Materials and Methods

Subjects

Humans: Twenty-two normal or corrected-to-normal vision adult human volunteers were recruited for Experiment 1 (seven male, fifteen female, aged 18–28 years), and eight were recruited for Experiment 2 (3 male, 5 female, aged 18–28 years). All participants are graduated students from the Chinese Academy of Sciences (CAS) in Shanghai. The ethical committee of the Institute of Neuroscience, CAS approved the experimental procedure. All subjects were naïve to the purpose of the study and were paid for their participation.

Animals: Four adult male rhesus monkeys (Macaca mulatta, monkey B, H, N, and S) weighing between 6 and 10 kg were used in the behavioral experiments. Two of them (H and N) were conducted electrophysiological recordings. Monkeys only obtained juice during experimental sessions as a reward for required behavioral responses. The Animal Care and Use Committee at the Institute of Neuroscience, CAS, approved animal care, behavioral and electrophysiological recording procedures.

Experimental setup and tasks

Apparatus and tracking system. Participants were required to sit in front of a chest-height table on which a lab-made virtual visual platform was placed (Fig. 1). During the whole experiment, their right arm (and the left arm in the case of monkey N, who was left-handed) was placed on the platform and blocked from sight. The image of the participant's arm reflected in the 45° mirror was captured by the CCD

www.pnas.org/cgi/doi/10.1073/pnas.1902334116

camera (MV-VEM120SC, Microvision Co., China) and projected to the rear screen by a high-resolution projector (BenQ MX602, China). Thus, when the participant looked in the horizontal mirror suspended between the screen and the table, the visual arm image appeared to be their real arm on the table. The lower edge of the screen was aligned to the table edge. The participant's trunk just touched the edge of the table, and their right shoulder was aligned with the midline of the screen. The arm image and visual target were generated and manipulated using the OpenCV graphics libraries in C++ (Visual Studio 2010, Microsoft Co., WA, USA). Hand position was tracked and recorded using CinePlex Behavioral Research Systems (Plexon Inc., TX, USA), sampled at 80 Hz. The tracking color marker was painted onto the subject's first segment of the middle finger, which was not visible after adjusting light exposure settings of the video.

Task procedure. Both human participants and monkeys were asked to make reaching movements to the targets (red dots with a 1 cm diameter, Fig. 1A) from a fixed start position (a blue dot with a 1.5 cm diameter, Fig. 1A), which were both generated by a computer program. Subjects were required to place their hand on the starting position for 1000 ms to initiate a trial, and were instructed not to move. Movement of the arm at any time during the initial holding period automatically ended the trial. After the initiation period, the starting point disappeared and the virtual (visual) arm was rotated (within one video frame, 16.7 ms) in the visual-proprioceptive conflict (VPC) condition (described below), and this mismatch "arm" was maintained for 500 ms as the preparation period. The reaching target was presented as a "go" signal.

Subjects had to reach to the visual target (chosen from T1 to T9 randomly trial by trial, Fig. 1A) within 2500 ms and placed their hand in the target area for 500 ms (monkeys received a reward for doing so). Any arm movement during the target-holding period automatically terminated the trial. The mismatch "arm" was maintained throughout the whole trial along with the arm movement. The target area is defined in details below. The inter-trial interval was 2000 ms, after which the subject was allowed to start the next trial. During the inter-trial-interval, the visual scene was blank.

To make sure that subjects used their proprioceptive sense of their arm, human participants were directly instructed to touch the targets with their veridical (proprioceptive) arm, and monkeys were first trained to perform the task with the vision-proprioception (VP) congruent stimuli to learn the task procedure, i.e., initiation, reaching, and holding the position. Then the proprioception-only (P) condition was followed, whereby the visual signal was blocked and subjects had to touch the target using only proprioceptive information. Once the animals' performance reached a certain level (after the intensive training, usually 1-2 weeks, the standard deviation of the monkeys' proprioception reached to 2° to 3°, centered at each target position), the visual-proprioceptive conflict (VPC) condition was introduced and randomly mixed with the P condition, block by block. In the VPC condition, across trials, the visual arm was randomly presented with a disparity of 0°, ±10°, ±20°, or ±30° (+: clockwise (CW); −: counterclockwise (CCW) direction; ±35° and ±45° were used in the case of monkeys N, H and S) from the subject's proprioceptive

arm, with their shoulder as the center point (Fig. 1C). In addition, ±90° was tested in the monkey N, H and S during some of the electrophysiological recording sessions. The starting point was fixed 25 cm away from the subject's shoulder. Target position was selected randomly trial by trial from one out of nine possible positions located on an arc (for monkey N and H, a ±4° jitter was added to the original positions trial by trial to ensure monkeys did not perform the task by memorizing all the target positions; for human experiments, fewer target positions were used, see the description below) (Fig. 1A). The arc diameter was about 50 cm for human subjects and about 40 cm for monkeys, as their arm lengths are different. The centre point was aligned with the subjects' shoulder, and the spatial visual (from T1 to T9) targets were arranged at 0°, ±10°, ±20°, ±30°, and ±40° from the midline.

Behavioral experiments in human and monkey

Human Experiment 1. Twenty-two participants were recruited for Experiment 1 (data from five participants were excluded due to incomplete tasks), in which subjective illusion statements were assessed using a questionnaire about ownership and agency. Only visual-proprioceptive conflict (VPC) condition was given in this experiment. The combinations of target position and rotation degree were balanced in the trials. Some combinations were excluded because the visual arm could not be presented outside of the screen (e.g., for the rightmost target, it was impossible to rotate the visual virtual arm clockwise any further). Five target positions (T2, T3, T5, T7 and T8,

Fig. 1A) were used in this experiment. Trials were grouped into 10-trial mini-blocks according to the cue disparity (rotation degree). Each mini-block contained only one type of cue disparity, and the mini-block order was randomized within subject. Block order was also counter-balanced across subjects. After each mini-block, participants were asked to complete the questionnaire (Table S1). In this experiment, the wood condition was tested, in which the real-time visual arm image was replaced by a piece of wood (Fig. 5A) (1). For each human, the size and length of the wood were identical to their veridical arm. The questionnaire was also completed after this wood-condition blocks. Thus, each subject completed 90 trials (10 trials $\times$ 9 disparities) for the VPC condition and 90 trials for the wood condition. The test order of arm and wood condition was counter-balanced across subjects.

