ actions elicits differential cortical activity as a function of the visual and/or motor conditions under which those actions were learned. If, as outlined above, correlated sensorimotor experience is crucial for learning a rich representation of actions that can be generalized to perceiving others, then learning to perform an action while simultaneously seeing oneself perform it should lead to significant “mirror system” activation (vision + action condition). This can be compared to learning conditions in which motor output and visual input are not correlated, such as where actions are learned “blind,” without visual feedback (action condition), or where actions performed by someone else are passively viewed (vision condition). In this study, we tested this hypothesis by training participants to learn novel sequences of hand movements under these three conditions. The sequences consisted of three hand postures (see Method for details). While the individual hand postures were probably not novel to the participants (i.e., it is likely that the participants had previously experienced the individual motor states of the postures), the specific sequence of these motor states was novel and it is

these action sequences that are the unit of assessment in this study. Participants were subsequently scanned while passively viewing movies of another person performing the learned action sequences, as well as novel action sequences for which the participants had neither visual nor motor experience. We predicted differential activations in visuomotor areas, in the form of greater response to the action sequences that were learned in the “vision + action” condition compared to the single modality conditions of vision or action. This would support the sensorimotor learning hypothesis.

An important aspect of this study is that during learning, the visual input from the action sequences (where present) was perceived in the normal first-person perspective. That is, the hand movements were shown at the viewing angle and orientation from which we normally perceive our own movements. At test, however, the action sequences were shown in the opposite orientation—the normal view from which we see others’ actions. This is important to test the predictions outlined earlier: Correlated motor and visual input can produce action learning that generalizes to the perception of other individuals. Previous fMRI studies have either not controlled for how the actions were learned as the actions were in the participants’ repertoire before the experiment (e.g., Calvo-Merino et al., 2005, 2006) or presented “third-person” views of the to-be-learned actions (i.e., movies of another person performing the action) in both the training and test phases (Cross, Hamilton, et al., 2009; Cross, Kraemer, et al., 2009). Furthermore, previous studies emphasizing the importance of correlated sensorimotor experience for mirror learning (e.g., Catmur et al., 2008; Catmur, Walsh, & Heyes, 2007) have not directly tested for this transfer from first-person learning to third-person observation. Thus, the current fMRI study is novel in two important ways: (i) It includes multiple types of learning (action only, vision only, and vision + action) as well as a novel (untrained) condition and (ii) it tests whether the learning of action sequences transfers from a first-person to a third-person perspective.

METHOD

Participants

Twelve participants (mean age = 24.5 years; range = 20–32 years; 7 women, 5 men) were all students at Bangor University and took part in exchange for £25. All participants gave written informed consent prior to taking part, and the study procedures were approved by the Ethics Committee of the School of Psychology, Bangor University.

Stimuli and apparatus

The stimuli used were 12 novel action sequences, each consisting of three hand postures. (Note that several of the hand postures were taken from the American Sign Language alphabet. However, all participants tested were from the United Kingdom and were unfamiliar with the hand postures.) Each individual hand posture (performed by one of the experimenters) was photographed and printed (example shown in Figure 1); these were used in the instruction phase described below.

Each action sequence was also performed by one of the experimenters and filmed from two different perspectives: a first-person and a third-person perspective (Figure 2). This provided 24 movies in total (12 action sequences \times 2 perspectives) each of 4 s duration. The first-person movies were used in the training session; the third-person movies were converted to gray scale and used as visual stimuli during the fMRI scan (see below). To minimize the visual differences between the participants' and the experimenter's hands, the experimenter performing the actions wore a yellow glove, which was also worn by the participants during the training sessions.

Training

A computer monitor was positioned at the edge of a desk and the participant was seated closely in front of the monitor with his or her hand positioned behind the monitor (Figure 3). The input to the monitor could be switched between one of two sources. The input came either from a camera (MRC Systems GmbH, Heidelberg, Germany) placed immediately behind the monitor, facing away from the participant and pointing downward toward the participant's hand at an angle of approximately 45° . When participants positioned their hand behind the monitor, they were able to view their hand from a first-person perspective, that is, as if the

monitor was not there and they were looking directly at their hand through a frame. Alternatively, the monitor input came from a computer (iMac G4, Apple, Cupertino, CA) that was operated by the experimenter and was used to display the prerecorded stimuli of the experimenter performing the action sequences.

Design and procedure

The 12 action sequences were randomly split into four groups of three sequences. Each set of action sequences was assigned to one of four training conditions. This allocation was counterbalanced across participants such that for 3 out of the 12 participants the same three action sequences were assigned to the same condition.

Training

Prior to scanning, participants completed two training sessions in which they learned 9 of the 12 action sequences. Each session lasted approximately 40 min. The two training sessions took place 1 day apart, with the scanning session taking place immediately after the second training session. During training, participants wore a yellow rubber glove on their right hand. This was done in order to minimize low-level visual differences between conditions (see below).

The visual input and/or motor output experienced during the training phase varied across three experimental training conditions. For three of the action sequences, participants executed the sequence while being able to view their movements in real time on the computer monitor in front of them (with input provided by the camera). The encoding and learning of these action sequences therefore involved both visual stimulation and motor activation (vision + action condition). A further three action sequences were learned by performing the sequence without receiving any visual feedback. Thus, the participants'



Figure 1. Example hand action. The action progressed from posture 1 to posture 2 to posture 3. See Appendix for a complete list of hand postures/actions used.

