What is simulated in the action observation network when we observe actions?

Alice Neal and **James M. Kilner**
The Wellcome Trust Centre for Neuroimaging, IoN, UCL, 12 Queen Square, London WC 1N 3BG, UK

Abstract

In the last decade there has been a great amount of research investigating the role of simulation in our ability to infer the underlying intentions of any observed action. The majority of studies have focussed on the role of mirror neurons and the network of cortical areas active during action observation (AON) in inferring the goal of an observed action. However, it remains unclear what precisely is simulated when we observe an action and how such simulations can enable the observer to infer the underlying intention of that action. In particular it is not known how simulation in the AON enables the inference of the same goal when the kinematics observed to achieve that goal differ, such as when reaching to grasp an object with the left or right hands. Here we performed a behavioural study with healthy human subjects to address this question. We show that the subjects were able to detect very subtle changes in the kinematics of an observed action. In addition, we fitted the behavioural responses with a model based on the predictive coding account of mirror neurons. This is a Bayesian account of action observation that can be explained by the free-energy principle. Here we show that we can model all the effects observed when the action observation system is considered within a predictive coding framework.

Keywords

action observation; Bayesian; human; mirror neurons; motor

Introduction

When we observe other people executing an action we are able to estimate their underlying intentions and goals. Many believe that this ability is made possible by activity of a particular class of neurons, mirror neurons, in the observer's own motor system during action observation (Di Pellegrino *et al.*, 1992; Gallese *et al.*, 1996; Rizzolatti *et al.*, 2001; Rizzolatti & Craighero, 2004; Fogassi *et al.*, 2005). However, despite nearly two decades of research on mirror neurons and the action observation network (AON) it remains unclear how the visual information from an observed action maps onto the observer's own motor system and how the goal of that action is inferred (Gallese *et al.*, 2004; Iacoboni, 2005; Jacob & Jeannerod, 2005; Saxe, 2005). Implicit in many descriptions of the AON is the idea that visual information is transformed as it is passed by forward connections along the AON from low-level representations of the movement kinematics to high-level representations of intentions subtending the action. Previously we have shown that such a scheme can not work in the case of action observation as the same sensory input can have many causes, and therefore the process of generating the sensory inputs from the causes are non-invertible

(Kilner *et al.*, 2007a,b). Instead, we have proposed that the AON is best considered within a predictive coding framework. In the predictive coding framework, each level of a hierarchy employs a generative model to predict representations in the level below. This generative model uses backward connections to convey the prediction to the lower level where it is compared with the representation in this subordinate level to produce a prediction error. The notion in predictive coding that we employ our own motor system to predict the observed action is aligned with the idea that mirror neuron activity reflects a simulation of the observed action (Jeannerod, 1994; Prinz, 1997; Gallese & Goldman, 1998; Wohlschläger *et al.*, 2003; Rizzolatti & Craighero, 2004; Grafton, 2009).

Observed actions can be described, and therefore must be simulated, at many levels (Kilner *et al.*, 2007a,b; Hamilton & Grafton, 2009). The vast majority of research on action observation has focussed on the intention and goal level. However, any system that is capable of inferring the goal and intention of an observed action must do so with only access to the observed kinematics. This means that the predictive coding and simulation account of mirror neuron activity must resolve how one unique goal of an action can be inferred when the observed kinematics are different. For example, when we reach and grasp an object with either our right or left hands the goal is the same, however, the observed kinematics differ (Tretriluxana *et al.*, 2008). In such cases do we generate or simulate the observed kinematics in an identical way to which we execute it, with a different generative model for the left and right hands as simulation theory would imply, or do we employ the same generative model for both.

Here we designed an experiment to disambiguate between these two accounts. Subjects observed videos of left- or right-handed actions. Half of all the videos were manipulated by reflecting the video around the vertical midpoint such that an original right-handed action would now appear like a left-handed action and *vice versa*. Subjects had to decide whether each video was manipulated or not. We show that subjects' responses are inconsistent with how the simulation account is commonly interpreted, but demonstrate that the results can be modelled by a predictive coding account of perceptual inference (Kilner *et al.*, 2007a,b).

Materials and methods

Subjects

Seventeen subjects (eight female) took part in the study (aged 26 ± 5.4 years). Informed consent was collected from each subject prior to the study and the study had local ethical approval.