The questionnaire statements were adopted from Kalckert and Ehrsson's original report (2) and translated into Mandarin Chinese. To simplify and minimize interruption of the experiment, only one statement was chosen for each type of question (Table S1). Similar to the original questionnaire, Q1 served as the ownership questions, and Q2 served as the control questions. The questions were presented on the projection screen one by one, and participants were asked to rate their subjective experience with their non-experimental (left) hand using a 7-point Likert scale. The Likert scale ranged from "-3" (strongly disagree) to "0" (uncertain) to "+3" (strongly agree).

Human Experiment 2. To avoid the motor adaptation within each block (10 trials with identical cue disparity) and to ensure the consistency with the behavioral design

in monkeys (describe below), another eight human participants were recruited for the second behavioral experiment. In this experiment, no questionnaire statements were given. The procedure was the same as for the Experiment 1, except that trials with different cue disparities were randomized (not the block-design). Specifically, there were 158 trials in total, which were separated into four blocks, which included 42 trials (6 trials $\times$7 targets) for the VP block, 42 trials (6 trials $\times$7 targets) for the P block, and 74 trials for the VPC block (37 trials per block). The VP and P experiments were additionally conducted and data were used for the model fitting (see Data analysis below). Seven target positions (T1, T2, T3, T5, T7, T8 and T9, Fig. 1A) were used in this experiment. Note that there were no significant differences in spatial drift between the block-design (Experiment 1) and random-trial experiment (Fig. S9).

Monkey experiment. To ensure that monkeys indeed performed the reaching to target task, in the VPC condition, the reaching target area (with reward) was defined as follows: the radial distance from the hand to the center of the target was less than 5 cm to ensure that the monkey did reach out to the target; with the target as the center, the azimuth range was set from [-7- rotation degree]° to +7° when the rotation degree was positive (clockwise), and from -7 to [+7°- rotation degree]° when the rotation degree was negative (counter-clockwise) (green zone in Fig. 1C). Because it was reasonable to assume that the monkeys' estimation of the arm would be located somewhere between the positions of the pure visual arm (completely bias to the visual information) and the pure proprioceptive arm (completely bias to the

proprioceptive arm). The additional azimuth 7° was introduced to cover the variance of vision and proprioception and the behavioral bias of each animal (7° is three times of standard deviations of proprioception of monkeys after intensive training). Animals were always rewarded no matter where their arms were located within the target area. Therefore, no feedback was given to subjects about their response (integration). Trials in which the animals' arm was outside of the target area were considered as errors, as animals might not have been engaged in the task. With the large reward area, the performance of monkeys reached 95% on average. Therefore, the pattern of the spatial drift was not due to the training protocol. All the nine target positions were used for the monkey H, N and S, and seven target positions (except T4 and T6, Fig. 1A) were used for the monkey B. Only correct trials were included in the analysis.

In one training or recording session (one experimental day), usually one or two P blocks were given first to ensure monkeys to perform the task with their proprioceptive arm, and then in the following blocks, VP, P, and VPC conditions were randomly mixed. The wood condition was randomly mixed with other arm conditions (VP, P and VPC) on monkey H and N in the recording sessions. In the wood experiment, the real-time visual arm image was replaced by a piece of wood. For each monkey, the size and length of the wood were identical to their veridical arm. The wood image (e.g. texture and color) changed every week during training and recording sessions.

Multisensory integration Model

As each target location displayed essentially identical pattern of drift, we then pooled behavioral data across all the target locations. All the statistics and model fittings were performed using Matlab (Mathworks, MA, USA).

Optimal integration model (forced-fusion model). The classical cue integration model assumes that the central nervous system estimates the source location (s) by combining information from multiple modalities (3). In the present task, the arm location can be estimated from vision ($\hat{s}_V$) and proprioception ($\hat{s}_P$), and each estimate is corrupted by neuronal (processing) noise σ_V and σ_P. If the noises are independent, then the maximum-likelihood estimation of s will be as follows:

$$\hat{s} = \frac{\frac{\hat{s}_V}{\sigma_{V^2}}}{\frac{1}{\sigma_V^2}+\frac{1}{\sigma_P^2}} + \frac{\frac{\hat{s}_P}{\sigma_{P^2}}}{\frac{1}{\sigma_V^2}+\frac{1}{\sigma_P^2}}$$

The variance of combined information will also be lower than that of each single estimator, as described by the following equation:

$$\frac{1}{\sigma_{VP}^2} = \frac{1}{\sigma_V^2}+\frac{1}{\sigma_P^2}$$

Bayesian Causal Inference model. We adopted the Bayesian Causal Inference (BCI) model described in a visual-auditory integration study (4). Basically, in the present study, the BCI framework included three models, as follows: (1) the full-segregation model, which assumes that visual and proprioceptive estimates of the arm's location are processed independently; (2) the forced-fusion model, which assumes that visual and proprioceptive estimates are integrated optimally, that is, weighted by their reliabilities regardless of the hidden causal structure; and (3) the Bayesian Causal

Inference model, which computes a final proprioceptive estimate by averaging the spatial estimates under forced-fusion and full-segregation assumptions weighted by the posterior probabilities of each causal structure. Thus, the model assumes that both visual and proprioceptive location information (s_V and s_P) were represented as x_V and x_P in the neural system, which were drawn from the normal distribution with sensory noise ($N(s_V, \sigma_V)$, $N(s_P, \sigma_P)$). The source inference is determined by the joint distribution of two sensory signals and the prior probability of a common source (*Prior*). Thus, according to the Bayesian rule, the one source (posterior) probability (P_{com}) is obtained as follows:

$$p(C = 1|x_V, x_P) = \frac{p(x_V, x_P|C = 1)Prior}{p(x_V, x_P|C = 1)Prior + p(x_V, x_P|C = 2)(1 - Prior)}$$

and the two sources probability is $p(C = 2|x_V, x_P) = 1 - p(C = 1|x_V, x_P)$. If the system completely "believes" there is only one common source for the two sensory signals (forced-fusion situation), then the estimate of arm position is determined by the optimal integration rule, as follows:

$$\hat{S}_{VP,C=1} = \frac{\frac{x_V}{\sigma_V^2} + \frac{x_P}{\sigma_P^2} + \frac{\mu_{P_r}}{\sigma_{P_r}^2}}{\frac{1}{\sigma_V^2} + \frac{1}{\sigma_P^2} + \frac{1}{\sigma_{P_r}^2}}$$

where the $N(\mu_{P_r}, \sigma_{P_r})$ stands for a prior distribution of arm location. In this experiment, we set the μ_{P_r} to be equal to the half of disparity, and set $\sigma_{P_r} = 10000$ to approximate a uniform distribution. If the system completely "believes" the two sensory signals are from different sources (full-segregation situation), the proprioceptive arm position is estimated independently from the visual information, as follows:

$$\hat{S}_{P,C=2} = \frac{\frac{x_P}{\sigma_P^2} + \frac{\mu_{P_I}}{\sigma_{P_I}^2}}{\frac{1}{\sigma_P^2} + \frac{1}{\sigma_{P_I}^2}}$$