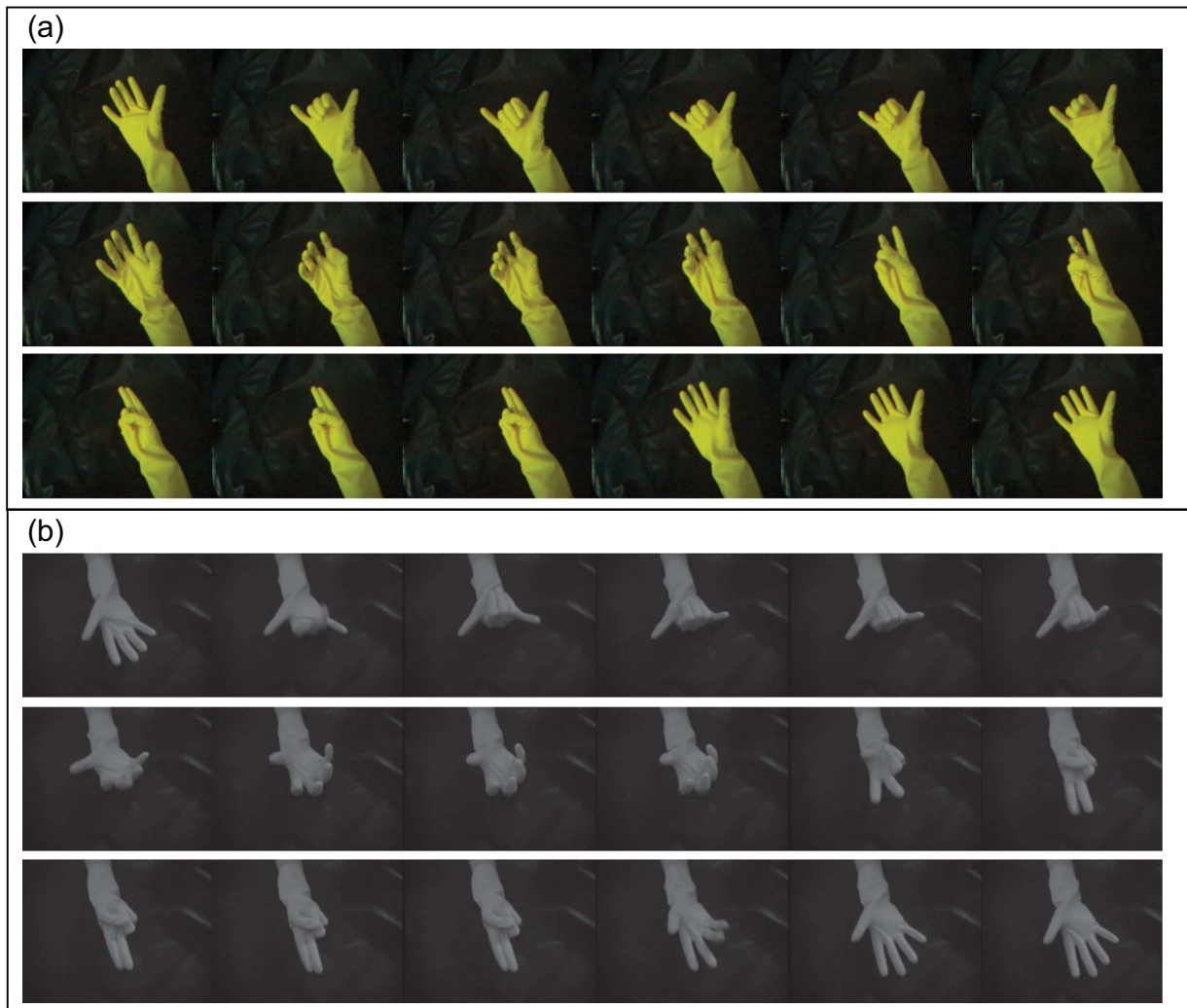


Figure 2. An example of one of the actions presented shown from a first-person perspective (a) and a third-person perspective (b). Movies like the one shown in (a) were presented in the vision-only condition of the training phase; movies shown in (b) were shown during scanning. The example shows a subset of frames (18 out of 120) from the full movie.

movements were not presented on the screen. Instead, a yellow fixation cross was presented on a black background. The encoding and learning of these action sequences was therefore from motor output only, and the visual stimulation was not related to the action sequence being performed (action condition). The next three action sequences were learned by observation only. Participants were required to keep their hand motionless palm down on the table behind the monitor and simply watch the movies presented on screen. These were the prerecorded video clips of the action sequences from a first-person perspective performed by the experimenter. Videos were presented using iMovie. Encoding and learning of these action sequences was therefore from visual input only (vision condition). The remaining three action sequences were not presented to the participants during the

training session and were only used in the scanning session. These action sequences were therefore completely novel to the participant during the scan (novel condition).

The order of training conditions (vision + action, action, and vision) was counterbalanced within and across participants. Each of the conditions was encountered one after the other (e.g., vision followed by action followed by vision + action). This was repeated 3 times with a different action sequence in each repetition, such that each of the nine action sequences was encountered once. This block of nine action sequences was repeated 6 times over the two training sessions (three in each). In each block, the order in which the conditions were encountered was counterbalanced, such that each of the six possible combinations was presented once in each of the three

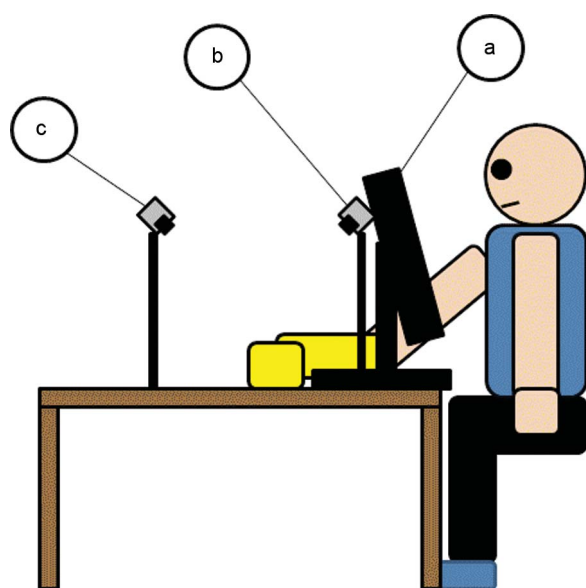


Figure 3. Experimental setup. During the training session, participants sat close to the desk and placed their right hand behind the monitor (a). They were able to view their hand from a first-person position on screen from input provided by a camera in position (b). Camera position (c) was not used during training; it was used only to record the movies of the hand actions (performed by an experimenter) filmed from a third-person perspective that were used in the scanner.

repetitions (e.g., $3 \times [V + A, V, A]$, $3 \times [V + A, A, V]$, $3 \times [V, A, V + A]$, $3 \times [V, V + A, A]$, $3 \times [A, V + A, V]$, $3 \times [A, V, V + A]$). Furthermore, the order in which these six combinations of the training conditions were presented was counterbalanced across participants. Overall, each of the 9 action sequences was encountered 6 times, resulting in 54 trials (27 in each session). These counterbalancing measures resulted in the action sequences being presented in a pseudorandom order.