Experimental design

Prior to the experiment subjects were told that they would be presented with short video clips of reach and grasp movements. They were informed that half of these movements had been altered using video software. Subjects were given no information about which aspect of the movement had been altered. The videos consisted of four actors making reach and grasp movements to a small object (Fig. 1). To make the videos each actor sat on a wooden stool facing a table covered by a black cloth. Actors were given a small cylindrical object and asked to place it at a comfortable arms reach. Starting from a pinch grip on a pin placed in the table directly in front of the actors, actors reached for the object, grasped it using a pinch grip and brought it back to the start position. Each movement was repeated eight times: four times the action was recorded from an allocentric or third-person viewpoint; and four times the action was recorded from an egocentric or first-person viewpoint. For the egocentric conditions, the camera was mounted on a tripod and placed just above the actors'

right shoulder when the subject made right-handed movements, and above the left shoulder for left-handed movements.

All videos were manipulated in the same way, they were reflected about the vertical midpoint using Windows Movie Maker, and edited with AVID LIQUID PRO v.7 to remove shadows and block potential flipping cues. In this way, for example, a left-handed action would now appear as a right hand, but would reach and grasp with identical kinematics as the original video.

There were eight video conditions that formed a $2 \times 2 \times 2$ factorial design (Fig. 1). The factors were: (i) observed hand (right or left); (ii) observed kinematics (left or right); and (iii) viewpoint (egocentric or allocentric). Subjects performed two blocks of the experiment. In one block subjects observed actions recorded from an egocentric viewpoint, and in the other they viewed the actions recorded from an allocentric viewpoint. The order of the viewpoint factor was random across subjects. In each block subjects observed 64 different videos. These were the four movements of the right and left hands for each of the four actors, presented in both the unmanipulated form and also after they had been manipulated. These videos were presented in a pseudo random sequence within each block.

Subjects were presented with a fixation cross followed by a video of a movement about 1 s in duration. Immediately after the video came the response screen, at the top of which was the question 'Manipulated?'. Subjects were required to press the 'n' key for 'no not manipulated' or the 'b' key for 'yes manipulated' before moving on to the next trial. Subjects were given a break between blocks.

Following the video blocks subjects were asked to complete a computer-based version of the Edinburgh Handedness Inventory. Handedness was measured using the Edinburgh Handedness Inventory and analysed using MATLAB v.6. The response screen and the handedness inventory were created using GSCND v1.254. Stimuli were presented and responses recorded on a PC using COGENT2000 version1.25.

Analysis

Prior to analysis the responses of each subject were transformed to percentage of 'no' answers – the percentage of times the subjects responded that they thought that the video had not been manipulated. At this stage one subject was eliminated from further analyses as he / she constantly responded that the videos appeared unmanipulated or natural. It was unclear whether this subject had not understood the task and so was eliminated prior to any further analyses. All further analyses were based on the responses of the remaining 16 subjects.

Model

The predictive coding account of goal inference of observed actions can be shown to be formally equivalent to empirical Bayesian inference (Kilner *et al.*, 2007a,b). The empirical Bayesian perspective on perceptual inference suggests that the role of backward connections is to provide contextual guidance to lower levels through a prediction of the lower level's inputs. Given this conceptual model, a stimulus-related response can be decomposed into two components. The first encodes the conditional expectation of perceptual causes, μ. The second encodes prediction error, ε. Here we were interested in whether the observed pattern of behavioural responses could be explained by modulations in the prior mean (μ) and the prior precision (Π). To this end we assumed that the percentage of times the subjects responded that the observed videos was natural was negatively correlated with the prediction error, ε, such that

$$\varepsilon = \Delta \Pi$$

where Δ is the difference between the observed data (y) and the prior mean (μ), and Π is the prior precision, the reciprocal of the prior variance (σ, see Fig. 4 for schematic). Note that here the prediction error (ε) is dependent upon both the prior mean (μ) and the prior precision (Π).