The final output of the arm position was obtained by the weighted sum of the forced-fusion result and the proprioceptive unisensory estimate, as follows:

$$\hat{S}_P = p(C=1|x_P, x_V)\hat{S}_{VP,C=1} + \big(1 - p(C=1|x_P, x_V)\big)\hat{S}_{P,C=2}$$

In the model simulation, we fixed the proprioceptive arm position at the end of trial as zero ($s_P = 0$), so that the visual arm position would be the visual-proprioceptive disparity ($s_V = Disparity$). In the present task, both human subjects and monkeys were required to report their proprioceptive arm position, thus only the proprioceptive estimate was simulated.

Model fitting analysis and comparison. For both humans and monkeys, the causal inference model was fitted to the data in the VPC condition. We set the visual and proprioceptive standard deviation (σ_V and σ_P) and the prior probability of common cause (*Prior*) as free parameters to be optimized. Here, 5000 trials per disparity were simulated to form a distribution for each optimization step, and the likelihood of the observed data given the model (i.e., the product of probability of each observed data point given the simulated distribution from the model) was calculated for each disparity. Then, we minimized the sum of negative log-likelihood of all observations, and the parameters were optimized by genetic algorithm. The procedure was the same as the optimal integration model, except that there were no causal structures and only two free parameters (visual deviation σ_V and proprioceptive deviation σ_P) need to be optimized. For the model fitting of the wood condition, we simulated the

distributions in both wood and arm conditions with *Priors* of each condition (*Prior-wood* and *Prior-arm*), then minimized the sum of negative log-likelihood of all observations in both conditions. All simulation and optimization processes were performed in Matlab (Mathworks, MA, USA). Only correct trials were included.

The model goodness of fit was evaluated using the coefficient of determination R^2, as follows:

$$R^2 = 1 - \exp\left[-\frac{2}{n}\{l(\hat{\beta}) - l(0)\}\right]$$

Where $l(\hat{\beta})$ and $l(0)$ denote the log likelihoods of the fitted and the null model, respectively, and n is the number of observations. For the null model, we assumed that subjects would report the perceived arm position randomly over the disparity range where form the leftmost to the rightmost. Thus, a uniform distribution over this span was predicted. We used the Bayesian information criterion (BIC) as an approximation to the model evidence. The relative Bayesian information criterion at the group level ($relBIC_{group}$) was the summation of all subjects' BIC relative to the causal inference model. Finally, we identified the better model at the group level by the exceedance probability based on all subjects' (sessions for monkeys) BIC (5).

Correlation between one source inference and body ownership. To explore whether the one source probability predicted by the causal inference model reflected the subjective rating of body illusion, we performed a correlation analysis between posterior probabilities of common source (P_{com}) and questionnaire scores in the human Experiment 1. Within subjects, a Pearson correlation was performed between P_{com} and ownership score of different disparities (Fig.2I). Across subjects, we

normalized the rating score to 0 to 1, then fitted both of them for each subject with a symmetric bell-shaped function, as follows:

$$S = \left(\frac{1+\cos\theta}{2}\right)^n$$

The half full-width-at-half-maximum of these fitted curves were used as the indices to determine the relationship between the subjective rating of the visual illusion, and the model predicted P_{com} (Fig.2J).

Electrophysiological Recordings

Implantation of recording chamber. After the two monkeys (H and N) were able to perform VP, P and VPC task at better than 95% correct rate, surgical operations were performed. Under strictly sterile conditions and general anesthesia with isoflurance, a cylindrical recording chamber (Crist instrument Co., Inc., Maryland, USA) of 22 mm diameter was implanted on the premotor cortex (Fig. S4). The location of the recording chamber on each animal was determined by the individual MRI atlas (3T, Institute of Neuroscience, Chinese Academy of Sciences) (6, 7).

Single-unit recordings. For each recording session, 2 or 3 glass-coated tungsten electrodes (1~2MΩ, Alpha Omega, Israel) were introduced into the premotor area through a multi-electrode driver (NAN electrode system, Plexon Inc., USA). The single-unit activity recorded extracellularly was collected by the Plexon system (Plexon Inc., Dallas, TX). For any given unit in the three conditions (VP, P and VPC), the eye movement throughout a trial was monitored and recorded for an off-line analysis to exclude the effect of eye movement on the unit firing (EyeLink 1000, SR

Research, ON, Canada). Trials in which the animal closed its eyes during visual cue period were excluded from analysis. Single units recorded through at least 8 completed trials with correct responses for each experimental condition (task conditions x spatial targets x sensory disparities, yield around 900 trials in total for each unit) were saved in the database for further analyses.

Data analysis

The following events in the VPC task were used for the data analysis: 1) onset of the visual arm rotation, 2) onset of the visual-spatial target, which is also the 'go' signal for the reaching movement, and the preparation period was defined as 500 ms before the event, 3) onset of the target holding period and 4) offset of the target holding period, which is also the onset of rewarding period. Baseline unit firing was calculated during the period of 500 ms prior to the visual arm rotation--event 1. Rotated visual arm (cue disparity) was presented for 500 ms after the event 1 with the real arm holding steadily at the start point position (the preparation period). The target holding period was lasted for 500 ms after the event 3 when the hand tracking system detected the velocity of arm movement dropped below 10 millimeter per second (8) and the position of hand was located in the target area (the green zone in Fig.1C). After off-line spike sorting by the Offlinesorter (Plexon Inc., Dallas, TX), files were exported to NeuroExplorer software (Plexon Inc., Dallas, TX). After data files in mat format generated by NeuroExplorer, the analysis was conducted in MATLAB (Mathworks, Natick, MA, USA).

Pre-processing. Timestamps of spiking events were re-sampled at 1kHz and converted into binary spike trains for single trials. To estimate continuous time-dependent firing rates, spike trains were convolved with a symmetric Hann kernel (MATLAB, Mathworks):

$$w(n) = A\left(1-\cos\left(2\pi\frac{n}{N}\right)\right),\ 0<=n<=N\ (N = L-1).$$

where A is a normalization factor ensuring the sum of the kernel values equals to 1. Window width L was set to 300ms. Single neurons were included in the analysis only if they had been recorded by a full set of conditions (VP, P and VPC condition with 9 disparities: 0°, ±10°, ±20°, ±35° and ±45°).