At the start of each trial, the participant was shown pictures of the three hand postures that constituted the to-be-learned action sequence (example shown in Figure 1). The participant was asked to memorize the action sequence only by looking at the pictures. He or she was not permitted to rehearse the action sequence, and the experimenter did not demonstrate the action sequence to the participant. This was to ensure that the role of visual input and motor output during the encoding of the actions was carefully controlled as per the designated experimental conditions. No time restrictions were placed on the memorization part. Once the participant felt he had memorized the action sequence, the trial began. In the action and vision + action conditions, participants performed the action sequence 8 times. The pace at which it was executed was slow and rhythmic, with approximately 1 s

between each of the separate postures. A rest posture (flat hand) separated each performance of the action sequence. If the participant forgot the action sequence or was not performing it to the desired level of accuracy, he or she was asked to look at the pictures again. However, as before, neither the participant nor the experimenter performed the action sequence while the participant was looking at the pictures. When the participant began the action sequence again, the count continued from before.

For the vision + action and action conditions, participants were told that no feedback from the experimenter was indicative of accurate performance of the action sequence. Only when the participant forgot the action or was not performing it to the desired level of accuracy was feedback given. Thus, feedback took the form of correcting errors rather than reinforcing accurate execution of the action sequence. As much feedback as necessary was given to ensure that the action sequence was executed correctly 8 times during a single block. Therefore, although we did not record the amount of feedback given or the number of times the participant needed to look at the action pictures again, we were confident that the number of correct performances was the same for all action sequences across both the vision + action and the action condition.

The video clips in the vision condition depicted the action sequence being performed at the same pace with a black screen of 1 s duration separating each individual performance of the action sequence (eight repetitions). Each action sequence was therefore either seen or performed 48 times in total.

fMRI scan

Participants were scanned immediately after the second training session. Three runs investigated the central hypothesis that the perception of others' actions elicits differential cortical activity in premotor and parietal areas as a function of the visual/motor conditions under which those actions were learned. Participants passively viewed the movies of all the 12 prerecorded action sequences, filmed from a third-person perspective (as shown in Figure 2b). Passive viewing was selected to avoid requiring participants to select and make motor responses that could contaminate the activation of the regions of interest.

Each of the three runs contained 16 blocks (see Figure 4a), consisting of 4 blocks per condition (vision + action, action, vision, novel). Within each block, four movies (each 4 s, 30 frames/second) were presented: for example, four movies of action sequences learned in the vision + action condition, followed by

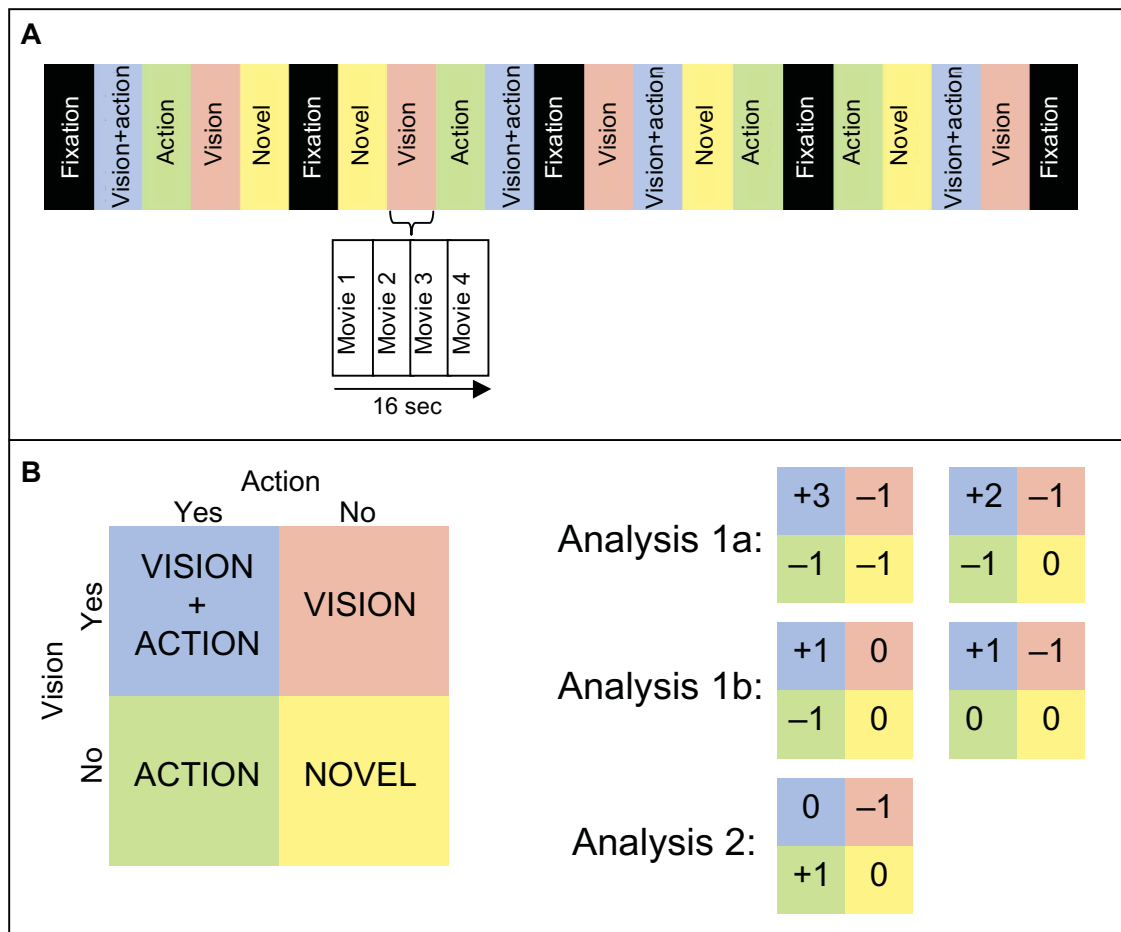


Figure 4. Illustration of the blocked design paradigm (a) and the design and summary of analyses (b).

