Active inference under intersensory conflict: Simulation and empirical results

Short title: Active inference under intersensory conflict

Jakub Limanowski and Karl Friston

Wellcome Centre for Human Neuroimaging, Institute of Neurology, University College London, London, UK

Correspondence

Jakub Limanowski

Wellcome Centre for Human Neuroimaging at UCL, 12 Queen Square, London WC1N 3BG

Email: j.limanowski@ucl.ac.uk

Phone: +44 (0)20 3448 4362

Abstract

It has been suggested that the brain controls hand movements via internal models that rely on visual and proprioceptive cues about the state of the hand. In active inference formulations of such models, the relative influence of each modality on action and perception is determined by how precise (reliable) it is expected to be. The 'top-down' affordance of expected precision to a particular sensory modality presumably corresponds to attention. Here, we asked whether increasing attention to (i.e., the precision of) vision or proprioception would enhance performance in a hand-target phase matching task, in which visual and proprioceptive cues about hand posture were incongruent. We show that in a simple simulated agent—using a neurobiologically informed predictive coding formulation of active inference—increasing either modality's expected precision improved task performance under visuo-proprioceptive conflict. Moreover, we show that this formulation captured the behaviour and self-reported attentional allocation of human participants performing the same task in a virtual reality environment. Together, our results show that selective attention can balance the impact of (conflicting) visual and proprioceptive cues on action—rendering attention a key mechanism for a flexible body representation for action.

Author summary

When controlling hand movements, the brain can rely on seen and felt hand position or posture information. It is thought that the brain combines these estimates into a multisensory hand representation in a probabilistic fashion, accounting for how reliable each estimate is in the given context. According to recent formal accounts of action, the expected reliability or 'precision' of sensory information can—to an extent—also be influenced by attention. Here, we tested whether this mechanism can improve goal-directed behaviour. We designed a task that required tracking a target's oscillatory phase with either the seen or the felt hand posture, which were decoupled by introducing a temporal conflict via a virtual reality environment. We first simulated the behaviour of an artificial agent performing this task, and then compared the simulation results to behaviour of human participants performing the same task. Together, our results showed that increasing attention to the seen or felt hand was accompanied by improved target tracking. This suggests that, depending on the current behavioural demands, attention can balance how strongly the multisensory hand representation is relying on visual or proprioceptive sensory information.

Keywords: action, active inference, attention, body representation, multisensory integration, precision, predictive coding

Introduction

Controlling the body's actions in a constantly changing environment is one of the most important tasks of the human brain. The brain solves the complex computational problems inherent in this task by using internal probabilistic (Bayes-optimal) models (Wolpert et al., 1998; Körding & Wolpert, 2004; Kilner et al., 2007; Shadmehr & Krakauer, 2008; Friston et al., 2010; Friston, 2011). These models allow the brain to flexibly estimate the state of the body and the consequences of movement, despite noise and conduction delays in the sensorimotor apparatus, via iterative updating by sensory prediction errors from multiple sources. The state of the hand, in particular, can be informed by vision and proprioception. Here, the brain makes use of an optimal integration of visual and proprioceptive signals, where the relative influence of each modality—on the final estimate—is determined by its relative reliability or precision, depending on the current context (van Beers et al., 1999, 2002; Foulkes & Miall, 2000; Ingram et al., 2000; Sober & Sabes, 2005; Friston et al., 2010; Friston, 2012; Samad et al., 2015; Rohe & Noppeney, 2016).

These processes can be investigated under an experimentally induced conflict between visual and proprioceptive information. The underlying rationale here is that incongruent visuo-proprioceptive cues about hand position or posture have to be integrated (provided the incongruence stays within reasonable limits), because the brain's body model entails a strong prior belief that information from both modalities is generated by one and the same external cause; namely, one's hand. Thus, a partial recalibration of one's unseen hand position towards the position of a (fake or mirror-displaced) hand seen in an incongruent position has been interpreted as suggesting an (attempted) resolution of visuo-proprioceptive conflict to maintain a prior body representation (Botvinick & Cohen, 1998; Pavani et al., 2000; Holmes et al., 2004, 2006; Tsakiris & Haggard, 2005; Makin et al., 2008; Heed et al., 2011; Limanowski & Blankenburg, 2016).

Importantly, spatial or temporal perturbations can be introduced to visual movement feedback *during action*—by displacing the seen hand position in space or time, using video recordings or virtual reality. Such experiments suggest that people are surprisingly good at adapting their movements to this kind of perturbations; i.e., they adjust to the novel visuo-motor mapping by means of visuo-proprioceptive recalibration or adaptation (e.g. Foulkes & Miall, 2000; Ingram et al., 2000; Balslev et al., 2004; Grafton et al., 2008; Bernier et al., 2009). Brain imaging studies have shown that during motor tasks involving the resolution of a visuo-proprioceptive conflict, one typically observes an increase in visual and multisensory activity (cf. Grefkes et al., 2004; Ogawa et al., 2007; Limanowski et al., 2017). Furthermore, the remapping required for this resolution is thought to be augmented by attenuation of proprioceptive activity (Balslev et al., 2004; Bernier et al., 2009; cf. Taub & Goldberg, 1974; Ingram et al., 2000). The conclusion generally drawn from these results is that visuo-proprioceptive recalibration (or visuo-motor adaptation) relies on temporarily adjusting the weighting of conflicting visual and proprioceptive information to enable adaptive action under specific prior beliefs about one's 'body model'.

The above findings—and their interpretation—can be accommodated within a hierarchical predictive coding formulation of active inference as a form of Bayes-optimal motor control, in which proprioceptive *as well as* visual prediction errors can update higher-level beliefs about the state of the body and thus influence action (Friston, 2010, 2012; Adams et al., 2013; Vasser et al., 2019). Hierarchical predictive coding rests on a probabilistic mapping from unobservable causes (hidden states) to observable consequences (sensory states), as described by a hierarchical generative model, where each level of the model encodes conditional expectations (*'beliefs'*) about states of the world that best explains states of affairs encoded at lower levels (i.e., sensory input). The causes of sensations are inferred via model inversion. In other words, the model's beliefs are updated to accommodate or 'explain away' ascending prediction error (a.k.a. Bayesian filtering or predictive coding, Rao & Ballard, 1999; Friston & Kiebel, 2009; Bastos et al., 2012). Active inference extends hierarchical predictive coding from the sensory to the motor domain in that the agent is now able to fulfil its model predictions via *action* (Perrinet et al., 2014). In brief, movement occurs because high-level multi- or amodal beliefs about state transitions predict proprioceptive and exteroceptive (visual) states that would ensue if e.g. a particular grasping movement was performed. Prediction error is then suppressed throughout the motor hierarchy (Kilner et al., 2007; cf. Grafton & Hamilton, 2007), ultimately by spinal reflex arcs that enact the predicted movement. This also implicitly minimises exteroceptive prediction error; e.g. the predicted visual consequences of the action (Adams et al., 2013, Friston, 2011, Shipp et al., 2013). Crucially, all ascending prediction errors are precision-weighted based on model predictions (where precision corresponds to the inverse variance), so that a prediction error that is expected to be more precise has a stronger impact on belief updating. The 'top-down' affordance of precision can be associated with *attention* (Feldman & Friston, 2010; Edwards et al., 2012; Brown et al., 2013). The ensuing attentional set should have a fundamental implication for behaviour, as action should also be more strongly informed by prediction errors 'selected' by attention. In other words, the impact of visual or proprioceptive prediction errors on multisensory beliefs driving action should be regulated via 'top-down' affordance of precision. One might therefore expect that the impact of visual or proprioceptive cues on action (and perception) should not only depend on factors like sensory noise, but may also be changed by directing the focus of selective attention to one or the other modality.

Here, we used a predictive coding scheme (cf. Friston et al., 2010; Perrinet et al., 2014) to test this assumption. We simulated behaviour, under active inference, in a simple manual action task (Fig. 1) that required hand-target phase matching with prototypical grasping movements—based on visual or proprioceptive cues under visuo-proprioceptive conflict. Crucially, we included a condition in which proprioception had to be adjusted to maintain visual task performance and a converse condition, in which proprioceptive task performance had to be maintained in the face of conflicting visual information. This enabled us to address the effects reported in the visuo-motor adaptation studies reviewed above and studies showing automatic biasing of one's own movement execution by incongruent action observation (Brass et al., 2001; Kilner et al., 2003). In our simulations, we asked whether changing the relative precision afforded to vision versus proprioception—which, presumably, corresponds to attention—would improve task performance (i.e., target matching with the respective instructed modality, vision or proprioception) in each case. We implemented this 'attentional' manipulation by adjusting the inferred precision of each modality, thus changing the degree with which the respective prediction errors drove model updating and action (see below). We then compared the results of our simulation with the actual behaviour and subjective ratings of attentional focus of healthy participants performing the same task in a virtual reality environment. We anticipated that participants, in order to comply with task instructions, would adopt an ‘attentional set’ (Posner et al., 1976; 1978; cf. Rohe & Noppeney, 2018) prioritizing the respective instructed target tracking modality over the task-irrelevant one. In other words, the instructed tracking or response modality should become “situationally dominant” by attentional allocation (Kelso et al., 1975; cf.Warren & Cleaver, 2001; Redding et al., 1985).