PSTH (peri-stimulus time histogram). For the PSTH in Figure 3, the smoothed firing rates were averaged across trials for a given condition aligned to the onset of periods of interest: 1) the onset of visual-arm rotation (event 1), 0.5s prior to target onset (for the preparation period); 2) the onset of target holding (event 3), 0.5s prior to trial end (for the target-holding period).

Probabilistic decoding analysis. Neurons showing the selectivity to either condition (VP and P) or spatial target were included in this analysis. For time periods of target-holding, firing rates were first averaged over the time period (500 ms) on individual trials. The firing rates in the trials collected from the VP and P conditions were then submitted to a two-way ANOVA (analysis of variance) (with two factors: *Condition (two levels: VP and P)* × *Spatial Target (nine levels: T1-T9)*). Neurons selective to the presented conditions (main effect of *Condition, p* < 0.05) or spatial target (main effect of *Spatial Target, p* < 0.05) were chosen for further probabilistic

decoding analysis.

For each neuron, neural responses of single trials from both VP and P conditions were first fitted to von Mises function (9), as their estimated spatial tuning curves for VP or P arm location respectively.

$$d = b + \alpha \times \cos(\theta - \mu).$$

where b, α, μ are parameters to fit, θ is the absolute arm location in the 2D platform (in radian, from -π/2 to π/2 as the position goes from the leftmost to the rightmost horizontal, with the midline front being zero, Fig.1A) varied trial by trial, and d is the corresponding firing rate. We assumed neurons follow Poisson-like statistics (10, 11) , therefore, a probability of observing current spike counts in a trial given the estimated firing rate from either VP or P condition can be derived from:

$$P(r|\theta) = \frac{e^{-f(\theta)} f(\theta)^r}{r!}$$

where r is the response of a single neuron on a single trial where the arm is at a specific location θ. $f(\theta)$ is the spatial tuning curve of the neuron, which maps every possible arm location ϑ to an estimated firing rate λ. Then for the same neuron in the VPC condition, we mapped, for a given spatial target (estimated arm location) in a single trial, the proprioceptive arm (θ_P) and the illusive visual arm (θ_{VP}) position (from behavior) onto the VP and P tuning curves respectively, to get the probabilities (P_{vp} and P_p) (Fig. 3D). The relative weight of integration (*VP-weight*) was further normalized as being divided by the sum of the two:

$$Pr_{VP} = P_{VP}/(P_{VP} + P_{\mathrm{p}})$$

Population Neurons. For the Figure 3G, based on the disparity levels and drifts, for

each neuron, we first divided all the trials in the VPC condition into 29 clusters. Continuous drift values were discretized into nine clusters: < -35°, [-35° -25°], [-25° -15°], [-15° -6°], [-6° +6°], [+6° +15°], [+15° +25°], [+25° +35°], > +35°. Note that ±6° covers approximately 99% of drift distribution under the VP and P condition. Thus, for the disparity 0°, there was only one cluster [-6° +6°]. The larger the disparity, the more clusters would be assigned, because the distribution of drift becomes wider (higher variance). For example, for the disparity ±45°, there were five clusters of drifts. The overall pattern of the assignment is depicted in Fig. S5B. To ensure each neuron contributed equally to each cluster, the VP-weights from single units were bootstrapped 50 times for each cluster. Then, each trial in the clusters at the population level was the average of 303 trials from 303 neurons, where one trial was randomly chosen from one neuron. For the demonstration and statistics, 50 trials were generated according to the above procedure as the population activity in each cluster for the further regression analyses. Note that we also pooled all the trials from the 303 neurons (without the bootstrap) and assigned them into the 29 clusters. A similar casual inference pattern of population activity was observed, as shown in Figure 3G. Note that the causal inference pattern was defined by using the linear regression between the VP-weight and P_{com} of the 29 clusters, for both individual and population neural activities.

Decoding analysis and statistical significance for decoding accuracies. Decoding was done with support vector machine classifiers implemented using the scikit-learn toolbox (12). We trained the classifier to identify P_{com} (8 bins from 1 to 8) based on

neural activity (VP-weight) from individual animals, pooled across recording days. Based on the drift from each monkey, a P_{com} matrix with 29 clusters was generated from the simulation (100 artificial trials for each group) of the BCI model (Fig. S5B). Then the 29 clusters were grouped into 8 classes (k-mean) and were labelled from low P_{com} (1) to high P_{com} (8) to represent their relative levels of illusion. The cluster number was determined by the *Silhouette* test (13). Population neural activity was defined by the VP-weights of cells of the corresponding 29 clusters in Figure 3G. Threefold cross-validation was then implemented by splitting the neural data into 3 subsamples, each randomly drawn from entire dataset. Decoders were then trained on 2 of the subsamples and tested on the remaining subsample, with this process repeated using all 3 subsamples used as test data once. Decoding accuracy was taken as the averaged accuracy across the 3 trained decoders. This cross-validation process was repeated 1000 times and the overall decoding accuracy taken as the mean across the 1000 repetitions. The decoding analysis was conducted with the population activity during the target-holding and baseline periods, individually. Statistical significance for decoding accuracy was determined by comparing mean decoding accuracy from the original data from each interested period (target-holding and baseline) against the accuracy from the shuffled data from the corresponding period.

Supplementary references:

1. Manos Tsakiris, Carpenter L, James D, & Fotopoulou A (2010) Hands only illusion: multisensory integration elicits sense of ownership for body parts but not for non-corporeal objects.
2. Kalckert A & Ehrsson HH (2014) The spatial distance rule in the moving and classical rubber hand illusions. *Conscious Cogn* 30:118-132.