four from the action condition, followed by four from the vision condition, followed by four movies from the novel condition (one of the three actions per condition was presented twice in each block). There were four designs of the experimental runs in which the presentation order of the action sequences from the different training conditions were counterbalanced across each of the blocks. Furthermore, the presentation of each block was counterbalanced across the designs. This ensured that the order in which the training conditions were presented was counterbalanced across the blocks within a run, across the runs within a participant, and across participants. Each major block was separated by a 16 s baseline fixation period that was also presented at the start and end of the run. In total, each run lasted 5.6 min ($4 \text{ s} \times 4 \text{ movies} \times 4 \text{ conditions} \times 4 \text{ blocks} + 5 \text{ fixation baseline periods}$).

Data acquisition. The data were acquired using a 3T Philips MRI scanner with a SENSE phased-array head coil. For functional imaging, a single-shot echo planar imaging sequence was used (T2*-weighted,

gradient echo sequence; echo time, 35 ms; flip angle, 90°). The scanning parameters were as follows: repetition time 2000 ms; 31 off-axial slices; voxel dimensions $3 \text{ mm} \times 3 \text{ mm} \times 3 \text{ mm}$, field of view (FOV) 240×240 , matrix 80×80 , phase encoding direction A-P. Seven dummy scans were acquired before each functional scan to reduce possible effects of T1 saturation. Parameters for T1-weighted anatomical scans were 288×232 matrix; 1 mm isotropic voxels; TR = 8.4 ms, TE = 3.8 ms; flip angle = 8° .

Data analysis. Preprocessing and statistical analyses of the MRI data were performed using BrainVoyager QX 1.9 (Brain Innovation, Maastricht, The Netherlands). Functional data were motion corrected, and low-frequency drifts were removed with a temporal high-pass filter (0.006 Hz). The data from the main experiment were spatially smoothed using a 6 mm full-width-half-maximum kernel. Functional data were manually co-registered with the anatomical T1 scans. The three-dimensional anatomical scans were transformed into Talairach space (Talairach &

Tournoux, 1988), and the parameters from this transformation were subsequently applied to the co-registered functional data.

For each participant, general linear models (GLMs) were created. One boxcar predictor, convolved with a two-gamma hemodynamic response function (HRF) function to model the hemodynamic response, modeled each condition of interest. Regressors of no interest were also included to account for differences in the mean MR signal across scans. Regressors were fitted to the MR time series in each voxel, and the resulting beta parameter estimates were used to estimate the magnitude of response to each of the conditions in the main experiment.

RESULTS

Did learning in the “vision + action” condition lead to the most widespread activation in visuomotor areas?

The central question of this study was to test whether having motor and visual input during the learning of new action sequences compared to action-only or visual-only learning subsequently led to greater brain activity when viewing someone else perform the same actions. The prefrontal and parietal regions of the brain were of particular interest in this and subsequent analyses.

A whole-brain random-effects analysis for the contrast vision + action > [action, vision, novel]

(+3,−1,−1) was carried out (Analysis 1a in Figure 4). This revealed activation in the frontal and parietal areas mainly in the left hemisphere (see Figure 5), including the inferior frontal gyrus (IFG), the postcentral gyrus, and the superior temporal sulcus. Activations significant at a cluster size threshold of $p < .05$ (corresponding to 5 voxels; uncorrected threshold $p < .005$) are further listed in Table 1. Overall, these analyses reveal more widespread activation when the participants had previous simultaneous visual and motor experience of the action sequences they were passively viewing in the scanner.

To rule out the possibility that this result was due to the novel condition not activating these areas at all, rather than the vision + action condition activating them more than the other learning conditions, we repeated the analysis leaving out the novel condition. The contrast vision + action > [action, vision] (+2,−1,−1,0) resulted in many of the same activations as (+3,−1,−1,−1), with the exception of the activation in the left precuneus and the right cingulate (see Table 1; cluster size threshold of $p < .05$; corresponding to 5 voxels; uncorrected threshold $p < .005$). Additionally, the new analysis resulted in two activations in the superior temporal sulcus (one in each hemisphere) and a further activation in the left precentral gyrus. Importantly, all activations in the core regions of interest (prefrontal and parietal visuomotor regions) were significant whether the novel condition was included in the analysis or not.

To further test whether action + vision led to greater activation than either action alone or vision