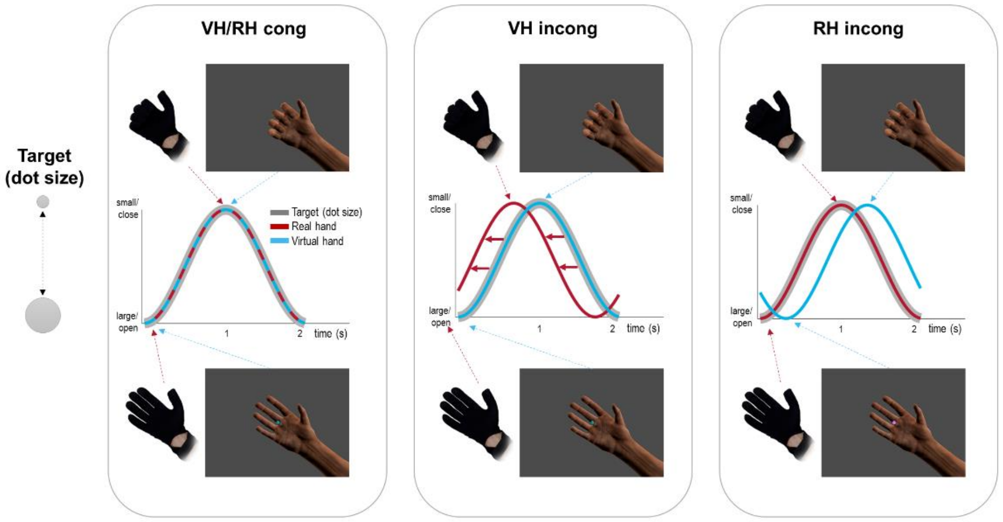

Figure 1. Task design and behavioural requirements. We used the same task design in the simulated and behavioural experiments, focusing on the effects of attentional modulation on hand-target phase matching via (near-)stationary, prototypical (i.e., well-trained) oscillatory grasping movements at 0.5 Hz. Participants (or the simulated agent) controlled a virtual hand model (seen on a computer screen) via a data glove worn on their unseen right hand. The virtual hand (VH) therefore represented seen hand posture (i.e., vision), which could be uncoupled from the real hand posture (RH; i.e., proprioception) by introducing a temporal delay (see below). The task required matching the phase of one’s right-hand grasping movements to the oscillatory phase of the fixation dot (‘target’), which was shrinking-and-growing sinusoidally at 0.5 Hz. In other words, participants had to rhythmically close the hand when the dot shrunk and to open it when the dot expanded. Our design was a balanced 2 x 2 factorial design: The task was completed (or simulated) under congruent or incongruent hand movements: the latter were implemented by adding a lag of 500 ms to the virtual hand movements (Factor ‘congruence’). Furthermore, the participants (or the simulated agent) performed the task with one of two goals in mind: to match the movements of the virtual hand (VH) or of those of the real hand (RH) to the phase of the dot (Factor ‘instructed modality’). Note that whereas in the congruent conditions (VH *cong*, RH *cong*) both hand positions were identical, and therefore both hands’ grasping movements could simultaneously be matched to the target’s oscillatory phase

(i.e., the fixation dot's size change), only one of the hands' (virtual or real) movements could be phase-matched to the target in the incongruent conditions—necessarily implying a phase mismatch of the other hand's movements. In the VH *incong* condition, participants had to adjust their movements to counteract the visual lag; i.e., they had to phase-match the virtual hand's movements (i.e., vision) to the target by shifting their real hand's movements (i.e., proprioception) out of phase with the target. Conversely, in the RH *incong* condition, participants had to match their real hand's movements (i.e., proprioception) to the target's oscillation, and therefore had to ignore the fact that the virtual hand (i.e., vision) was out of phase. The curves show the performance of an ideal participant (or simulated agent).

Results

Simulation results

The simulated agent had to match the phasic size change of a central fixation dot (target) with the grasping movements of the unseen real hand (proprioceptive hand information) or the seen virtual hand (visual hand information). Under visuo-proprioceptive conflict (i.e., a phase shift between virtual and real hand movements introduced via temporal delay), only one of the hands could be aligned with the target's oscillatory phase (see Fig. 1 for a detailed task description). The aim of our numerical analyses or simulations was to test whether—in the above manual phase matching task under perceived intersensory conflicts—increasing the expected precision of sensory prediction errors from the instructed modality (vision or proprioception) would result in improved task performance, whereas increasing the precision of prediction errors from the 'distractor' modality would impair it. Such a result would demonstrate that in an active inference scheme, behaviour under intersensory conflict can be augmented via top-down precision control; i.e., selective attention (cf. Feldman & Friston, 2010; Edwards et al., 2012; Brown et al., 2013).

Figures 2-3 show the results of these simulations, in which an active inference agent performed the target matching task under the two kinds of instruction (virtual hand or real hand task; i.e., the agent had a strong prior belief that the visual or proprioceptive hand posture would track the target's oscillatory size change) under congruent or incongruent visuo-proprioceptive mappings (i.e., where incongruence was realized by temporally delaying the virtual hand's movements with respect to the real hand). In this setup, the virtual hand corresponds to hidden states generating visual input, while the real hand generates proprioceptive input.

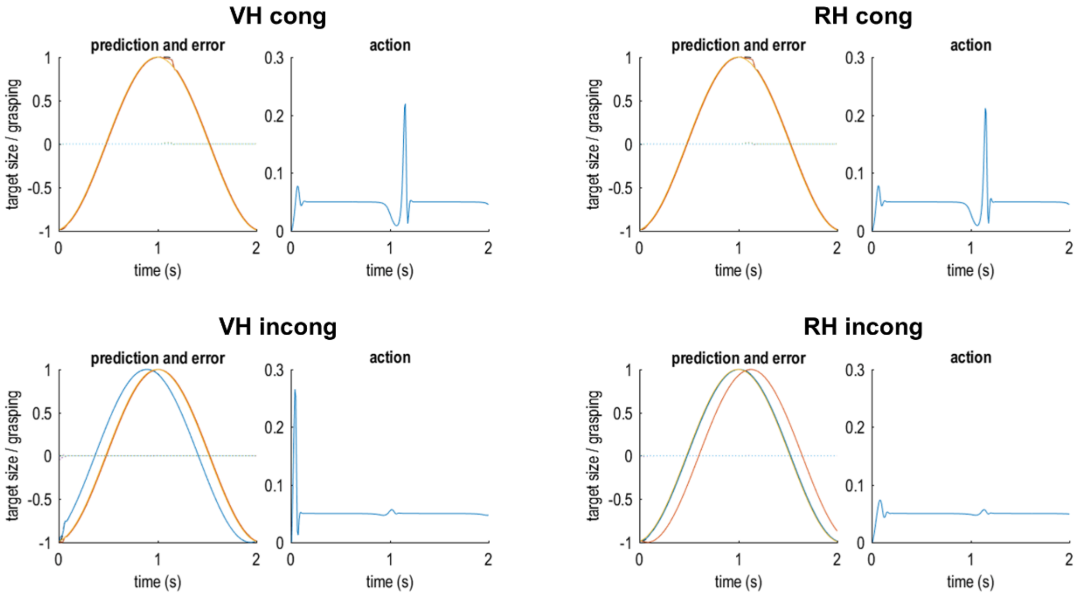

Figure 2. Simulated behaviour of an agent performing the hand-target phase matching task under ideally adjusted model beliefs. Each pair of plots shows the simulation results for an agent with a priori 'ideally' adjusted model beliefs about visuo-proprioceptive congruence; i.e., in the congruent tasks, the agent believed that its real hand generated matching seen and felt postures, whereas it believed that the same hand generated mismatching postures in the incongruent tasks. Each pair of plots shows the simulation results for one grasping movement in the VH and RH tasks under congruence or incongruence; the left plot shows the predicted sensory input (solid coloured lines; yellow = target, red = vision, blue = proprioception) and the true, real-world values (broken black lines) for the target and the visual and proprioceptive hand posture, alongside the respective sensory prediction errors (dotted coloured lines; blue = target, green = vision, purple = proprioception); the right plot (blue line) shows the agent's action (i.e., the rate of change in hand posture, see Methods). Note that target phase matching is near perfect and there is practically no sensory prediction error (i.e., the dotted lines stay around 0).

Under congruent mapping (i.e., in the absence of visuo-proprioceptive conflict) the simulated agent showed near perfect tracking performance (Fig. 2). We next simulated an agent performing the task under incongruent mapping, while equipped with the prior belief that its seen and felt hand postures were in fact unrelated, i.e., never matched. Not surprisingly, the agent easily followed the task instructions and again showed near perfect tracking with vision or proprioception, under incongruence (Fig. 2). However, as noted above, it is reasonable to assume that human participants would have the strong prior belief—based upon life-long learning and association—that their manual actions generated matching seen and felt postures (i.e., a prior belief that modality specific sensory consequences have a common cause). Our study design assumed that this association would be very hard to update, and that consequently performance could only be altered via adjusting expected precision of vision vs proprioception (see Methods).

Therefore, we next simulated the behaviour (during the incongruent tasks) of an agent embodying a prior belief that visual and proprioceptive cues about hand state were in fact congruent. As shown in Fig. 3A, this introduced notable inconsistencies between the agent's model predictions and the true states of vision and proprioception, resulting in elevated prediction error signals. The agent was still able to follow the task instructions, i.e., to keep the (instructed) virtual or real hand more closely matched to the target's oscillatory phase, but showed a drop in performance compared with the 'ideal' agent (cf. Fig. 2).

We then simulated the effect of our experimental manipulation, i.e., of increasing precision of sensory prediction errors from the respective task-relevant (constituting increased attention) or task-irrelevant (constituting increased distraction) modality on task performance. We expected this manipulation to affect behaviour; namely by how strongly the respective prediction errors would impact model belief updating and subsequent performance (i.e., action). The key result of these simulations (Fig. 3A) was that increasing the log precision of vision or proprioception—the respective instructed tracking modality—resulted in reduced visual or proprioceptive prediction errors. This can be explained by the fact that these 'attended' prediction errors were now more strongly suppressed by model belief updating—and action. Conversely, one can see a complementary increase of prediction errors from the 'unattended' modality.