Results

In general subjects had a significant bias in reporting that they thought that the videos were natural or not manipulated (Fig. 2; $P < 0.05$). This is unsurprising as all the kinematics were natural. Despite the fact subjects reported that they were guessing, the within-subject analysis of the behavioural results indicated the interaction between observed hand and observed kinematics was significant ($F_{1,15} = 5.267$, $P < 0.05$; Fig. 2). All other main effects and interactions did not meet the required significance criterion, $P < 0.05$ (main effect of viewpoint $F_{1,15} = 1.660$, $P = 0.217$; main effect of observed kinematics $F_{1,15} = 0.141$, $P = 0.141$; interaction between viewpoint and observed kinematics $F_{1,15} = 0.142$, $P = 0.141$; main effect of observed hand $F_{1,15} = 4.260$, $P = 0.057$; interaction between viewpoint and observed hand $F_{1,15} = 0.009$, $P = 0.927$; interaction between viewpoint, observed kinematics and observed hand $F_{1,15} = 0.152$, $P = 0.482$). *Post hoc t*-tests revealed that the interaction was driven by an increase in the percentage of times that the subject responded that the right hand observed with right-hand kinematics was natural, or not manipulated, compared with the other conditions. This was particularly evident when the subjects observed the actions from a third-person perspective (Fig. 2). The interaction was not a crossover interaction. In other words, subjects did not think that actions with left-hand kinematics performed by a left hand were natural, even though these videos were not manipulated, and subjects can quite as easily reach and grasp an object with their left hands as with their right hands.

Subsequent analyses focussed on between-subject variance. For each of the eight conditions we tested whether the subjects' perception of the movements observed were correlated with the degree to which they were right-handed. To this end we performed a linear regression analysis between each individual's index of laterality and the percentage of times they rated an observed action as natural or unmanipulated. Only one of the eight conditions was significantly correlated with the index of laterality, the right-hand kinematics when viewed as a right hand from the first-person perspective, which was negatively correlated with the laterality index ($P < 0.05$, $R^2 = 0.31$; Fig. 3A and C). Of the seven non-significant correlations one showed a trend to significance ($P = 0.07$, $R^2 = 0.21$), this was the right-hand kinematics when viewed as a left hand from a first-person perspective, which was negatively correlated with the laterality index. In other words for both of these correlations the more right-handed a person was the less natural they thought the actions were. None of the correlations between laterality and ratings to actions shown in a third-person perspective was close to significance (Fig. 3).

Discussion

Here we have shown that: (i) when we observe actions we are sensitive to subtle and non-perceived differences in the kinematics of the observed action (Fig. 2); and (ii) that the degree to which subjects thought that action with right-handed kinematics was natural was inversely correlated with how right-handed the subject was (Fig. 3). Both of these results are difficult to explain by a simple 'simulation theory' interpretation of the role of the AON and mirror neurons in action understanding. If we simulate the observed action in exactly the

same way that we would execute it then why did subjects not perceive left-handed actions performed by the left as natural? And why would the perception of the action be inversely correlated with how right-handed the subject was? The pattern of responses observed here, however, can not be explained by this simulation account as there were no significant differences when the left-hand kinematics were viewed either in their natural or manipulated forms. Such discrepancies between the simulation theory and the results here can be resolved by considering simulation theory within the recently proposed framework of predictive coding (Kilner *et al.*, 2007a,b). The essence of the predictive coding account is that, given a prior expectation about the goal / intention of the person we are observing, we use our own motor system to generate a model of how we perform the same action to predict the observed action kinematics of others. The comparison of the predicted kinematics with the observed kinematics generates prediction errors. These prediction errors are used to update our representation of the person's motor commands and inferred goals. By minimizing the prediction error at all the levels of a cortical hierarchy engaged when observing actions, the most likely cause of the action will be inferred at all levels (intention, goal, motor and kinematic). The larger the prediction error between the predicted and observed kinematics the more likely the subjects would be to rate the action as unnatural or manipulated.