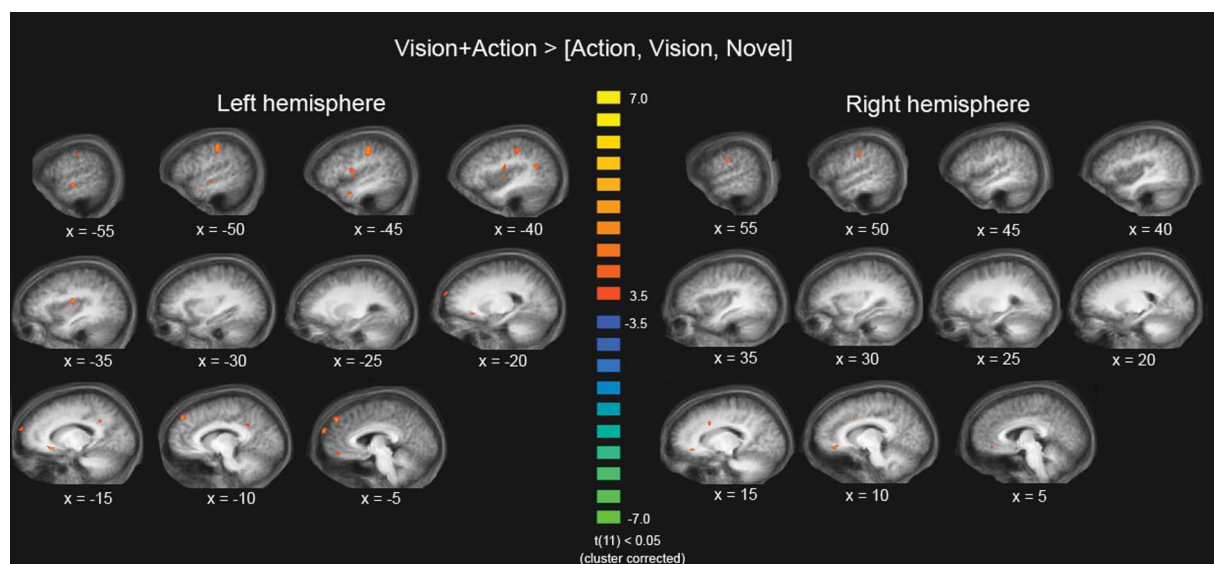


Figure 5. Activations from the whole-brain random-effects group-average analysis of vision + action > (action, vision, novel) overlaid on the average anatomical scan from all participants. Activations significant at a cluster size threshold of $p < .05$ (corresponding to 5 voxels) are shown. Sagittal slices of both hemispheres are shown in 5 mm increments from $x = \pm 55$ to $x = \pm 5$.

TABLE 1
Summary of *t*-test whole-brain analysis for vision + action > [action, vision, novel] (+3,-1,-1,-1)

Region	Extent (mm ³)	Peak coordinates			Max (<i>t</i>)
		X	Y	Z	
<i>Vision + action vs. (action, vision, novel)</i>					
L postcentral gyrus ¹	1379	−45	−25	43	5.75
L superior frontal gyrus ¹	682	−3	65	26	6.24
L superior frontal gyrus ¹	534	−6	48	40	6.54
L inferior frontal gyrus ¹	407	−48	−1	10	5.05
L insula ¹	401	−36	−10	16	6.24
L posterior superior temporal sulcus ¹	343	−42	−54	20	5.03
L anterior superior temporal sulcus ¹	260	−51	−16	−4	5.37
L orbital gyrus ¹	200	−15	20	−7	5.68
L precuneus	180	−15	−48	31	4.31
L middle temporal gyrus ¹	175	−48	0	−20	5.33
R superior frontal gyrus ¹	559	15	36	−5	5.59
R postcentral gyrus ¹	307	54	−16	29	5.56
R cingulate gyrus	165	12	11	34	5.31

Note: ¹Activations also significant for vision + action > [action, vision] (+2,-1,-1,0).

alone, we performed two additional simple contrasts (Analysis 1b in Figure 4). The contrast vision + action > action resulted in a number of activations mostly in the superior frontal gyrus and the IFG in the left hemisphere. The vision + action > vision alone resulted in widespread activation, again mainly in the left hemisphere. This included activations in ventral premotor cortex, motor cortex, precentral sulcus, inferior parietal sulcus, superior temporal sulcus, and inferior temporal sulcus. Activations significant at a cluster size threshold of $p < .05$ (corresponding to 5 voxels, uncorrected threshold $p < .005$) are listed in Table 2.

Is there a difference between (unimodal) action and vision?

The above analyses suggest that the pattern of activation found for action only is more similar to action + vision than that found for vision only. This would suggest that it is specifically motor input during learning that is crucial for the subsequent activation of visuomotor areas during the viewing of third-person actions. However, so far we have not examined the effects of either type of learning by itself. Following on from the previous analyses, we predict that learning in the action-only condition will lead to more activation in visuomotor areas than learning in the vision-only condition. To test this, a whole-brain random-effects analysis for the contrast action > vision was carried out (Analysis 2 in Figure 4). Although this revealed less widespread activation than the previous analyses,

learning in the action condition led to subsequent significant activations, whereas learning in the vision condition did not. The activations were primarily in the intraparietal sulcus (IPS) and in the cerebellum in both hemispheres (see Figure 6). Activations significant at a cluster size threshold of $p < .05$ (corresponding to 5 voxels, uncorrected threshold $p < .005$) are also listed in Table 3. The analysis revealed no areas that were more active for vision compared to that for action. Overall, this analysis supports the notion that having previous motor experience of the actions is crucial for the engagement of visuomotor brain areas during passive viewing of third-person actions, and furthermore that motor experience alone is sufficient for the engagement of some of these areas.

DISCUSSION

The aim of this study was to test whether the brain's response to viewing actions is modulated by the conditions under which the actions were previously learned. Learning with both vision and action cues, relative to either cue alone, resulted in significant activation of the frontal and parietal regions mainly in the left hemisphere that have been repeatedly implicated in action production and perception. This finding supports the idea that correlated motor and visual activities during action learning lead to the creation of stronger representations in these fronto-parietal regions. In addition to that result, we also found evidence for increased engagement of parietal

TABLE 2
Summary of *t*-test whole-brain analysis for vision + action > action (+1,-1,0,0), and vision + action > vision (+1,0,-1,0)

Region	Extent (mm ³)	Peak coordinates			Max (t)
		X	Y	Z	
<i>Vision + action vs. action</i>					
L superior frontal gyrus	698	-6	38	46	6.12
L insula	195	-36	-10	19	5.39
L superior frontal sulcus	186	-18	26	37	6.67
L middle frontal gyrus	172	-33	11	52	5.17
L middle frontal gyrus	160	-24	62	13	4.29
L precentral sulcus	146	-42	8	46	5.01
R cingulate gyrus	260	9	-4	37	5.09
R lateral sulcus	146	27	-25	7	4.83
<i>Vision + action vs. vision</i>					
R postcentral sulcus	1399	51	-16	31	5.82
R medial occipitotemporal gyrus	205	39	-25	-8	5.84
R cerebellum	199	33	-40	-23	5.06
R cerebellum	142	33	-52	-23	4.73
L postcentral sulcus	4542	-42	-28	43	7.93
L precentral sulcus	530	-21	-10	52	5.36
L superior frontal gyrus	497	-9	-44	43	6.77
L superior temporal sulcus	484	-33	-61	22	6.52
L thalamus	447	-12	-13	-5	7.25
L superior temporal sulcus	433	-45	-37	1	4.65
L cingulate gyrus	376	-15	-1	31	6.42
L inferior frontal gyrus	356	-36	14	22	5.79
L inferior temporal sulcus	286	-51	-70	1	5.68
L lateral sulcus	142	-36	10	16	5.23