Importantly, the above ‘attentional’ alterations substantially influenced hand-target phase matching performance (Fig. 3B). Thus, increasing the precision of the instructed task-relevant sensory modality’s prediction errors led to improved target tracking (i.e. a reduced phase shift of the instructed modality’s grasping movements from the target’s phase). In other words, if the agent attended to the instructed visual (or proprioceptive) cues more strongly, its movements were driven more strongly by vision (or proprioception)—which helped it to track the target’s oscillatory phase with the respective modality’s grasping movements. Correspondingly, increasing the precision of the ‘irrelevant’ (not instructed) modality in each case led to worse simulated tracking performance.

The simulations also show that the amount of action itself was comparable across conditions (blue plots in Figs. 2-3; i.e., movement of the hand around the mean stationary value of 0.05), which means that the kinematics of the hand movement per se were not biased by attention. Action was particularly evident in the initiation phase of the movement and after reversal of movement direction (open-to-close). At the point of reversal of movement direction, conversely, there was a moment of stagnation; i.e., changes in hand state were temporarily suspended (with action nearly returning to zero). In our simulated agent, this briefly increased uncertainty about hand state (i.e., which direction the hand was moving), resulting in a slight lag before the agent picked up its movement again, which one can see reflected by a small ‘bump’ in the true hand states (Figs. 2-3). These effects were somewhat more pronounced during movement under visuo-proprioceptive incongruence and prior belief in congruence—which indicates that the fluency of action depended on sensory uncertainty.

In sum, these results show that attentional effects of the sort we hoped to see can be recovered using a simple active inference scheme; in that precision control determined the influence of separate sensory modalities—each of which was generated by the same cause, i.e., the same hand—on behaviour by biasing action towards cues from that modality.

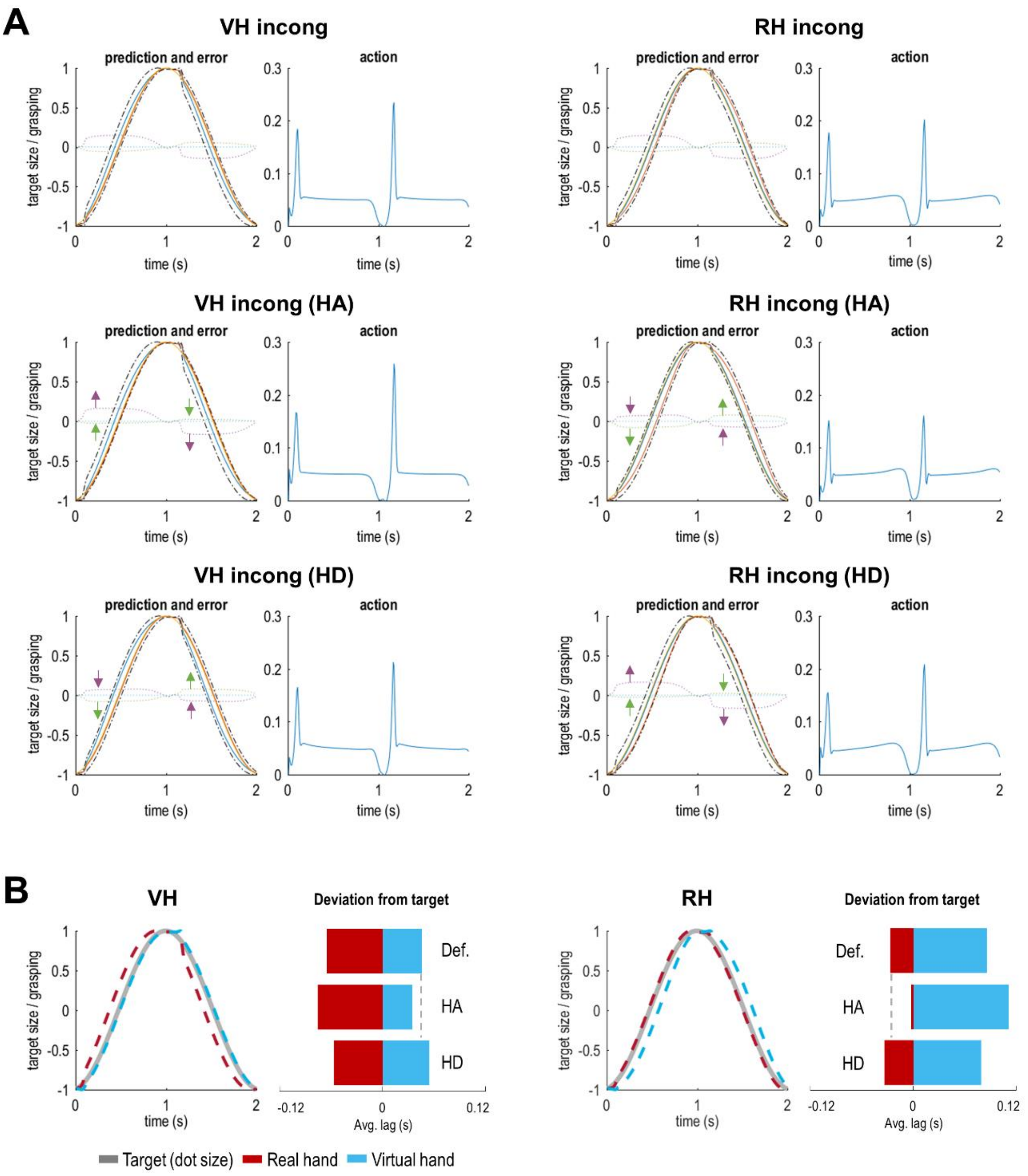

Figure 3. Simulated behaviour of a 'realistic' agent performing the hand-target phase matching task. Here we simulated an agent performing the incongruent tasks under the prior belief that its hand generated matching visual and proprioceptive information; i.e., under perceived intersensory conflict. **(A)** The plots follow the same format as in Fig. 2. Note that, in these results, one can see a clear divergence of true from predicted visual and proprioceptive postures, and correspondingly increased prediction errors. The top row shows the simulation results for the default weighting of visual and proprioceptive information; the middle row shows the same agent's behaviour when precision of the respective task-relevant modality (i.e., vision in the VH task and proprioception in the RH task) was increased (HA: high attention); the bottom row shows the analogous results when the precision

of the respective other, irrelevant modality was increased (HD: high distraction). Note how in each case, increasing (or decreasing) the log precision of vision or proprioception resulted in an attenuation (or enhancement) of the associated prediction errors (indicated by green and purple arrows for vision and proprioception, respectively). Crucially, these ‘attentional’ effects had an impact on task performance, as evident by an improved hand-target tracking with vision or proprioception, respectively. This is shown in panel **(B):** The curves show the tracking in the HA conditions. The bar plots represent the average deviation (phase shift or lag, in seconds) of the real hand’s (red) or the virtual hand’s (blue) grasping movements from the target’s oscillatory size change in each of the simulations shown in panel (A). Note that under incongruence (i.e., a constant delay of vision), reducing the phase shift of one modality always implied increasing the phase shift of the other modality (reflected by a shift of red *and* blue bars representing the average proprioceptive and visual phase shift, respectively). Crucially, in both RH and VH incong conditions, increasing attention (HA; i.e., in terms of predictive coding: the precision afforded to the respective prediction errors) to the task-relevant modality enhanced task performance (relative to the default setting, Def.), as evident by a reduced phase shift of the respective modality from the target phase. The converse effect was observed when the agent was ‘distracted’ (HD) by paying attention to the respective task-irrelevant modality.

Empirical results

We first analysed the post-experiment questionnaire ratings of our participants (Fig. 4) to the following two questions: “How difficult did you find the task to perform in the following conditions?” (Q1, answered on a 7-point visual analogue scale from “very easy” to “very difficult”) and “On which hand did you focus your attention while performing the task?” (Q2, answered on a 7-point visual analogue scale from “I focused on my real hand” to “I focused on the virtual hand”). For the ratings of Q1, a Friedman’s test revealed a significant difference between conditions ($\chi^2_{(3,69)} = 47.19$, $p < 0.001$). Post-hoc comparisons using Wilcoxon’s signed rank test showed that, as expected, participants reported finding both tasks more difficult under visuo-proprioceptive incongruence (VH *incong* > VH *cong*, $z_{(23)} = 4.14$, $p < 0.001$; RH *incong* > RH *cong*, $z_{(23)} = 3.13$, $p < 0.01$). There was no significant difference in reported difficulty between VH *cong* and RH *cong*, but the VH *incong* condition was perceived as significantly more difficult than the RH *incong* condition ($z_{(23)} = 2.52$, $p < 0.05$). These results suggest that, per default, the virtual hand and the real hand instructions were perceived as equally difficult to comply with, and that in both cases the added incongruence increased task difficulty—more strongly so when (artificially shifted) vision needed to be aligned with the target's phase.

For the ratings of Q2, a Friedman's test revealed a significant difference between conditions ($\chi^2_{(3,69)}$ = 35.83, $p < 0.001$). Post-hoc comparisons using Wilcoxon's signed rank test showed that, as expected, participants focussed more strongly on the virtual hand during the virtual hand task and more strongly on the real hand during the real hand task. This was the case for congruent (VH *cong* > RH *cong*, $z_{(23)}$ = 3.65, $p < 0.001$) and incongruent (VH *incong* > RH *incong*, $z_{(23)}$ = 4.03, $p < 0.001$) movement trials. There were no significant differences between VH *cong* vs VH *incong*, and RH *cong* vs RH *incong*, respectively. These results show that participants focused their attention on the instructed target modality, irrespective of whether the current movement block was congruent or incongruent. This supports our assumption that participants would adopt a specific attentional set to prioritize the instructed target modality.