3. Ernst MO & Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*.
4. Kording KP*, et al.* (2007) Causal inference in multisensory perception. *PloS one* 2(9):e943.
5. Maloney LT, Wozny DR, Beierholm UR, & Shams L (2010) Probability Matching as a Computational Strategy Used in Perception. *PLoS Computational Biology* 6(8):e1000871.
6. Graziano M (2009) The Intelligent Movement Machine: An Ethological Perspective on the Primate Motor System.
7. Graziano MS (1999) Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position. *Proceedings of the National Academy of Sciences of the United States of America* 96(18):10418-10421.
8. McGuire LM & Sabes PN (2011) Heterogeneous representations in the superior parietal lobule are common across reaches to visual and proprioceptive targets. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31(18):6661-6673.
9. Amirikian B & Georgopoulos AP (2000) Directional tuning profiles of motor cortical cells. *Neurosci Res* 36(1):73-79.
10. Fetsch CR, Pouget A, DeAngelis GC, & Angelaki DE (2011) Neural correlates of reliability-based cue weighting during multisensory integration. *Nature neuroscience* 15(1):146-154.
11. Jazayeri M & Movshon JA (2006) Optimal representation of sensory information by neural populations. *Nat Neurosci* 9(5):690-696.
12. Pedregosa F*, et al.* (2011) Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 12:2825-2830.
13. Rudolph R (1959) The Silhouette Test. *Nordisk Psykologi* 11(1):25-44.

Supplementary Figures and legends

Supplementary Figure. 1

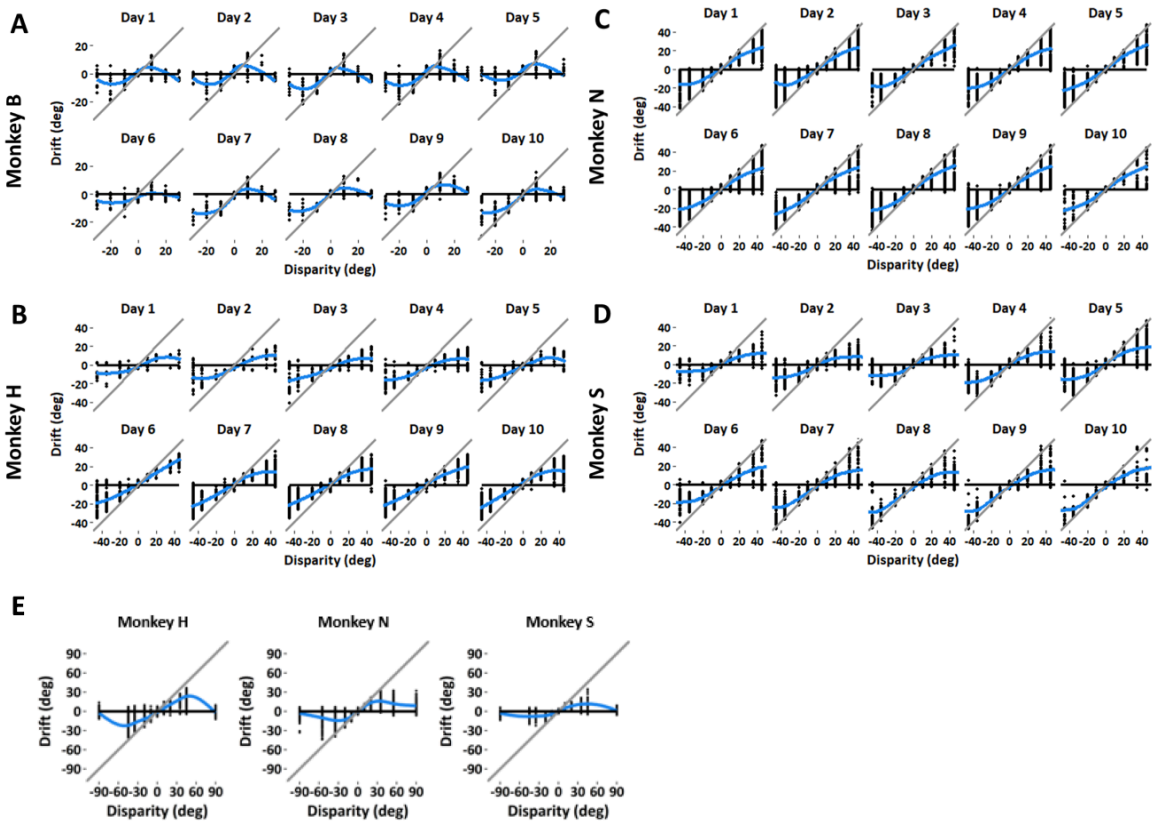


Figure S1. (A-D) Behavioral performance of ten consecutive sessions (days) from the four monkeys. The pattern of drift was consistent across all sessions for each monkey. Each black dot indicates a single trial and the blue line indicates the smoothed mean drift across trials. Monkey B was only involved in the behavioral experiments and in total completed 15 days (sessions) of the VPC task. For the monkey H, N and S, we only showed 10 consecutive sessions as examples to demonstrate the consistency of their behavior. **(E)** For the monkey H, N and S, the 90° disparity was also tested. One session data was demonstrated on each monkey.