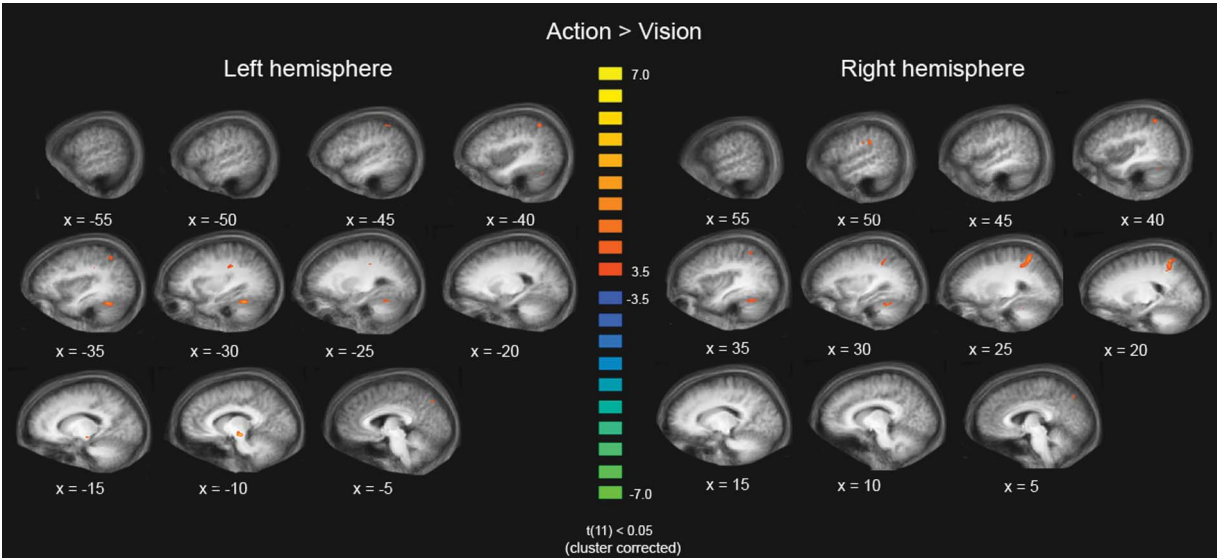


Figure 6. Activations from the whole-brain random-effects group-average analysis of action > vision overlaid on the average anatomical scan from all participants. Activations significant at a cluster size threshold of $p < .05$ (corresponding to 5 voxels) are shown. Sagittal slices of both hemispheres are shown in 5 mm increments from $x = \pm 55$ to $x = \pm 5$.

TABLE 3
Summary of *t*-test whole-brain analysis for action > vision

Region	Extent (mm ³)	Peak coordinates			Max (t)
		X	Y	Z	
<i>Action > vision</i>					
L cerebellum	817	−33	−49	−20	5.95
L intraparietal sulcus	359	−42	−52	46	4.71
L thalamus	315	−9	−19	−8	7.86
L cuneus	293	−3	−72	40	5.74
L intraparietal sulcus	234	−30	−30	31	4.81
R intraparietal sulcus	1412	21	−49	49	6.05
R cerebellum	518	33	−46	−23	4.55
R supramarginal gyrus	248	48	−28	31	4.78
R intraparietal sulcus	217	39	−49	49	5.24

regions by action sequences for which the participant had only prior motor experience compared to action sequences learned visually.

The activations we found included IFG and/or IPS. (Note that the contrasts we performed are not entirely orthogonal with respect to each other, and so it is not meaningful to directly compare the resulting activation maps, e.g., in order to identify overlapping regions.) These regions correspond to the core regions of the so-called human mirror system (Rizzolatti & Craighero, 2004), which are commonly activated during action observation as well as action execution (Buccino, Vogt, et al., 2004; Chong, Cunnington, Williams, Kanwisher, & Mattingley, 2008; Dinstein et al., 2007, 2008; Gazzola & Keysers, 2009; Grezes & Decety, 2001; Hamilton & Grafton, 2006, 2008; Héту, Mercier, Eugène, Michon, & Jackson, 2011; Kilner et al., 2009; Lingnau et al., 2009). The regions responding to action observation seen in fMRI studies also typically include the superior temporal sulcus (Buccino et al., 2001; Cross et al., 2006; Cross, Kraemer, et al., 2009; Grezes & Decety, 2001) and the cerebellum (Buccino, Vogt, et al., 2004; Calvo-Merino et al., 2006; Cross, Kraemer, et al., 2009), which we also identified in this study.

The present findings advance the fMRI literature on the influence of motor repertoire on visual action processing. Previous studies have found stronger activations of fronto-parietal areas when observing actions that are in the observers' repertoire, for example, observing conspecifics versus nonconspecifics (Buccino, Lui, et al., 2004), or male versus female ballet dancers (Calvo-Merino et al., 2006). Furthermore, Cross and colleagues have shown that these expertise effects can be built *de novo* and that the activity evoked by actions is modulated not only by motor

experience but also by subjective ratings of ability to perform an action (Cross et al., 2006). In a further study, Cross and colleagues addressed the question of whether "mirror" activation was modulated by how observed actions were previously learned (Cross, Kraemer, et al., 2009). They compared physical to observational learning (the to-be-learned actions were dance sequences that participants either watched and performed, or just watched). Although both types of learning led to similar brain activations in premotor and inferior parietal regions, directly comparing the observation of danced sequences to watched sequences resulted in two additional premotor activations. To our knowledge, this study is the first fMRI study to compare all three types of learning (vision + action, vision, action), and our results lend further support to the proposal that physical learning results in relatively greater involvement of visuomotor brain areas than visual learning.