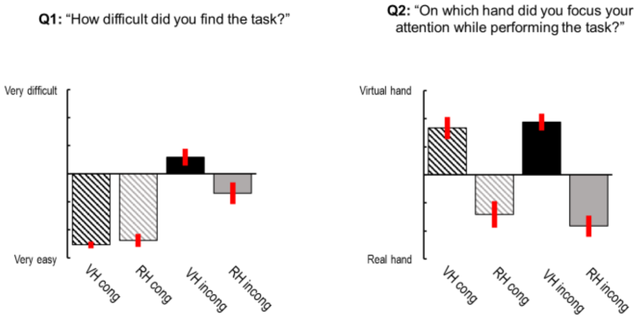

Figure 4. Self-reports of task difficulty and intersensory attentional focus given by our participants. The bar plots show the mean ratings for Q1 and Q2 (given on a 7-point visual analogue scale), with associated standard errors of the mean. On average, participants found the VH and RH task more difficult under visuo-proprioceptive incongruence—more strongly so when artificially shifted vision needed to be aligned with the target's phase (VH incong, Q1). Importantly, the average ratings of Q2 showed that participants attended to the instructed modality (irrespective of whether the movements of the virtual hand and the real hand were congruent or incongruent).

Next, we analysed the task performance of our participants; i.e., how well the virtual (or real) hand's grasping movements were phase-matched to the target's oscillation (i.e., the fixation dot's size change) in each condition. Note that under incongruence, better target phase-matching with the virtual hand implies a worse alignment of the real hand's phase with the target, and *vice versa*. As predicted (Fig. 1; and as confirmed by the simulation results, Figs. 2-3), we expected an interaction between task and congruence: participants should show a better target phase-matching of the virtual hand under visuo-proprioceptive incongruence, if the virtual hand was the instructed target modality (but no such difference should be significant in the congruent movement trials, since virtual and real hand movements were identical in these trials). All of our participants were well trained (see Methods), therefore our task focused on average performance benefits from attention (rather than learning or adaptation effects).

The participants' average tracking performance is shown in Figure 5. A repeated-measures ANOVA on virtual hand-target phase-matching revealed significant main effects of task ($F_{(1,22)} = 31.69$, $p = 0.00001$) and congruence ($F_{(1,22)} = 173.42$, $p = 3.38\text{e-}12$) and, more importantly, a significant interaction between task and congruence ($F_{(1,22)} = 50.69$, $p = 0.0000003$). Post-hoc *t*-tests confirmed that there was no significant difference between the VH *cong* and RH *cong* conditions ($t_{(23)} = 1.19$, $p = 0.25$), but a significant difference between the VH *incong* and RH *incong* conditions ($t_{(23)} = 6.59$, $p = 0.000001$). In other words, in incongruent conditions participants aligned the phase of the virtual hand's movements significantly better with the dot's phasic size change when given the 'virtual hand' than the 'real hand' instruction. Furthermore, while the phase shift of the *real* hand's movements was larger during VH *incong* > VH *cong* ($t_{(23)} = 9.37$, $p = 0.000000003$)—corresponding to the smaller phase shift, and therefore better target phase-matching, of the virtual hand in these conditions—participants also exhibited a significantly larger shift of their real hand's movements during RH *incong* > RH *cong* ($t_{(23)} = 4.31$, $p = 0.0003$). Together, these results show that participants allocated their attentional resources to the respective instructed modality (vision or proprioception), and that this was accompanied by significantly better target tracking in each case—as expected based on the active inference formulation, and as suggested by the simulation results.

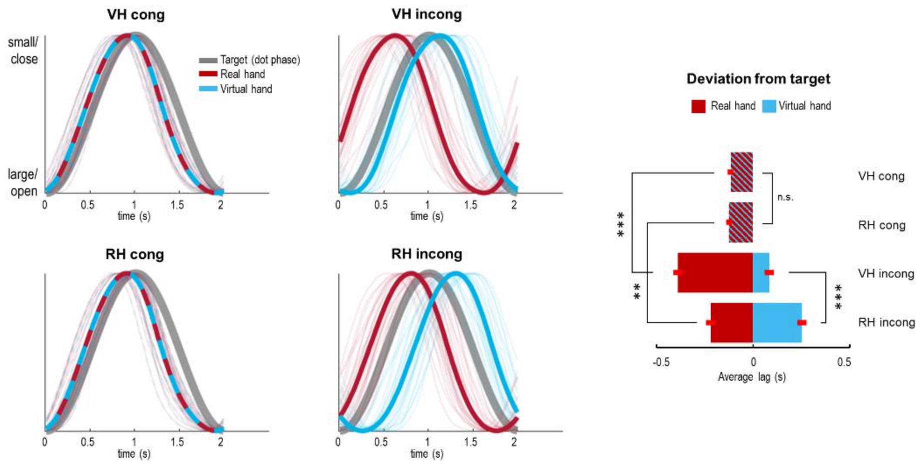

Figure 5. Task performance of our participants. Left: Average normalized trajectories of the real hand's (red) and the virtual hand's (blue) grasping movements relative to the oscillation of the target (pulsating fixation dot, grey) in each condition. The individual participant's average normalized trajectories in each condition are shown as thin lines. In the congruent conditions, the virtual hand's and the real hand's movements were identical, whereas the virtual hand's movements were delayed by 500 ms in the incongruent conditions. Right: The bar plot shows the corresponding average deviation (lag in seconds) of the real hand (red) and the virtual hand (blue) from the target in each condition, with associated standard errors of the mean. Crucially, there was a significant interaction effect between task and congruence; participants aligned the virtual hand's movements better with the target's oscillation in the VH *incong* > RH *incong* condition (and correspondingly, the real hand's movements in the RH *incong* > VH *incong* condition), in the absence of a significant difference between the congruent conditions. Bonferroni-corrected significance: $^{**}p < 0.01$, $^{***}p < 0.001$.

Discussion

We have shown that behaviour in a manual hand-target phase matching task, under visuo-proprioceptive conflict, benefits from adjusting the balance of visual versus proprioceptive precision by increased attention to either task-relevant modality. Our results generally support a predictive coding formulation of active inference, where visual and proprioceptive cues affect multimodal beliefs that drive action—depending on the relative precision afforded to each modality (Friston et al., 2010; Brown et al., 2013). Firstly, a simulated agent exhibited better phase matching when the expected sensory precision of the instructed 'task-relevant' modality (i.e., attention to vision or proprioception) was increased relative to the 'task-irrelevant' modality. This effect was reversed when attention was increased to the 'task-irrelevant' modality, effectively corresponding to cross-modal distraction. These results suggest that more precise sensory prediction errors have a greater impact on belief updating—which in turn guide goal-directed action. Our simulations also suggested that intersensory conflict—and its possible partial resolution—was based on a prior belief that one's hand movements generate matching visual and proprioceptive sensations. In an agent holding the unrealistic belief that visual and proprioceptive postures are per default unrelated, no evidence for an influence of intersensory conflict on target tracking was observed. Secondly, the self-report ratings of attentional allocation and the behaviour exhibited by human participants performing the same task, in a virtual reality environment, suggested an analogous mechanism: Our participants reported shifting their attention to the respective instructed modality (vision or proprioception)—and they were able to correspondingly align either vision or proprioception with an abstract target (oscillatory phase) under intersensory conflict. Together, our results suggest a tight link between precision control, attention, and multisensory integration in action—conforming to the principles of reciprocal message passing under hierarchical predictive coding for active inference, whereby the brain can choose how much to rely on specific sensory cues, and how strongly to resolve these prediction errors by action, in a given context.

Previous work on causal inference models has shown that Bayes-optimal cue integration can explain a variety of multisensory phenomena under intersensory conflict, including the recalibration of the less precise modality onto the more precise one (van Beers et al., 1999; Deneve et al., 2001; Ernst & Banks, 2002; Körding et al., 2007; Ma & Pouget, 2008; Kayser & Shams, 2015; Samad et al., 2015; Rohe & Noppeney, 2016, 2018). Our work advances on these findings by showing that fine-tuning the expected precision of two conflicting sources of bodily information (i.e., seen or felt hand posture) enhances the accuracy of goal-directed action (i.e., target tracking) with the respective 'attended' modality. Thus, we did not model hierarchical causal *perceptual* inference (we implemented a prior belief about a single cause of seen and felt hand postures in our model), but a simple case of *active* inference. Specifically, we showed that action itself was influenced by instructed attentional allocation, via augmentation of the impact of sensory prediction errors on model estimates of the 'attended' modality relative to the 'unattended' one.

Previous simulation studies—using a predictive coding formulation of active inference—have demonstrated that cued reaching movements to visual targets rely on a flexible balance of visual and proprioceptive precision (Friston et al., 2010; Friston, 2012; cf. Perrinet et al., 2014). Recently, experimental evidence for context dependent precision-modulation during action has been provided by a demonstration of a compensatory increase in sensory precision when participants were eye-tracking a noisy visual target stimulus (Adams et al., 2015). Our results complement these demonstrations by showing a context-dependent effect of attentional alteration of visual versus proprioceptive precision on behaviour under intersensory conflict.

More generally, our results support the notion that an endogenous attentional 'set' (Posner et al., 1978) can influence the precision afforded to vision or proprioception during action, and thus to prioritize either modality for a current behavioural context. Several studies have shown that visuo-proprioceptive recalibration is context dependent in that either vision or proprioception may be the 'dominant' modality—with corresponding recalibration of the 'non-dominant' modality (Warren & Cleaves, 1971; Kelso et al., 1975; Posner et al., 1976; Redding et al., 1985; Foulkes & Miall, 2000; Ingram et al., 2000; Foxe & Simpson, 2005; Cressman & Henriques, 2009; Rand & Heuer, 2019). Thus, our results lend (at least tentative) support to arguments that visuo-proprioceptive (or visuo-motor) adaptation and recalibration can be enhanced by increasing the precision of visual information (attending to vision; cf. Kelso et al., 1975; Posner et al., 1976). Notably, our results also suggest that the reverse can be true; i.e., that visuo-proprioceptive recalibration can be counteracted by increasing one's attention to proprioception. In sum, our results suggest that updating the predictions of a 'body model' affects goal-directed action. However, as it has been suggested that prediction updating may happen without control updating (Mathew et al., 2018), future work could establish whether the effects observed in our study can have long-lasting impact on the (generalizable) learning of motor control.