Supplementary Figure. 2

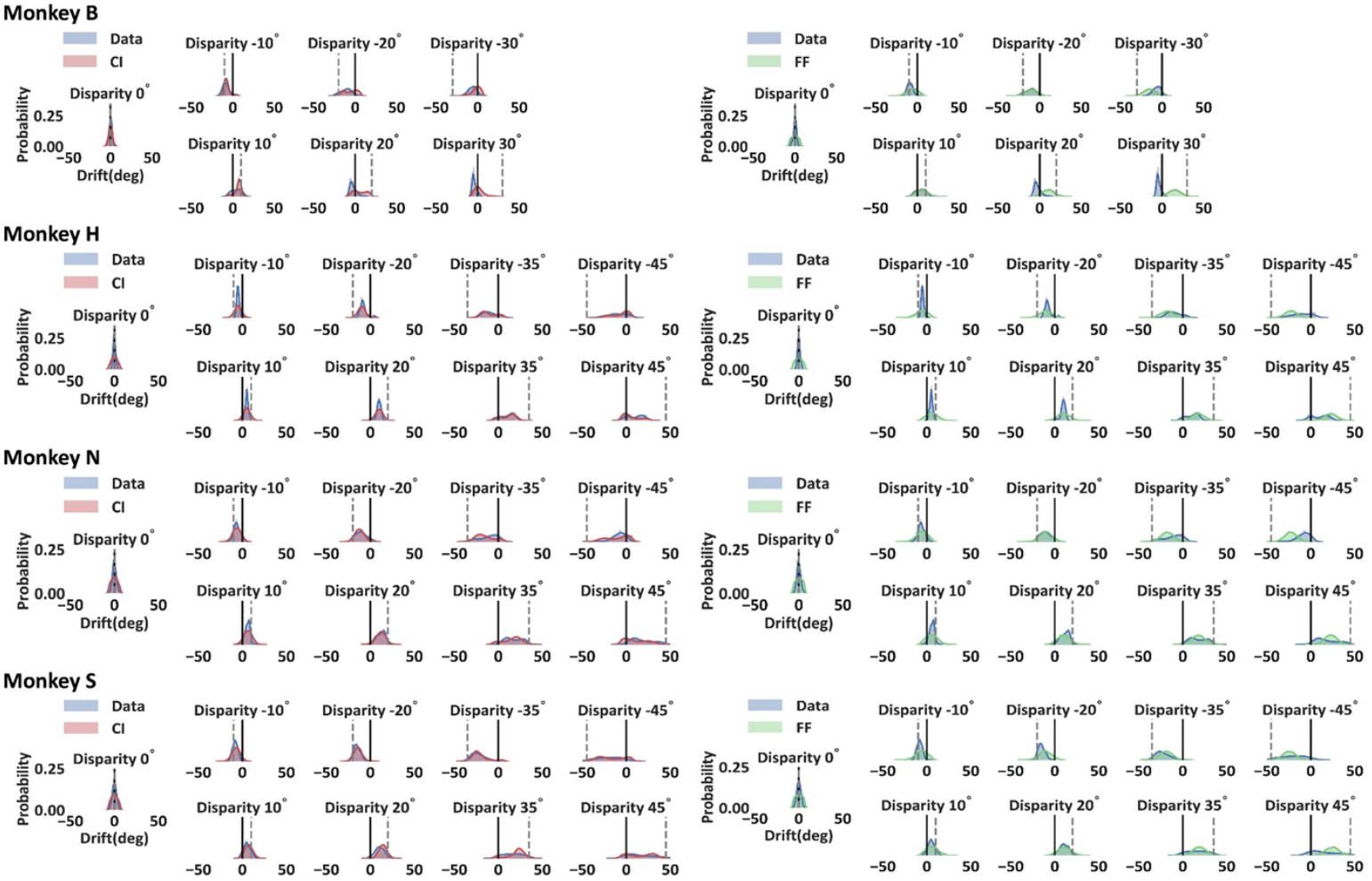


Figure S2. Histograms of behavior and model simulated results in individual monkeys. The predictions of the Causal Inference model (left panel, red histograms) significantly better characterized monkeys' behavior data (blue histograms) than the predictions of forced-fusion (optimal integration) model (right panel, green histograms). The details of model comparisons were shown in Table S2. Data distributions were plotted using all trials from multiple sessions from four monkeys. Model results were generated from 5000 trials per disparity for the individual session and were plotted the distribution using all these simulated trials. Vertical black solid and dotted gray lines indicate the normalized pure proprioceptive bias and pure visual bias respectively.

Supplementary Figure. 3

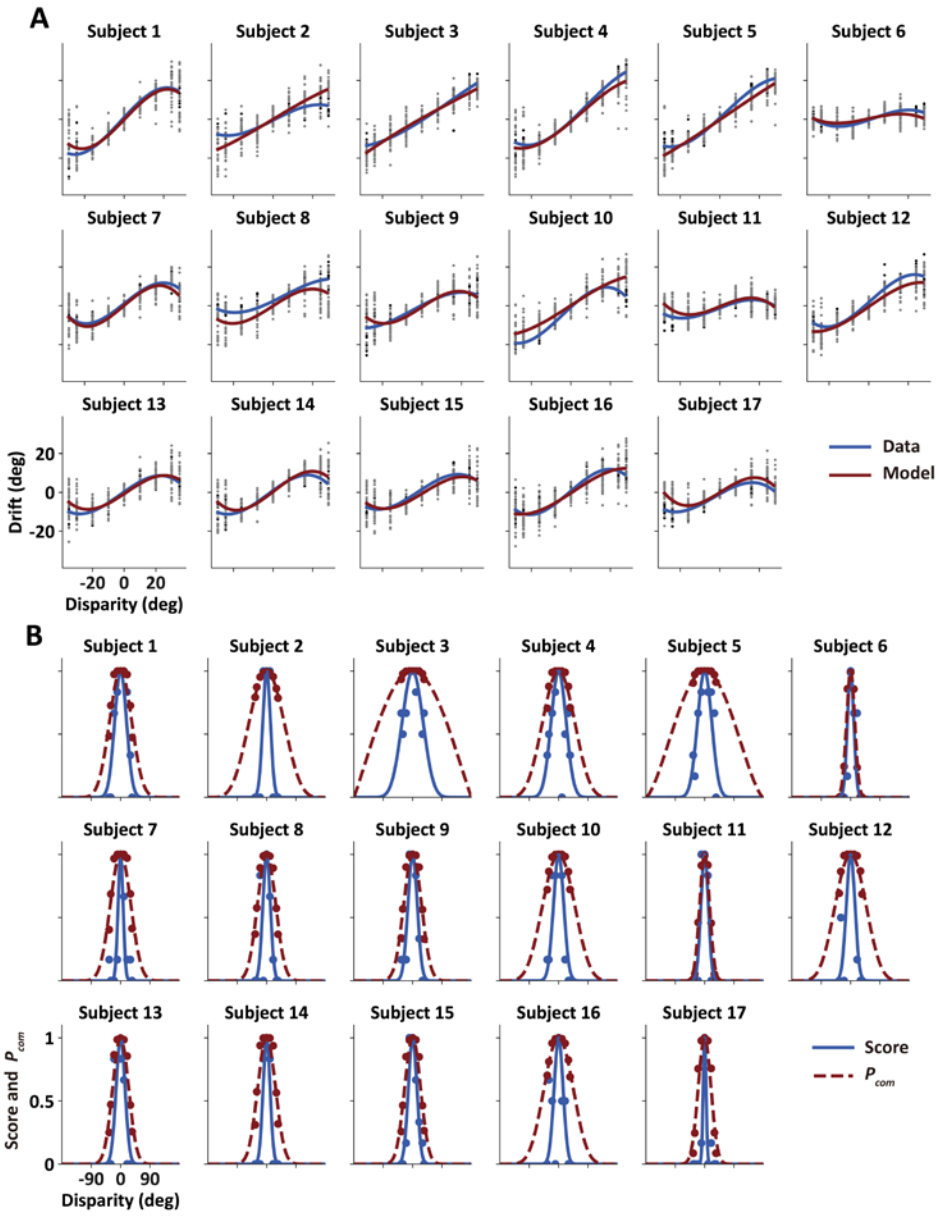


Figure S3. (A) The causal inference (model average) model simulation results (red line) faithfully characterized the profile of the mean drift (blue line) for each human participant (human Experiment 1). Black and gray dots represent behavioral and model simulated trials. **(B)**. The normalized questionnaire rating scores and the posterior probability of common source (P_{com}) for each human participant were fitted by a symmetric bell-shaped function. Blue dots and solid lines represent the rating scores at each level of disparity and fitted results. Red dots and dashed line indicate the P_{com}. The half-full-widths at half-maximum of these fitted curves were used as the indices to determine the relationship between the subjective rating of the visual illusion, and the model predicted P_{com} in Figure 2J.