A partial explanation for the results described here can be achieved by assuming that we always simulate the observed action as if we were executing the action with our dominant hand. In other words, right-handers would simulate any observed action as if it were being performed by the right hand. In this case, when subjects observed an action executed by the left hand with left-handed kinematics there would be a prediction error between the model generated with right-hand kinematics and the observed left-handed kinematics, and the subjects would consequently have a higher probability of rating these videos as manipulated. Indeed, it could be argued that as left-handed actions are more difficult to simulate than right-handed actions, because they are less frequently produced, one might expect this result. However, this can only be a partial explanation of the results as this model would predict that both the right-handed manipulated and non-manipulated videos should appear natural as they both have the kinematics of a right-handed action. This was not the case. Furthermore, this account offers no explanation of the negative correlation between the subjects' laterality index and their responses. Although it might seem logical to argue that right-handed actions are more easily modelled by right-handers because the subjects see these actions more frequently, the purpose of this paper was to go further and provide a model that could explain such results. It should be noted that the predictive coding account and an experience-based account are not mutually exclusive. Indeed, in the predictive coding account the parameters of the generative models must be learned through perceptual learning (Friston, 2003; Kilner & Blakemore, 2007), and this process will clearly be experience dependent. The point here is that whereas some of the results are consistent with an experience-based account, not all of them are.

Within the predictive coding account the prediction error can be weighted so that it is dependent not only on the parameters of the generative model but also the prior precision of these parameters (the confidence that these parameters are likely to be correct; Friston *et al.*, 2006; Friston, 2009; Friston & Kiebel, 2009). This is shown schematically in Fig. 4. Here we propose that when subjects observed the actions they always simulated the action using a generative model of how they would perform the same action with their dominant hand, and in addition they modulated their confidence that the generative model would produce the correct kinematics such that it was increased when observing actions that appeared as if they were right-handed and decreased for actions that appeared as if they were left-handed. In other words, in each situation subjects used the same prior means but modulated the prior precision depending upon whether they were observing a right hand or a left hand. By

always modelling the observed kinematics with a generative model of our dominant hand with the precision on the parameters of this model that are dependent upon the hand observed, it is possible to simulate the results observed here and presented in Fig. 2 (Fig. 4).

Modulating the precision of parameters in our generative model when observing actions can also be used to explain the correlation between the individuals' index of laterality and their ratings. Such correlations were only significant for actions with right-handed kinematics when viewed from the first-person perspective. The majority of times we view a hand in the first-person perspective it is our own hand. When we observe our own self-generated movements the precision of the parameters of the generative model predicting the produced kinematics will be high as we have learnt how we move our own limbs. It seems reasonable to assume that the precision will not be the same for each subject and will be dependent upon how well the subject thinks they can perform the action. If we take the laterality index as a rough surrogate for the precision that each subject has on the parameters of their own generative model for executed actions, we can account for the negative correlation between the subjects' rating and their laterality index by assuming that when the subjects observe an action in the first-person viewpoint they then model the kinematics as if it were their own hand. Previous studies have shown that the mere observation of a hand observed from this viewpoint is sufficient for us to believe that that hand is our own (Ehrsson *et al.*, 2004, 2005, 2007; Makin *et al.*, 2008). In this case the confidence in the parameters of the generative model of any observed action would be correlated with how right-handed the subjects were, such that if a subject had a high laterality index then they would have a high precision that the parameters of the model were correct. However, the actions were not performed by them, so that subjects that had a high precision would have a larger precision-weighted prediction error and therefore they would have a higher incidence of reporting the observed action as non-natural (Fig. 4). When subjects observe actions from a third-party viewpoint they then a *priori* know that the kinematics observed would not be fitted with their own model and so have a broader pattern of prior precisions. In this case no significant negative correlations are observed. Although speculative, this suggests that subjects can use the prior precision to modulate self from other generated actions (Gallagher, 2007).

Summary

Here we have shown that: (i) when we observe actions we are sensitive to subtle and non-perceived differences in the kinematics of the observed action; and (ii) that the degree to which subjects thought that right-handed videos were natural was inversely correlated with how right-handed the subject was. We argue that these results can only be explained within the predictive coding framework of simulation theory of observed actions. This account suggests that we always generate a simulation of an observed action as if we would execute the action with our dominant hand, and we modulate the confidence or prior precision of this model. Such a model creates the testable hypothesis that some neuronal populations that are active during action observation should be functionally correlated not with the prior mean of the simulated action but with the prior precision.

Acknowledgments

Abbreviation

AON   action observation network.