A noteworthy difference between our simulation results and the result of the behavioural experiment was that our participants exhibited a more pronounced shift of their real movements in the 'real hand' condition (which partly aligned the delayed virtual hand with the target's phase). This effect was reminiscent of the behaviour of our simulated agent under 'high distraction' (i.e., attention to the task-irrelevant modality) and occurred despite the fact that, as indicated by the ratings, participants focused on their real hand and tried to comply with the task instructions. Interestingly, however, our participants reported the 'real hand' task to be easier than the 'virtual hand' task under visuo-proprioceptive incongruence—which suggests that they did not *notice* their 'incorrect' behavioural adjustment. In contrast, the simulated agent even showed slightly better RH than VH alignment—this can be explained by the fact that proprioception was 'naturally' the modality driving movement, while the vision was experimentally delayed (which had to be inferred by the agent).

One tentative interpretation of the much stronger visual bias in the behavioural experiment is possible in light of predictive coding formulations of shared body representation and self-other distinction; i.e., the relative balance between visual and proprioceptive prediction errors to decide whether 'I am observing an action' or whether 'I am moving' (Kilner et al., 2003, 2007; Friston, 2012; cf. Vasser et al., 2019). Generally, visual prediction errors have to be attenuated during action observation to prevent actually performing (i.e., mirroring) the observed movement (Friston, 2012). However, several studies have demonstrated 'automatic' imitative tendencies during action observation, reminiscent of 'echopraxia', which are extremely hard to inhibit—for example, seeing an incongruent finger or arm movement biases participants' own movement execution (Brass et al., 2001; Kilner et al., 2003). In a predictive coding framework, this can be formalized as an 'automatic' update of multimodal beliefs driving action by precise (not sufficiently attenuated) visual body information (cf. Kilner et al., 2007). Such an interpretation would be in line with speculations that participants in visuo-motor conflict tasks attend to vision, rather than proprioception, if not instructed otherwise (Kelso et al., 1975; Posner et al., 1976; Kelso, 1979; cf. Redding et al., 1985). Whether or not such effects—an 'automatic' influence of a seen visual hand posture that is incongruent (note that this could mean leading or lagging in our case) to the felt one—can account for our behavioural results could be clarified by future work. Likewise, an interesting question is whether these effects could perhaps be reduced by actively ignoring or 'dis-attending' (Clark, 2015; Limanowski, 2017) away from vision. An analogous mechanism has been tentatively suggested by observed benefits of proprioceptive attenuation—thereby increasing the relative impact of visual information—during visuo-motor adaptation and visuo-proprioceptive recalibration (Taub & Goldberg, 1974; Ingram et al., 2000; Balslev et al., 2004; Bernier et al., 2009; cf. Limanowski et al., 2015a,b; Zeller et al., 2016). These questions should best be addressed by combined behavioural and brain imaging experiments, to illuminate the neuronal correlates of the (supposedly attentional) precision weighting in the light of recently proposed implementations of predictive coding in the brain (Bastos et al., 2012; Shipp et al., 2013; Shipp, 2016).

It should be noted that our results need to be validated by future work using more complicated movement tasks (here, we focused on a simple, well-trained grasping movement), different target modalities (we used a visual, albeit non-spatial target), and more biophysically realistic models of motor (hand movement) control. Moreover, we interpret our simulation and empirical results in terms of evidence for top-down precision modulation, which corresponds to the process of 'attention' within the active inference account of predictive coding (Feldman & Friston, 2010; Edwards et al., 2012; Brown et al., 2013). This interpretation needs to be applied with some caution to the behavioural results, as we can only infer any attentional effects from the participants' self-reports. We assume that participants monitored their behaviour continuously, but with the present data we cannot rule out that movements might have been executed automatically between discrete time points at which behaviour was monitored. Future work could therefore use explicit measures of attention, perhaps supplemented by forms of supervision, to validate behavioural effects. Finally, our experimental design is not able to disentangle the (likely interdependent) effects of sensory noise and attention (i.e., expected precision). Therefore, another important question for future research is the potential attentional compensation of experimentally added sensory noise (e.g., via jittering or blurring the visual hand or via tendon vibration in the proprioceptive domain, cf. Jaeger et al., 1979), whereby it should be remembered that these manipulations may in themselves be 'attention-grabbing' (Beauchamp et al., 2010).

Methods

Task design

We used the same task design in the simulations and the behavioural experiment (see Fig. 1). For consistency, we will describe the task as performed by our human participants, but the same principles apply to the simulated agent. We designed our task as a non-spatial modification of a previously used hand-target tracking task (cf. Limanowski et al., 2017). The participant (or simulated agent) had to perform repetitive grasping movements paced by sinusoidal fluctuations in the size of a central fixation dot (sinusoidal oscillation at 0.5 Hz). Thus, this task was effectively a phase matching task, which we hoped to be less biased towards the visual modality due to a more abstract target quantity (oscillatory size change vs spatially moving target, as in previous studies). The fixation dot was chosen as the target to ensure that participants had to fixate the centre of the screen (and therefore look at the virtual hand) in all conditions. Participants (or the simulated agent) controlled a virtual hand model via a data glove worn on their unseen right hand (details below). In this way, vision (*seen* hand position via the virtual hand) could be decoupled from proprioception (*felt* hand position). In half of the movement trials, temporal delay of 500 ms between visual and proprioceptive hand infromation was introduced by delaying vision (i.e., the seen hand movements) with respect to proprioception (i.e., the unseen hand movements performed by the participant or agent). In other words, the seen and felt hand positions were always incongruent (phase-shifted) in these conditions. Crucially, the participant (agent) had to perform the phase matching task with one of two goals in mind: to match the target's oscillatory phase with the seen virtual hand movements (vision) or with the unseen real hand movements (proprioception). This resulted in a 2 x 2 factorial design with the factors 'visuo-proprioceptive congruence' (congruent, incongruent) and 'instructed modality' (vision, proprioception).

Simulations

We based our simulations on predictive coding formulations of active inference as situated within a free energy principle of brain function, which has been used in many previous publications to simulate perception and action (e.g. Friston et al., 2010; Friston, 2012; Brown & Friston, 2013; Perrinet et al., 2014; Adams et al., 2015). Here, we briefly review the basic assumptions of this scheme (please see the above literature for details).

Hierarchical predictive coding rests on a probabilistic mapping of hidden causes to sensory consequences, as described by a hierarchical generative model, where each level of the model encodes conditional expectations ('beliefs'; which here refer to subpersonal or non-propositional Bayesian beliefs in the sense of Bayesian belief updating and belief propagation; i.e., posterior probability densities) about states of the world that best explains states of affairs encoded at lower levels or—at the lowest sensory level—sensory input. Thus, the hierarchy provides a deep model of how current sensory input is generated from causes in the environment; where increasingly higher-level beliefs represent increasingly abstract (i.e., hidden or latent) states of the environment. The generative model therefore maps from unobservable causes (hidden states) to observable consequences (sensory states). Model inversion corresponds to inferring the causes of sensations; i.e., mapping from consequences to causes. Operationally, this inversion rests upon the minimisation of free energy or 'surprise' approximated in the form of prediction error. In other words, prediction errors are used to update expectations to accommodate or 'explain away' ascending prediction error. This corresponds to Bayesian filtering or predictive coding (Rao & Ballard, 1999; Friston & Kiebel, 2009; Bastos et al., 2012)—which, and the linear assumptions, is formally identical to linear quadratic control in motor control theory (Todorov, 2008). In such an architecture, descending connections convey predictions suppressing activity in the cortical level immediately below, and ascending connections return prediction error (i.e., sensory data not explained by descending predictions). Crucially, the ascending prediction errors are precision-weighted (where precision corresponds to the inverse variance), so that a prediction error that is afforded a greater precision has a stronger impact on belief updating.

Active inference extends hierarchical predictive coding from the sensory to the motor domain; i.e., by equipping standard Bayesian filtering schemes (a.k.a. predictive coding) with classical reflex arcs that enable action (e.g., a hand movement) to fulfil predictions about hidden states of the world. In brief, desired movements are specified in terms of prior beliefs about state transitions (policies), which are then realised by action; i.e., by sampling or generating sensory data that provide evidence for those beliefs (Perrinet et al., 2014). Thus, action is also driven by optimisation of the model via suppression of prediction error: movement occurs because high-level multi- or amodal prior beliefs about behaviour predict proprioceptive and exteroceptive (visual) states that would ensue if the movement was performed (e.g., a particular limb trajectory). Prediction error is then suppressed throughout a motor hierarchy; ranging from intentions and goals over kinematics to muscle activity (Kilner et al., 2007; cf. Grafton & Hamilton, 2007). At the lowest level of the hierarchy, spinal reflex arcs suppress proprioceptive prediction error by enacting the predicted movement, which also implicitly minimises exteroceptive prediction error; e.g. the predicted visual consequences of the action (Adams et al 2013, Friston 2011, Shipp et al 2013). Thus, via embodied interaction with its environment, an agent can reduce its model's free energy ('surprise' or, under specific assumptions, prediction error) or, in other words, maximise Bayesian model evidence. Put succinctly, all action is in the service of self-evidencing (Hohwy, 2016).