It should be noted that the individual hand postures of the action sequences used here may well have been experienced previously, and it is likely that even the novel condition activated visuomotor systems to some extent. However, we would argue that actions are made of sequences of discrete motor states, such as when a hand moves through space to grasp and manipulate a tool. Similarly, in previous studies using novel dance sequences, for example, Cross et al. (2006) and Cross, Kraemer, et al. (2009), it is likely that the various individual body postures had been previously experienced. However, it is the sequence of actions that is the goal achieved during dance, and it is this sequence that is novel. Similarly, therefore, our unit of assessment was the sequence of action states, and these sequences were almost certainly novel for our participants.

An important novel aspect of this study was that we tested for transfer from first-person learning to third-person observation. In the training session, the visual input (where present) was seen from a first-person perspective, whereas subsequent action observation was seen from a third-person perspective, as when observing someone else's behavior. Therefore, an individual's visuomotor experience transfers, in a way that depends on the learning conditions, to the observation of another person's actions. This finding of transfer is remarkable as all viewed action sequences were visually novel, because none of them had previously been seen from a third-person view. Further, the participants passively viewed the actions while in the scanner, and so there was no explicit requirement for motor engagement during the test phase. Interestingly, our results show increased activation in frontal visuomotor areas during third-person action observation only when the actions had previously been learned in the vision + action condition (compared to all other conditions), not when comparing action-only to vision-only learning. Caggiano et al. (2011) have recently shown that there are view-dependent as well as view-independent mirror neurons in the macaque premotor cortex. The authors argued that view-dependent neurons represent an intermediate step in the formation of view independence. Also, a recent human fMRI study (Héту et al., 2011) suggests that the activity in frontal and parietal visuomotor areas is equally strong for first- and third-person visual perspectives of hands executing several different types of actions (transitive, nontransitive, and meaningless hand movements). However, using multivoxel pattern analysis, Oosterhof, Tipper, and Downing (2012) found action-specific cross-modal visual-motor representations in the ventral premotor cortex only for first-person but not for third-person views of actions. Evidence for cross-modal encoding regardless of perspective was found in the parietal and occipitotemporal cortices. Thus, current evidence (including this study) supports viewpoint invariance in premotor areas for viewed actions, whereas the extension of this invariance to common coding with performed actions remains less clear, in part because of the diversity of approaches (e.g., univariate, multivariate) taken in previous studies.

We interpret our results in line with frameworks (Heyes, 2001; Keyser & Perrett, 2004) in which correlated and contingent visual and motor signals during learning combine to most effectively engage frontoparietal action representations. This is consistent with the finding of greatest visuomotor responses to actions learned with both cues. But how can we account for the similar effects found in a comparison of actions

learned by performance alone and those learned by vision alone? One possibility is that viewing was a more passive task, and hence, less attention and shallower levels of encoding were undertaken when viewing compared to producing actions. However, there are two points to consider: First, the task of viewing actions has some ecological validity in our study. That is, in most situations we observe other people without specific motor intent, and it is possible that less attention is engaged with viewing other people than when producing one's own new actions. Hence, differences in attention may reflect the normal learning situation for the development of action representations, rather than being merely artifacts of our design. Second, there is in fact evidence for the weaker role of vision even in learning situations where there is rich visual input and where attention is actively oriented to vision. For example, male ballet dancers produce greater activation of visuomotor systems when observing dance movements that they would produce compared to moves typically made by female dancers (Calvo-Merino et al., 2006). Note, however, that in many situations the male dancers would have substantial visual exposure to the female dance movements and they would have to carefully attend to them as they time their own actions to those of their female dance partners. Hence, female dance movements are highly relevant to male dancers' own actions, but they are not internally generated motor states. Rather, it is the males' internal generation of their own motor output that plays the dominant role in action learning, enabling passive viewing of third-person actions to activate motor states.

Therefore, instead, our proposal is that in situations where a motor behavior is produced but its results cannot be seen, a visual representation of the ongoing action is spontaneously generated, as proposed by forward predictive models of vision and action. Because of delays in sensory processing, it is necessary to predict future states of visual information after motor responses. Furthermore, learning of new motor tasks, as in this study, requires a comparison between future visual states as predicted by motor output and what is actually perceived. While the action sequences used in this study were novel, participants will have had prior experience of a wide range of observations of the hand postures that made up the action sequences (as noted above). Specific motor postures will activate visual representations they were previously associated with and therefore as the hand moves to its next posture visual representations of this future state can be activated. So during motor responses, visual images are generated and mismatches between them and actual

visual input are used to fine-tune the visuomotor system in the development of skills¹ (e.g., Kawato, 1999).

Similarly, in accounts such as the ideomotor principle (Greenwald, 1970) or common event coding theory (Hommel, Musseler, Aschersleben, & Prinz, 2001), actions are automatically encoded in terms of the perceivable effects they generate (e.g., hand movements in this study). In sharp contrast, during passive visual perception of actions, no forward model is required. What is striking about our current results is that we typically assume the visual images generated when producing actions will reflect the egocentric first-person perspective of produced actions. However, our findings suggest that the generated images also generalize to subsequent views of allocentric third-person actions. Although we do not know the exact mechanisms by which the brain achieves this, it seems likely that having experience of mirror exposure, being imitated, and making synchronous movements with other people all contribute to the brain's ability to generalize across perspectives.