References

Di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G. Understanding motor events: a neurophysiological study. Exp. Brain Res. 1992; 91:176–180. [PubMed: 1301372]

Ehrsson HH, Spence C, Passingham RE. That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. Science. 2004; 305:875–877. [PubMed: 15232072]

Ehrsson HH, Holmes NP, Passingham RE. Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. J. Neurosci. 2005; 25:10564–10573. [PubMed: 16280594]

Ehrsson HH, Wiech K, Weiskopf N, Dolan RJ, Passingham RE. Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. Proc. Natl. Acad. Sci. USA. 2007; 104:9828–9833. [PubMed: 17517605]

Fogassi L, Ferrari PF, Gesierich B, Rozzi S, Chersi F, Rizzolatti G. Parietal lobe: from action organization to intention understanding. Science. 2005; 308:662–667. [PubMed: 15860620]

Friston KJ. Learning and inference in the brain. Neural Netw. 2003; 16:1325–1352. [PubMed: 14622888]

Friston K. The free-energy principle: a rough guide to the brain? Trends Cogn. Sci. 2009; 13:293–301. [PubMed: 19559644]

Friston K, Kiebel S. Predictive coding under the free-energy principle. Philos. Trans. R. Soc. Lond. B Biol. Sci. 2009; 364:1211–1221. [PubMed: 19528002]

Friston K, Kilner J, Harrison L. A free energy principle for the brain. J. Physiol. Paris. 2006; 100:70–87. [PubMed: 17097864]

Gallagher S. Simulation trouble. Soc. Neurosci. 2007; 2:353–365. [PubMed: 18633823]

Gallese V, Keysers C, Rizzolatti G. A unifying view of the basis of social cognition. Trends Cogn. Sci. 2004; 8:396–403. [PubMed: 15350240]

Gallese V, Goldman A. Mirror-neurons and the simulation theory of mind reading. Trends Cogn. Sci. 1998; 2:493–501. [PubMed: 21227300]

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. Brain. 1996; 119:593–609. [PubMed: 8800951]

Grafton ST. Embodied cognition and the simulation of action to understand others. Ann. N. Y. Acad. Sci. 2009; 1156:97–117. [PubMed: 19338505]

Hamilton, AF.; Grafton, ST. The motor hierarchy: from kinematics to goals and intentions. In: Rosetti, Y.; Kawato, M.; Haggard, P., editors. Attention and Performance xxii. Oxford University Press; Oxford, UK: 2009.

Iacoboni M. Neural mechanisms of imitation. Curr. Opin. Neurobiol. 2005; 15:632–637. [PubMed: 16271461]

Jacob P, Jeannerod M. The motor theory of social cognition: a critique. Trends Cogn. Sci. 2005; 9:21–25. [PubMed: 15639437]

Jeannerod M. The representing brain - neural correlates of motor intention and imagery. Behav. Brain Sci. 1994; 17:187–202.

Kilner JM, Blakemore SJ. How does the mirror neuron system change during development? Dev. Sci. 2007; 10:524–526. [PubMed: 17683337]

Kilner JM, Friston KJ, Frith CD. Predictive coding: an account of the mirror neuron system. Cogn. Process. 2007a; 8:159–166. [PubMed: 17429704]

Kilner JM, Friston KJ, Frith CD. The mirror-neuron system: a Bayesian perspective. Neuroreport. 2007b; 18:619–623. [PubMed: 17413668]

Makin TR, Holmes NP, Ehrsson HH. On the other hand: dummy hands and peripersonal space. Behav. Brain Res. 2008; 191:1–10. [PubMed: 18423906]

Prinz W. Perception and action planning. Eur. J. Cogn. Psychol. 1997; 9:129–154.

Rizzolatti G, Craighero L. The mirror-neuron system. Annu. Rev. Neurosci. 2004; 27:169–192. [PubMed: 15217330]

Rizzolatti G, Fogassi L, Gallese V. Neurophysiological mechanisms underlying the understanding and imitation of action. Nat. Rev. Neurosci. 2001; 2:661–670. [PubMed: 11533734]

Saxe R. Against simulation: the argument from error. Trends Cogn. Sci. 2005; 9:174–179. [PubMed: 15808499]

Tretriluxana J, Gordon J, Winstein CJ. Manual asymmetries in grasp pre-shaping and transport-grasp coordination. Exp. Brain Res. 2008; 188:305–315. [PubMed: 18437369]