Following the above notion of active inference, one can describe action and perception as the solution to coupled differential equations describing the dynamics of the real world (boldface) and the behaviour of an agent (italics, cf. Friston et al., 2010 for details).

$$\begin{aligned} \boldsymbol{s} &= \boldsymbol{g}(\boldsymbol{x},\boldsymbol{v},\boldsymbol{a}) + \boldsymbol{\omega}_v \\ \dot{\boldsymbol{x}} &= \boldsymbol{f}(\boldsymbol{x},\boldsymbol{v},\boldsymbol{a}) + \boldsymbol{\omega}_x \\ \dot{a} &= -\partial_a F(\tilde{s},\tilde{\mu}) \\ \dot{\tilde{\mu}} &= D\tilde{\mu} - \partial_{\tilde{\mu}} F(\tilde{s},\tilde{\mu}) \end{aligned} \qquad (1)$$

The first pair of coupled stochastic (i.e., subject to random fluctuations $\boldsymbol{\omega}_x$, $\boldsymbol{\omega}_v$) differential equations describes the dynamics of hidden states and causes in the world and how they generate sensory states. Here, ***(s, x, v, a)*** denote sensory input, hidden states, hidden causes and action in the real world, respectively. The second pair of equations corresponds to action and perception, respectively—they constitute a (generalised) gradient descent on variational free energy, known as an evidence bound in machine learning (Winn & Bishop, 2005). The differential equation describing perception corresponds to generalised filtering or predictive coding. The first term is a prediction based upon a differential operator *D* that returns the generalised motion of conditional (i.e., posterior) expectations about states of the world, including the motor plant (vector of velocity, acceleration, jerk, *etc*.). Here, the variables $(\tilde{s}, \tilde{\mu}, a)$ correspond to generalised sensory input, conditional expectations and action, respectively. Generalised coordinates of motion, denoted by the ~ notation, correspond to a vector representing the different orders of motion (position, velocity, acceleration, *etc*.) of a variable. The differential equations above are coupled because sensory states depend upon action through hidden states and causes $(\boldsymbol{x}, \boldsymbol{v})$ while action $a(t) = \boldsymbol{a(t)}$ depends upon sensory states through internal states $\tilde{\mu}$. Neurobiologically, these equations can be considered to be implemented in terms of predictive coding; i.e., using prediction errors on the motion of hidden states—such as visual or proprioceptive cues about hand position—to update beliefs or expectations about the state of the lived world and embodied kinematics.

By explicitly separating hidden real-world states from the agent's expectations as above, one can separate the generative process from the updating scheme that minimises free energy. To perform simulations using this scheme, one solves Eq. 1 to simulate (neuronal) dynamics that encode conditional expectations and ensuing action. The generative model thereby specifies a probability density function over sensory inputs and hidden states and causes, which is needed to define the free energy of sensory inputs:

$$\begin{aligned}
s &= g^{(1)}(x^{(1)}, v^{(1)}) + \omega_v^{(1)} \\
\dot{x}^{(1)} &= f^{(1)}\left(x^{(1)}, v^{(1)}\right) + \omega_x^{(1)} \\
&\vdots \\
v^{(i-1)} &= g^{(i)}\left(x^{(i)}, v^{(i)}\right) + \omega_v^{(i)} \\
\dot{x}^{(i)} &= f^{(i)}\left(x^{(i)}, v^{(i)}\right) + \omega_x^{(i)} \\
&\vdots
\end{aligned} \qquad (2)$$

This probability density is specified in terms of nonlinear functions of hidden states and causes $(f^{(i)}, g^{(i)})$ that generate dynamics and sensory consequences, and Gaussian assumptions about random fluctuations $(\omega_x^{(i)}, \omega_v^{(i)})$ on the motion of hidden states and causes. These play the role of sensory noise or uncertainty about states. The precisions of these fluctuations are quantified by ($\Pi_x^{(i)}$, $\Pi_v^{(i)}$), which are the inverse of the respective covariance matrices.

Given the above form of the generative model (Eq. 2), we can now write down the differential equations (Eq. 1) describing neuronal dynamics in terms of prediction errors on the hidden causes and states as follows:

$$\begin{aligned}
\dot{\tilde{\mu}}_x^{(i)} &= D\tilde{\mu}_x^{(i)} + \frac{\partial \tilde{g}^{(i)}}{\partial \tilde{\mu}_x^{(i)}} \cdot \Pi_v^{(i)} \tilde{\varepsilon}_v^{(i)} + \frac{\partial \tilde{f}^{(i)}}{\partial \tilde{\mu}_x^{(i)}} \cdot \Pi_x^{(i)} \tilde{\varepsilon}_x^{(i)} - D \cdot \Pi_x^{(i)} \tilde{\varepsilon}_x^{(i)} \\
\dot{\tilde{\mu}}_v^{(i)} &= D\tilde{\mu}_v^{(i)} + \frac{\partial \tilde{g}^{(i)}}{\partial \tilde{\mu}_v^{(i)}} \cdot \Pi_v^{(i)} \tilde{\varepsilon}_v^{(i)} + \frac{\partial \tilde{f}^{(i)}}{\partial \tilde{\mu}_v^{(i)}} \cdot \Pi_x^{(i)} \tilde{\varepsilon}_x^{(i)} - \Pi_v^{(i+1)} \tilde{\varepsilon}_v^{(i+1)} \\
\tilde{\varepsilon}_x^{(i)} &= D\tilde{\mu}_x^{(i)} - \tilde{f}^{(i)}\left(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)}\right) \\
\tilde{\varepsilon}_v^{(i)} &= \tilde{\mu}_v^{(i-1)} - \tilde{g}^{(i)}(\tilde{\mu}_x^{(i)}, \tilde{\mu}_v^{(i)})
\end{aligned} \tag{3}$$

The above equation (Eq. 3) describes recurrent message passing between hierarchical levels to suppress free energy or prediction error (i.e., predictive coding, cf. Friston & Kiebel, 2009; Bastos et al., 2012). Specifically, error units receive predictions from the same hierarchical level and the level above. Conversely, conditional expectations (‘beliefs’, encoded by the activity of state units) are driven by prediction errors from the same level and the level below. These constitute bottom-up and lateral messages that drive conditional expectations towards a better prediction to reduce the prediction error in the level below—this is the sort of belief updating described in the introduction.

Finally, we can now add action as the specific sampling of predicted sensory inputs. As noted above, along active inference, high-level beliefs (conditional expectations) elicit action by sending predictions down the motor (proprioceptive) hierarchy to be unpacked into proprioceptive predictions at the level of (pontine) cranial nerve nuclei and spinal cord, which are then ‘quashed’ by movement so that predicted movements are enacted to ‘fulfil’ predictions.

$$\dot{a} = \; -\partial_a F = -(\partial_a \tilde{\varepsilon}_v^{(1)}) \cdot \Pi_v^{(1)} \tilde{\varepsilon}_v^{(1)} \tag{4}$$

In our case, the generative process and model used for simulating the target tracking task are straightforward (using just a single level) and can be expressed as follows:

$$\dot{\boldsymbol{x}} = \begin{bmatrix} \dot{\boldsymbol{x}}_t \\ \dot{\boldsymbol{x}}_h \end{bmatrix} = \begin{bmatrix} t_t \\ \dfrac{1}{t_a} a \end{bmatrix} + \boldsymbol{\omega}_v$$

$$\boldsymbol{s} = \begin{bmatrix} \boldsymbol{s}_t \\ \boldsymbol{s}_p \\ \boldsymbol{s}_v \end{bmatrix} = \begin{bmatrix} \sin(\boldsymbol{x}_t) \\ \sin(\boldsymbol{x}_h) \\ \sin(\boldsymbol{x}_h - \boldsymbol{v}) \end{bmatrix} + \boldsymbol{\omega}_x$$

$$\dot{x} = \begin{bmatrix} \dot{x}_t \\ \dot{x}_h \end{bmatrix} = \begin{bmatrix} t_t \\ x_t - x_h \end{bmatrix} + \omega_v$$

$$s = \begin{bmatrix} s_t \\ s_p \\ s_v \end{bmatrix} = \begin{bmatrix} \sin(x_t) \\ \sin(x_h) \\ \sin(x_h - v) \end{bmatrix} + \omega_x \tag{5}$$

The first pair of equations describe the *generative process*; i.e., a noisy sensory mapping from hidden states and the equations of motion for states in the real world. In our case, the real-world variables comprised two hidden states **x_t** (the state of the target) and **x_h** (the state of the hand), which generate sensory inputs; i.e., proprioceptive **s_p** and visual **s_v** cues about hand posture , and visual cues about the target's size **s_t**. Note that to simulate sinusoidal movements—as used in the experimental task—sensory cues pertaining to the target and hand are mapped via sine functions of the respective hidden states (plus random fluctuations). Both target and hand states change linearly over time, and become sinusoidal movements via the respective sensory mapping from causes to sensory data. We chose this solution in our particular case for a straightforward implementation of phase shifts (visuo-proprioceptive incongruence) via subtraction of a constant term from the respective sensory mapping (*v*, see below). Thus, the target state x_t is perturbed by hidden causes at a constant rate (t_t = 1/40), i.e., it linearly increases over time. This results in one oscillation of a sinusoidal trajectory via the sensory mapping sin($\boldsymbol{x_t}$)—corresponding to one growing-and-shrinking of the fixation dot, as in the behavioural experiment—during 2 seconds (the simulations proceeded in time bins of 1/120 seconds, see Fig. 2). The hand state is driven by action *a* with a time constant of t_a = 16.67 ms, which induced a slight 'sluggishness' of movement mimicking delays in motor execution. Action thus describes the rate of change of hand posture along a linear trajectory—at a rate of 0.05 per time bin—which again becomes an oscillatory postural change (i.e., a grasping movement) via the sinusoidal sensory mapping. The hidden cause *v* modelled the displacement of proprioceptive and visual hand posture information in a virtue of being subtracted within the sinusoidal sensory mapping from the hidden hand state to visual sensory information sin($\boldsymbol{x_t}$-$\boldsymbol{v}$). In other words, $\boldsymbol{v}$ = 0 when the virtual hand movements were congruent, and $\boldsymbol{v}$ = 0.35 (corresponding to about 111 ms delay) when the virtual hand's movements were delayed with respect to the real hand. Note that random fluctuations in the process generating sensory input were suppressed by using high precisions on the errors of the sensory states and motion in the generative process (exp(16) = 8886110). This can be thought of as simulating the average response over multiple realizations of random inputs, under any particular precisions assumed by the generative model. If random fluctuations are introduced into the generative process, each solution or realized response itself becomes a random variable. Therefore, the single movement we simulated in each condition may be interpreted as a participant-specific average over realizations; i.e., in which the effects of random fluctuations are averaged out (cf. Perrinet et al., 2014; Adams et al., 2015). This ensured that our simulations reflect systematic differences depending on the parameter values chosen to reflect alterations of sensory attention via changing parameters of the agent's model (as described below).