Supplementary Figure. 4

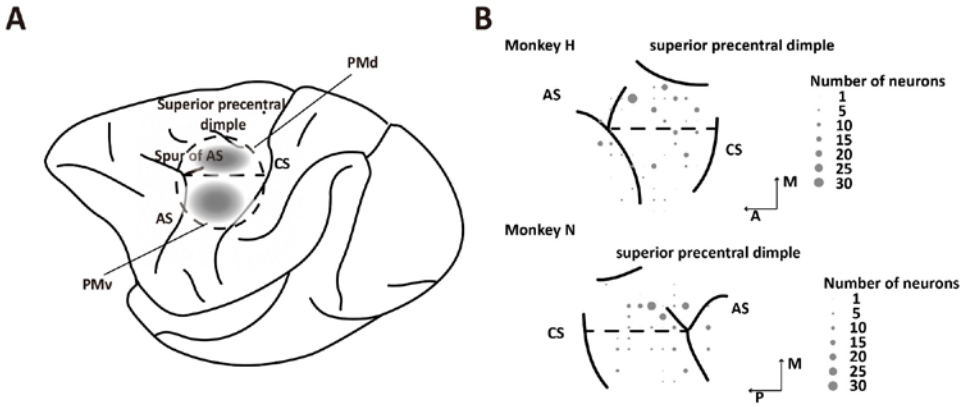


Figure S4. Anatomical locations of recording sites. AS: arcuate sulcus; C(e)S: central sulcus. The recording sites cover the arm areas in the F2, F4 and part of F5. The recording chamber was implanted in the left hemisphere of monkey H and right hemisphere of monkey N. The database included 275 neurons from dorsal premotor cortex (32 recording sites), 164 neurons from ventral premotor cortex (31 recording sites), and 87 neurons close to the boundary of AS or CS from 13 sites were unclassified. The most (249 out of 303, 82.2%) of selective neurons in the causal inference analysis were from either dorsal (164) or ventral (87) promotor cortex.

Supplementary Figure. 5

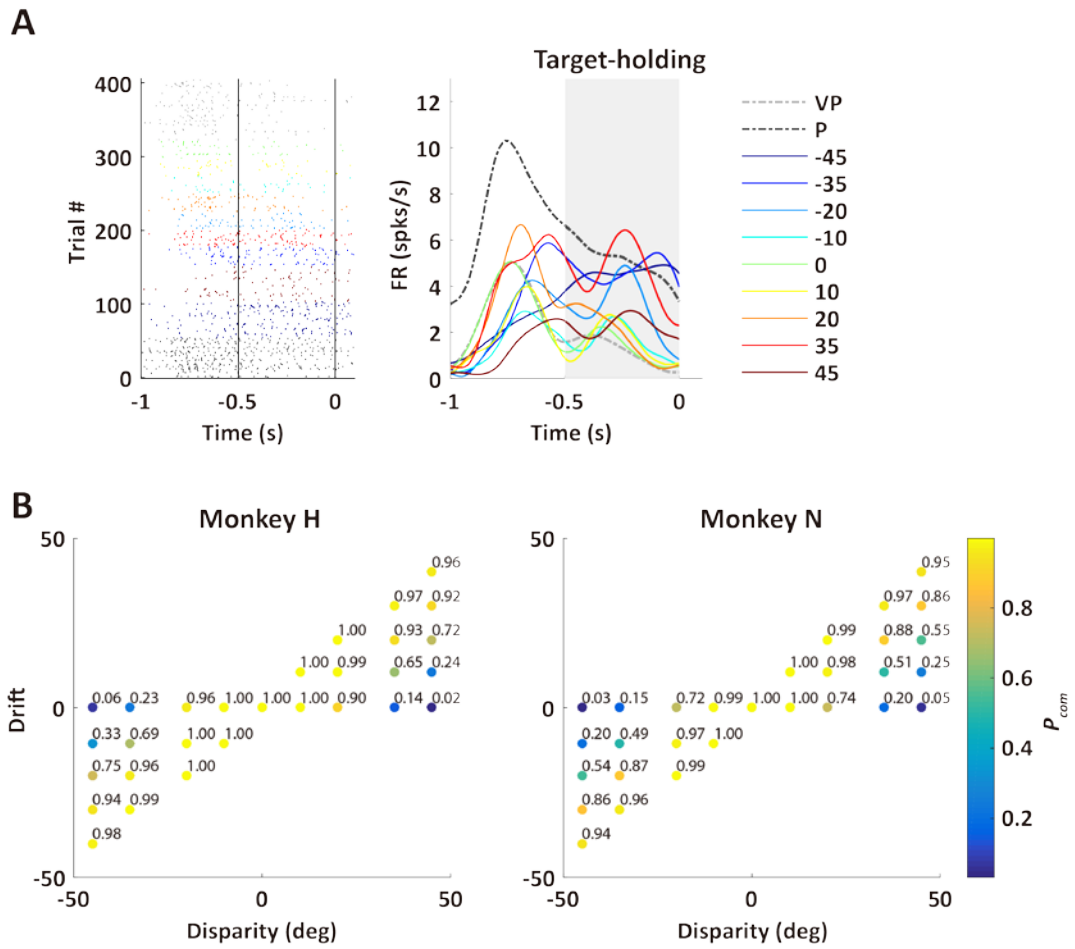


Figure S5. (**A**) Rasters and histograms of activity of the same example neuron in Fig. 3F, showing the P (segregation)-preference during the target-holding period. (**B**) The pattern of P_{com} from modelling of the drift of the two monkeys. Behavior data included all trials from the recording sessions (monkey H: n = 68 sessions; monkey N: n= 47 sessions). The 29 clusters were defined based on the distribution of drift (see Methods). Each value indicates the P_{com} fitted by the BCI model.