Wohlschläger A, Gattis M, Bekkering H. Action generation and action perception in imitation: an instance of the ideomotor principle. Philos. Trans. R. Soc. Lond. B Biol. Sci. 2003; 358:501–515. [PubMed: 12689376]
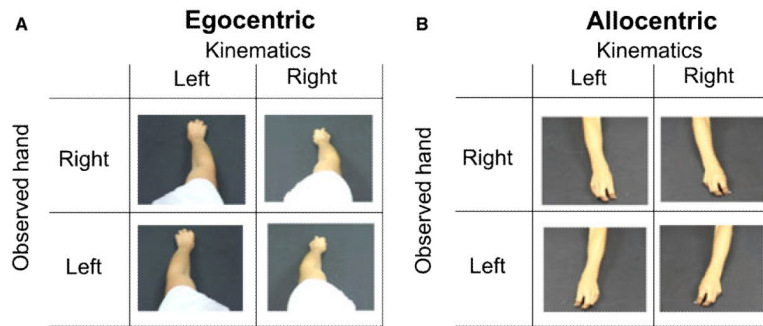
Fig. 1.
Experimental design. There were eight video conditions that formed a $2 \times 2 \times 2$ factorial design. The factors were observed hand (right or left), observed kinematics (right or left) and viewpoint [egocentric (A) and allocentric (B)].
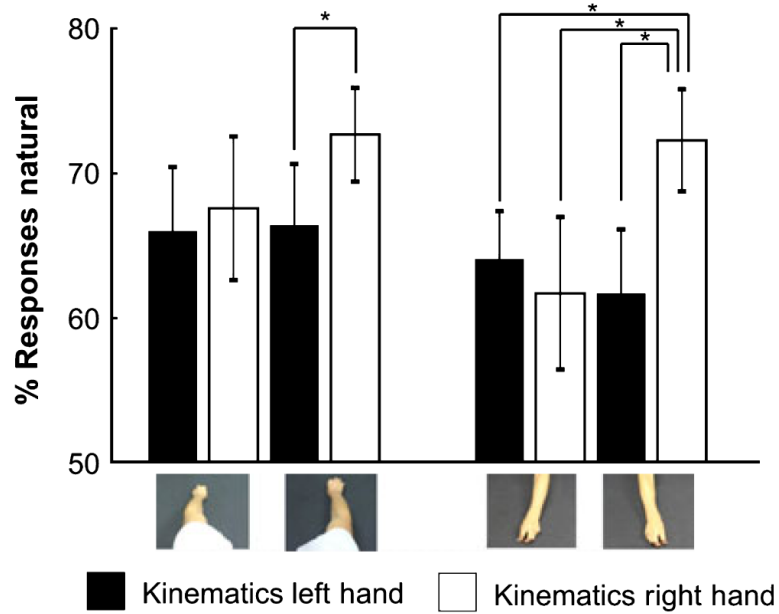
Fig. 2.
Behavioural responses. This figure shows the mean behavioural responses for each of the eight conditions. The error bars show standard error. The image of the hand depicts the observed hand, left or right. The white and the black bars show the responses to the observed kinematics for the right and left hands, respectively. Significant differences at $P < 0.05$ are shown (indicated by the *). All other comparisons were non-significant at this threshold.
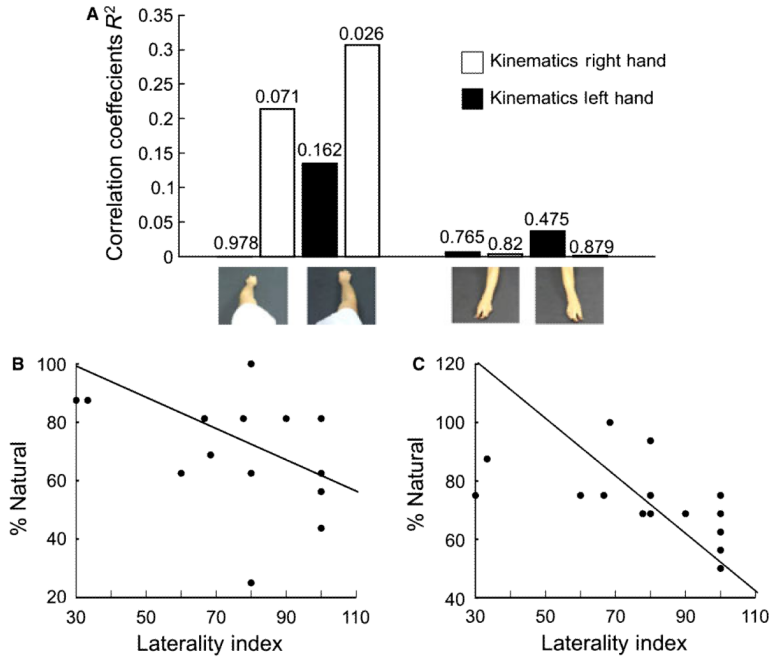
Fig. 3.
Correlations with laterality index. (A) The correlation coefficient R^2 for the correlation across subjects of the laterality index for each subject and their ratings that the videos were not manipulated. The image of the hand depicts the observed hand, left or right. The white and the black bars show the responses to the observed kinematics for the right and left hands, respectively. The significance of each correlation is shown above the appropriate bar. (B and C) Scatter plots of percentage of responses that were natural or non-manipulated against the laterality index. (B) The scatter for responses to videos of actions from an egocentric viewpoint that appeared to show a left hand, but right-hand kinematics. (C) The scatter for responses to videos of actions from an egocentric viewpoint that appeared to show a right hand, and right-hand kinematics.