The second pair of equations describe the agent's *generative model* of how sensations are generated using the form of Eq. 2. These define the free energy in Eq. 1 and specify behaviour (under active inference). The generative model has the same form as the generative process, with the important exceptions that there is no action and the state of the hand is driven by the displacement between the hand and the target $x_t - x_h$. In other words, the agent believes that its grasping movements will follow the target's oscillatory size change, which is itself driven by some unknown force at a constant rate (and thus producing an oscillatory trajectory as in the generative process). This effectively models (the compliance with) the task instruction, under the assumption that participants already know about the oscillatory phase of the target; i.e., they have been well trained. Importantly, this formulation models the 'real hand' instruction; under the 'virtual hand' instruction, the state of the hand was driven by $x_t - (x_h - v)$, reflecting the fact that any perceived visual delay (i.e., the inferred displacement of vision from proprioception v) should now also be compensated to keep the virtual hand aligned with the target's oscillatory phase under incongruence; the initial value for v was set to represent the respective information about visuo-proprioceptive congruence, i.e., 0 for congruent movement conditions and 0.35 for incongruent movement conditions. We defined the agent's model to entertain a prior belief that visual and proprioceptive cues are normally congruent (or, for comparison, incongruent). This was implemented by setting the prior expectation of the cause v to 0 (indicating congruence of visual and proprioceptive hand posture information), with a log precision of 3 (corresponding to about 20.1). In other words, the hidden cause could vary, a priori, with a standard deviation of about exp(-3/2) = 0.22. This mimicked the learned association between seen and felt hand positions (under a minimal degree of flexibility), which is presumably formed over a lifetime and very hard to overcome and underwrites phenomena like the 'rubber hand illusion' (Botvinick & Cohen, 1998; see Introduction).

Crucially, the agent's model included a precision-weighting of the sensory signals—as determined by the active deployment of attention along predictive coding accounts of active inference. This allowed us to manipulate the precision assigned to proprioceptive or visual prediction errors (Π_p, Π_v) that, per default, were given a log precision of 3 and 4, respectively (corresponding to 20.1 and 54.6, respectively). This reflects the fact that vision usually is afforded a higher precision than proprioception in hand position estimation (e.g. van Beers et al., 1999; cf. Kelso et al., 1975). To implement increases in task-related (selective) attention, we increased the log precision of prediction errors from the instructed modality (vision or proprioception) by 1 in each case (i.e., by a factor of about 2.7); in an alternative scenario, we tested for incorrect allocation of attention to the non-instructed or 'distractor' modality by increasing the precision of the appropriate prediction errors. We did not simulate increases in both sensory precisions, because our study design was tailored to investigate selective attention as opposed to divided attention. Note that in the task employed, divided attention was precluded, since attentional set was induced via instructed task-relevance; i.e., attempted target phase-matching. In other words, under incongruence, only one modality could be matched to the target. The ensuing generative process and model are, of course, gross simplifications of a natural movement paradigm. However, this formulation is sufficient to solve the active inference scheme in Eq. 1 and examine the agent's behaviour under the different task instructions and, more importantly, under varying degrees of selectively enhanced sensory precision afforded by an attentional set.

Experiment

26 healthy, right-handed volunteers (15 female, mean age = 27 years, range = 19-37, all with normal or corrected-to-normal vision) participated in the experiment, after providing written informed consent. Two participants were unable to follow the task instructions during training and were excluded from the main experiment, resulting in a final sample size of 24. The experiment was approved by the local research ethics committee (University College London) and conducted in accordance with the usual guidelines.

During the experiment, participants sat at a table wearing an MR-compatible data glove (5DT Data Glove MRI, 1 sensor per finger, 8 bit flexure resolution per sensor, 60 Hz sampling rate) on their right hand, which was placed on their lap under the table. The data glove measured the participant's finger flexion via sewn-in optical fibre cables; i.e., each sensor returned a value from 0 to 1 corresponding to minimum and maximum flexion of the respective finger. These raw data were fed to a photorealistic virtual right hand model (cf. Limanowski et al., 2017), whose fingers were thus moveable with one degree of freedom (i.e., flexion-extension) by the participant, in real-time. The virtual reality task environment was instantiated in the open-source 3D computer graphics software Blender (http://www.blender.org) using a Python programming interface, and presented on a computer screen at about 60 cm distance (1280 x 1024 pixels resolution).

The participants' task was to perform repetitive right-hand grasping movements paced by the oscillatory size change of the central fixation do, which continually decreased-and-increased in size sinusoidally (12 % size change) at a frequency of 0.5 Hz; i.e., this was effectively a phase matching task (Fig. 1).

The participants had to follow the size changes with right-hand grasping movements; i.e., to close the hand when the dot shrunk and to open the hand when the dot grew. In half of the movement trials, an incongruence between visual and proprioceptive hand information was introduced by delaying the virtual hand's movements by 500 ms with respect to the movements performed by the participant. In other words, the virtual hand and the real hand were persistently in incongruent (mismatching) postures in these conditions. The delay was clearly perceived by all participants.

Participants performed the task in trials of 32 seconds (16 movement cycles; the last movement was signalled by a brief blinking of the fixation dot), separated by 6 second fixation-only periods. The task instructions ('VIRTUAL' / 'REAL') were presented before each respective movement trials for 2 seconds. Additionally, participants were informed whether in the upcoming trial the virtual hand's movements would be synchronous ('synch.') or delayed ('delay'). The instructions and the fixation dot in each task were coloured (pink or turquoise, counterbalanced across participants), to help participants remember the current task instruction during each movement trial. Participants practised the task until they felt confident, and then completed two runs of 8 min length. Each of the four conditions 'virtual hand task under congruence' (VH cong), 'virtual hand task under incongruence' (VH incong), 'real hand task under congruence' (RH cong), and 'real hand task under incongruence' (RH incong) was presented 3 times per run, in randomized order.

To analyse the behavioural change in terms of deviation from the target (i.e., phase shift from the oscillatory size change), we averaged and normalized the movement trajectories in each condition for each participant (raw data were averaged over the four fingers, no further pre-processing was applied). We then calculated the phase shift as the average angular difference between the raw averaged movements of the virtual or real hand and the target's oscillatory pulsation phase in each condition, using a continuous wavelet transform. The resulting phase shifts for each participant and condition were then entered into a 2 x 2 repeated measures ANOVA with the factors task (virtual hand, real hand) and congruence (congruent, incongruent) to test for statistically significant group-level differences. Post-hoc t-tests (two-tailed, with Bonferroni-corrected alpha levels to account for multiple comparisons) were used to compare experimental conditions.

After the experiment, participants were asked to indicate—for each of the four conditions separately—their answers to the following two questions: "How difficult did you find the task to perform in the following conditions?" (Q1, answered on a 7-point visual analogue scale from "very easy" to "very difficult") and "On which hand did you focus your attention while performing the task?" (Q2, answered on a 7-point visual analogue scale from "I focused on my real hand" to "I focused on the virtual hand"). The questionnaire ratings were evaluated for statistically significant differences using a nonparametric Friedman's test and Wilcoxon's signed-rank test (with Bonferroni-corrected alpha levels to account for multiple comparisons) due to non-normal distribution of the residuals.

Acknowledgements: We thank Thomas Parr for helpful comments on the simulations.

References

Adams RA, Shipp S, Friston KJ. Predictions not commands: active inference in the motor system. Brain Structure and Function. 2013;218:611-643.

Adams RA, Aponte E, Marshall L, Friston KJ. Active inference and oculomotor pursuit: The dynamic causal modelling of eye movements. Journal of Neuroscience Methods. 2015;242:1-14.

Balslev D, Christensen LO, Lee JH, Law I, Paulson OB, Miall RC. Enhanced accuracy in novel mirror drawing after repetitive transcranial magnetic stimulation-induced proprioceptive deafferentation. Journal of Neuroscience. 2004;24(43):9698-9702.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. Canonical microcircuits for predictive coding. Neuron. 2012;76(4):695-711.

Beauchamp MS, Pasalar S, Ro T. Neural substrates of reliability-weighted visual-tactile multisensory integration. Frontiers in Systems Neuroscience. 2010;4:25.

Bernier PM, Burle B, Vidal F, Hasbroucq T, Blouin J. Direct evidence for cortical suppression of somatosensory afferents during visuomotor adaptation. Cerebral Cortex. 2009;19(9):2106-2113.

Botvinick M, Cohen J. Rubber hands ‘feel’touch that eyes see. Nature. 1998;391(6669):756.

Brass M, Zysset S, von Cramon DY. The inhibition of imitative response tendencies. Neuroimage. 2011;14(6):1416-1423.