Original manuscript received 24 October 2011

Revised manuscript accepted 30 March 2012

First published online 16 May 2012

REFERENCES

- Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*, 1109–1116.
- Bach, P., Peatfield, N. A., & Tipper, S. P. (2007). Focusing on body sites: The role of spatial attention in action perception. *Experimental Brain Research*, *178*, 509–517.
- Bach, P., & Tipper, S. P. (2007). Implicit action encoding influences personal-trait judgments. *Cognition*, *102*, 151–178.
- Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: Comparing symbolic, spatial and imitative cues. *Brain and Cognition*, *44*, 124–143.
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R. J., Zilles, K., Rizzolatti, G., & Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, *13*, 400–404.
- Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., Porro, C. A., & Rizzolatti, G. (2004a). Neural circuits involved in the recognition of actions performed by nonconspecifics: An fMRI study. *Journal of Cognitive Neuroscience*, *16*, 114–126.
- Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H.-J., & Rizzolatti, G. (2004b). Neural circuits underlying imitation learning of hand actions: An event-related fMRI study. *Neuron*, *42*, 323–334.
- Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J. K., Thier, P., Giese, M. A., & Casile, A. (2011). View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. *Current Biology*, *21*, 144–148.
- Calvo-Merino, B., Ehrenberg, S., Leung, D., & Haggard, P. (2010). Experts see it all: Configural effects in action observation. *Psychological Research*, *74*, 400–406.
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, *15*, 1243–1249.
- Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, *16*, 1905–1910.
- Casile, A., & Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Current Biology*, *16*, 69–74.
- Catmur, C., Gillmeister, H., Bird, G., Liepelt, R., Brass, M., & Heyes, C. (2008). Through the looking glass: Counter-mirror activation following incompatible sensorimotor learning. *European Journal of Neuroscience*, *28*, 1208–1215.
- Catmur, C., Walsh, V., & Heyes, C. M. (2007). Sensorimotor learning configures the human mirror system. *Current Biology*, *17*, 1527–1531.
- Chong, T. T., Cunnington, R., Williams, M. A., Kanwisher, N., & Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Current Biology*, *18*, 1576–1580.
- Cook, R., Press, C., Dickinson, A., & Heyes, C. (2010). Acquisition of automatic imitation is sensitive to sensorimotor contingency. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 840–852.
- Cross, E. S., Hamilton, A. F. de C., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *NeuroImage*, *31*, 1257–1267.
- Cross, E. S., Hamilton, A. F. de C., Kraemer, D. J. M., Kelley, W. M., & Grafton, S. T. (2009). Dissociable substrates for body motion and physical experience in the human action observation network. *European Journal of Neuroscience*, *30*, 1383–1392.
- Cross, E. S., Kraemer, D. J. M., Hamilton, A. F. de C., Kelley, W. M., & Grafton, S. T. (2009). Sensitivity of the action observation network to physical and observational learning. *Cerebral Cortex*, *19*, 315–326.
- Dinstein, I., Gardner, J. L., Jazayeri, M., & Heeger, D. J. (2008). Executed and observed movements have different distributed representations in human aIPS. *Journal of Neuroscience*, *28*, 11231–11239.
- Dinstein, I., Hasson, U., Rubin, N., & Heeger, D. J. (2007). Brain areas selective for both observed and

¹Note that this is likely to apply to most but not all types of motor responses. The successful learning of whole-body actions, for example, where visual input does not provide useful information about the location of unseen body parts is likely to rely more on other processes such as “external imagery” (imagining seeing the self from a third-person perspective) (e.g., Hardy & Callow, 1999).

- executed movements. *Journal of Neurophysiology*, 98, 1415–1427.
- Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, 91, 176–189.
- Etzel, J. A., Gazzola, V., & Keysers, C. (2008). Testing simulation theory with cross-modal multivariate classification of fMRI data. *PLoS One*, 3, e3690.
- Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, 73, 2608–2611.
- Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: Single-subject analyses of unsmoothed fMRI data. *Cerebral Cortex*, 19, 1239–1255.
- Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideomotor mechanism. *Psychological Review*, 77, 73–99.
- Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, 12, 1–19.
- Hamilton, A. F., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience*, 26, 1133–1137.
- Hamilton, A. F., & Grafton, S. T. (2008). Action outcomes are represented in human inferior frontoparietal cortex. *Cerebral Cortex*, 18, 1160–1168.
- Hardy, L., & Callow, N. (1999). Efficacy of external and internal visual imagery perspectives for the enhancement of performance on tasks in which form is important. *Journal of Sport & Exercise Psychology*, 21, 95–112.
- Héту, S., Mercier, C., Eugène, F., Michon, P. E., & Jackson, P. L. (2011). Modulation of brain activity during action observation: Influence of perspective, transitivity and meaningfulness. *PLoS One*, 6, e24728.
- Heyes, C. M. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, 5, 253–261.
- Heyes, C. M. (2005). Imitation by association. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From mirror neurons to memes* (pp. 157–176). Cambridge, MA: MIT Press.
- Heyes, C., Bird, G., Johnson, H., & Haggard, P. (2005). Experience modulates automatic imitation. *Cognitive Brain Research*, 22, 233–240.
- Heyes, C., & Ray, E. (2000). What is the significance of imitation in animals? *Advances in the Study of Behavior*, 29, 215–245.
- Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *Journal of Cognitive Neuroscience*, 21, 1229–1243.
- Hommel, B., Musseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24, 849–878.
- Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9, 718–727.
- Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: A Hebbian perspective. *Trends in Cognitive Sciences*, 8, 501–550.
- Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., & Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *Journal of Neuroscience*, 29, 10153–10159.
- Lingnau, A., Gesierich, B., & Caramazza, A. (2009). Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 9925–9930.
- Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology*, 20, 750–756.
- Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2012). Viewpoint (in)dependence of action representations: An MVPA study. *Journal of Cognitive Neuroscience*, 24, 975–989.
- Oosterhof, N. N., Wiggett, A. J., Diedrichsen, J., Tipper, S. P., & Downing, P. E. (2010). Surface-based information mapping reveals crossmodal vision–action representations in human parietal and occipitotemporal cortex. *Journal of Neurophysiology*, 104, 1077–1089.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–192.
- Romani, M., Cesari, P., Urgesi, C., Facchini, S., & Aglioti, S. M. (2005). Motor facilitation of the human cortico-spinal system during observation of biomechanically impossible movements. *NeuroImage*, 26, 755–763.
- Serino, A., De Filippo, L., Casavecchia, C., Coccia, M., Shiffrar, M., & Làdavas, E. (2010). Lesions to the motor system affect action perception. *Journal of Cognitive Neuroscience*, 22, 413–426.
- Strafella, A. P., & Paus, T. (2000). Modulation of cortical excitability during action observation: A transcranial magnetic stimulation study. *NeuroReport*, 11, 2289–2292.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. Stuttgart: Thieme.

APPENDIX: ACTION SEQUENCES