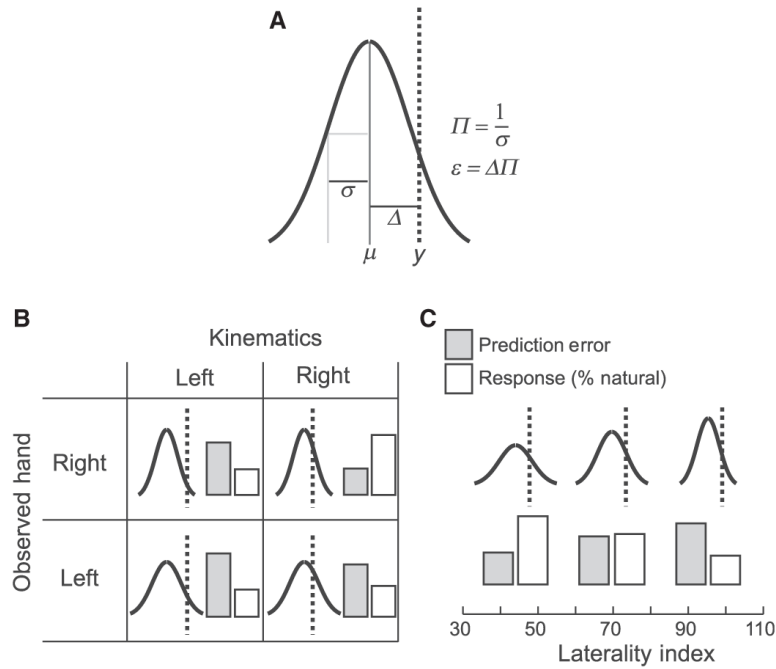
Fig. 4.
Schematic of the effect of prior mean and precision on prediction error. (A) A schematic outlining the relationship between the prior mean (μ), the prior precision (Π), the observed data (y) and the precision-weighted prediction error (ε). The prior precision is the reciprocal of the prior variance (σ). Note that here the prediction error (ε) is dependent upon the difference (Δ) between the observed data (y) and the prior mean (μ) and the prior precision (Π). (B) A schematic depicting the effect of prior mean and precision on the prediction error for each of the four conditions collapsed across the factor viewpoint. The Gaussian represents the precision of the prior for the observed action (see A). This is dependent upon the observed hand, being more precise for right-handed actions. Note that the prior mean does not change. The dashed line depicts the observed data. This is dependent upon the observed kinematics. Note that the prior mean is closer to the observed data for actions observed with right-handed kinematics than for those with left-handed kinematics. The grey bar depicts the size of the prediction error. The white bar depicts the corresponding percentage of natural responses for that condition. Note that when the prediction error is high the percentage of natural responses is low. (C) A schematic showing how changes in the prior precision can explain the negative correlation between subjects' responses and their laterality index. Note that the precision decreases with the subject's index of laterality, this resulted in a decrease in the prediction error and a subsequent increase in the percentage of 'natural' responses.