Brown H, Adams A, Parees I, Edwards M, Friston KJ. Active inference, sensory attenuation and illusions. Cognitive processing. 2013;14(4):411-427.

Clark A. Surfing uncertainty: Prediction, action, and the embodied mind. Oxford: Oxford University Press; 2015.

Cressman EK, Henriques DY. Sensory recalibration of hand position following visuomotor adaptation. Journal of neurophysiology. 2009;102(6):3505-3518.

Deneve S, Latham PE, Pouget A. Efficient computation and cue integration with noisy population codes. Nature neuroscience. 2001;4:826.

Edwards MJ, Adams RA, Brown H, Parees I, Friston K. A Bayesian account of ‘hysteria’. Brain. 2012;135(11):3495-3512.

Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. Nature. 2002;415:429.

Feldman H, Friston K. Attention, uncertainty, and free-energy. Frontiers in human neuroscience. 2010;4:215.

Foulkes AJM, Miall RC. Adaptation to visual feedback delays in a human manual tracking task. Experimental Brain Research. 2000;131(1):101-110.

Foxe JJ, Simpson GV. Biasing the brain’s attentional set: II. Effects of selective intersensory attentional deployments on subsequent sensory processing. Experimental brain research. 2005;166(3-4):393-401.

Friston K. The free-energy principle: a unified brain theory?. Nature reviews neuroscience. 2010;11(2):127.

Friston K. What is optimal about motor control? Neuron. 2011;72:488-98.

Friston K. Prediction, perception and agency. International Journal of Psychophysiology. 2012;83(2):248-252.

Friston K, Kiebel S. Predictive coding under the free-energy principle. Philosophical Transactions of the Royal Society B: Biological Sciences. 2009;364(1521):1211-1221.

Friston KJ, Daunizeau J, Kilner J, Kiebel SJ. Action and behavior: a free-energy formulation. Biological cybernetics. 2010;102(3):227-260.

Grafton ST, Hamilton AFDC. Evidence for a distributed hierarchy of action representation in the brain. Human movement science. 2007;26(4):590-616.

Grafton ST, Schmitt P, Van Horn J, Diedrichsen J. Neural substrates of visuomotor learning based on improved feedback control and prediction. Neuroimage. 2008;39(3):1383-1395.

Grefkes C, Ritzl A, Zilles K, Fink GR. Human medial intraparietal cortex subserves visuomotor coordinate transformation. Neuroimage. 2004;23:1494-1506.

Heed T, Gründler M, Rinkleib J, Rudzik FH, Collins T, Cooke E, O'Regan JK. Visual information and rubber hand embodiment differentially affect reach-to-grasp actions. Acta psychologica. 2011;138(1):263-271.

Holmes NP, Crozier G, Spence C. When mirrors lie:“Visual capture” of arm position impairs reaching performance. Cognitive, Affective, & Behavioral Neuroscience. 2004;4(2):193-200.

Holmes NP, Snijders HJ, Spence C. Reaching with alien limbs: Visual exposure to prosthetic hands in a mirror biases proprioception without accompanying illusions of ownership. Perception & Psychophysics. 2006;68(4):685-701.

Hohwy J. The Self-Evidencing Brain. Noûs. 2006;50:259-85.

Ingram HA, Van Donkelaar P, Cole J, Vercher JL, Gauthier GM, Miall RC. The role of proprioception and attention in a visuomotor adaptation task. Experimental Brain Research. 2000;132(1):114-126.

Jaeger RJ, Agarwal GC, Gottlieb GL. Directional errors of movement and their correction in a discrete tracking task. Journal of motor behavior. 1979;11:123-133.

Kayser C, Shams L. Multisensory causal inference in the brain. PLoS biology. 2015;13:e1002075.

Kelso JS. Motor-sensory feedback formulations: are we asking the right questions? Behavioral and Brain Sciences. 1979;2(1):72-73.

Kelso JA, Cook E, Olson ME, Epstein W. Allocation of attention and the locus of adaptation to displaced vision. Journal of Experimental Psychology: Human Perception and Performance. 1975;1(3):237.

Kilner JM, Paulignan Y, Blakemore SJ. An interference effect of observed biological movement on action. Current biology. 2003;13(6):522-525.

Kilner JM, Friston KJ, Frith CD. Predictive coding: an account of the mirror neuron system. Cognitive processing. 2007;8(3):159-166.

Körding KP, Wolpert DM. Bayesian integration in sensorimotor learning. Nature. 2004;427(6971):244.

Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. PLoS one. 2007;2:e943.

Limanowski J. (Dis-) attending to the Body: Action and Self-experience in the Active Inference Framework. In T. Metzinger & W. Wiese (Eds.), Philosophy and Predictive Processing. Frankfurt am Main: MIND Group; 2017.

Limanowski J, Blankenburg F. Network activity underlying the illusory self-attribution of a dummy arm. Human Brain Mapping. 2015a;36(6):2284-2304.

Limanowski J, Blankenburg F. That's not quite me: limb ownership encoding in the brain. Social cognitive and affective neuroscience. 2015b;11(7):1130-1140.

Limanowski J, Blankenburg F. Integration of visual and proprioceptive limb position information in human posterior parietal, premotor, and extrastriate cortex. Journal of Neuroscience. 2016;36(9):2582-2589.

Limanowski J, Kirilina E, Blankenburg F. Neuronal correlates of continuous manual tracking under varying visual movement feedback in a virtual reality environment. Neuroimage. 2017;146:81-89.

Ma WJ, Pouget A. Linking neurons to behavior in multisensory perception: A computational review. Brain research. 2008;1242:4-12.

Makin TR, Holmes NP, Ehrsson HH. On the other hand: dummy hands and peripersonal space. Behavioural brain research. 2008;191(1):1-10.

Mathew J, Bernier PM, Danion FR. Asymmetrical relationship between prediction and control during visuo-motor adaptation. eNeuro. 2018;5(6).

Ogawa K, Inui T, Sugio T. Neural correlates of state estimation in visually guided movements: an event-related fMRI study. Cortex. 2007;43:289-300.

Pavani F, Spence C, Driver J. Visual capture of touch: Out-of-the-body experiences with rubber gloves. Psychological science. 2000;11(5):353-359.

Perrinet LU, Adams RA, Friston KJ. Active inference, eye movements and oculomotor delays. Biological cybernetics. 2014;108(6):777-801.

Posner MI, Nissen MJ, Klein RM. Visual dominance: an information-processing account of its origins and significance. Psychological review. 1976;83(2):157.

Posner MI, Nissen MJ, Ogden WC. Attended and unattended processing modes: The role of set for spatial location. Modes of perceiving and processing information. 1978;137(158):2.

Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience.1999;2(1):79.

Rand MK, Heuer H. Visual and proprioceptive recalibrations after exposure to a visuomotor rotation. European Journal of Neuroscience. 2019.

Redding GM, Clark SE, Wallace B. Attention and prism adaptation. Cognitive psychology. 1985;17(1):1-25.

Rohe T, Noppeney U. Distinct computational principles govern multisensory integration in primary sensory and association cortices. Current Biology. 2016;26(4):509-514.

Rohe T, Noppeney U. Reliability-weighted integration of audiovisual signals can be modulated by top-down attention. Eneuro. 2018;5(1).

Samad M, Chung AJ, Shams L. Perception of body ownership is driven by Bayesian sensory inference. PloS one. 2015;10:e0117178.

Shadmehr R, Krakauer JW. A computational neuroanatomy for motor control. Experimental brain research. 2008;185(3):359-381.

Shipp S. Neural Elements for Predictive Coding. Frontiers in Psychology. 2016;7:1792.

Shipp S, Adams RA, Friston KJ. Reflections on agranular architecture: predictive coding in the motor cortex. Trends in Neuroscience. 2013;36:706-16.

Sober SJ, Sabes PN. Flexible strategies for sensory integration during motor planning. Nature neuroscience. 2005;8(4):490.

Taub E, Goldberg IA. Use of sensory recombination and somatosensory deafferentation techniques in the investigation of sensory-motor integration. Perception. 1974;3(4):393-408.

Todorov E. General duality between optimal control and estimation. In 47th IEEE Conference on Decision and Control (pp. 4286-4292). IEEE; 2008.

Tsakiris M, Haggard P. The rubber hand illusion revisited: visuotactile integration and self-attribution. Journal of Experimental Psychology: Human Perception and Performance. 2005;31(1):80.

Van Beers RJ, Sittig AC, Gon JJDVD. Integration of proprioceptive and visual position-information: An experimentally supported model. Journal of neurophysiology. 1999;81(3):1355-1364.

Van Beers RJ, Wolpert DM, Haggard P. When feeling is more important than seeing in sensorimotor adaptation. Current biology. 2002;12:834-837.

Vasser M, Vuillaume L, Cleeremans A, Aru J. Waving goodbye to contrast: self-generated hand movements attenuate visual sensitivity. Neuroscience of consciousness. 2019;1:niy013.

Warren DH, Cleaves WT. Visual-proprioceptive interaction under large amounts of conflict. Journal of Experimental Psychology. 1971;90:206-214.

Winn J, Bishop CM. Variational message passing. Journal of Machine Learning Research. 2005;6:661-94

Wolpert DM, Goodbody SJ, Husain M. Maintaining internal representations: the role of the human superior parietal lobe. Nature neuroscience. 1998;1(6):529.

Zeller D, Friston KJ, Classen JD. Dynamic causal modeling of touch-evoked potentials in the rubber hand illusion. Neuroimage. 2016;138:266-273.

Supporting Information Legends

Raw data recorded during the two experimental runs for each participant. Each data file contains the real and virtual finger values (recorded by the data glove) alongside timing and experimental condition information. Please see the file ‘VariableCoding’ for details on variable names.