Supplementary Figure. 6

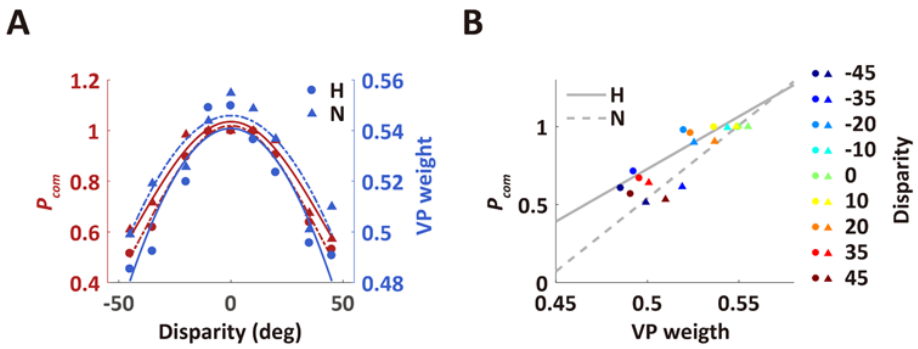


Figure S6. The changes of VP weights across disparities closely fitted with the profile of the P_{com} from behavior in the BCI model. (**A**) Averaged P_{com} (Fig. S5B) and VP-weight across disparities of each monkey (H and N). (**B**) Significant correlation between the behavior and neural activity of two monkeys (linear regression, monkey H: $R = 0.93$, $p < 0.001$; monkey N: $R = 0.93$, $p < 0.001$).

Supplementary Figure. 7

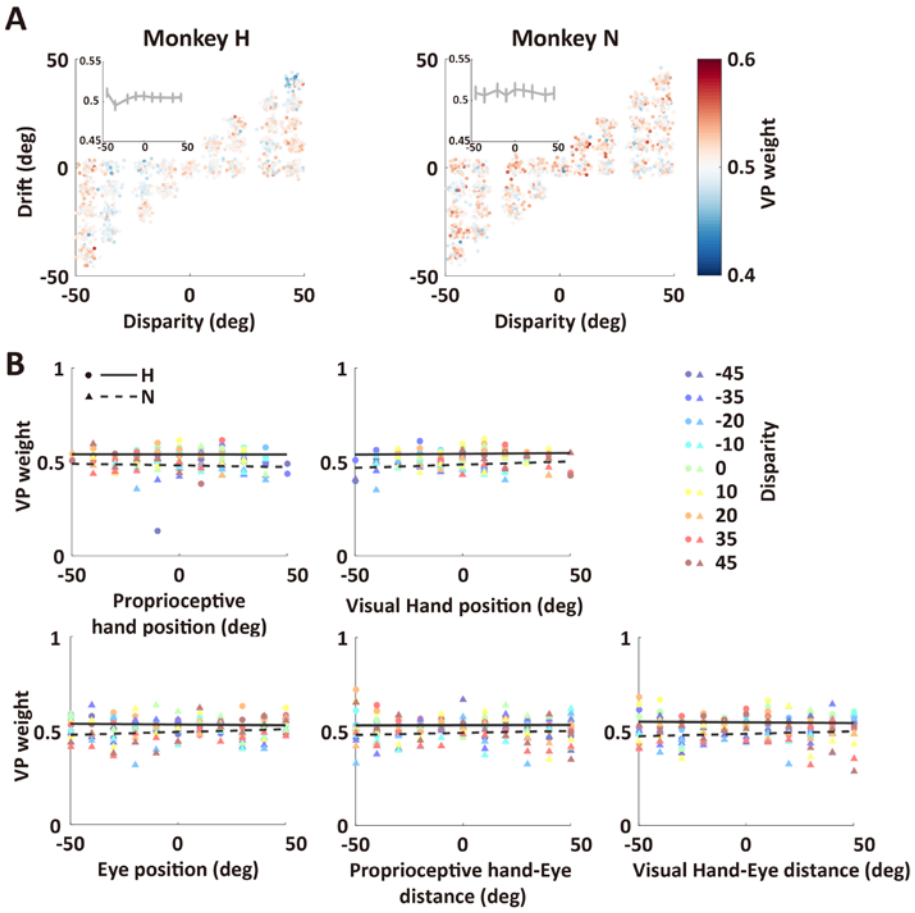


Figure S7. (**A**) VP weight pattern during the baseline period (based on the same population of neurons used in the target-holding period shown in Fig. 3G) showed no significant correlation with the disparity (Linear regression of mean VP weight against P_{com} under nine disparity levels (monkey H: R = 0.05, p = 0.9; N: R = 0.41, p = 0.27)). (**B**) The proprioceptive arm position, visual arm position, eye fixation position, the distance between eye fixation and proprioceptive arm position, and the distance between eye fixation and visual arm position did not show significant correlation with VP weight of premotor population neurons (linear regression, all ps > 0.1). All these behavioral data were binned into 10 bins in each disparity (arranged from -50° to +50° to cover all trials). In each bin, trials of each neuron were bootstrapped 50 times and were averaged to estimate the distance at population level. The eye-tracking data was imported into Matlab using EDF Converter (SR Research). The fixations and saccades in eye movements were separated with the default algorithm of the software with the velocity (30°/sec), acceleration (8000°/sec^2) and motion thresholds (0.1°) respectively. The fixation positions were averaged during the target-holding period of each trial.

Supplementary Figure. 8

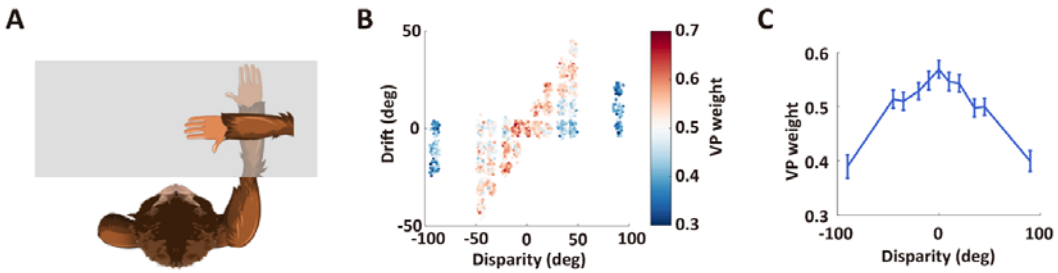


Figure S8. (**A**) Overview of the task-performing platform showing the 90° disparity. (**B**) The VP-weight pattern of the population neural activity from two monkeys (n = 66). (**C**) The VP-weight under the (±) 90 disparity was the lowest among all the disparities (VP-weight in ±90° vs. VP-weight in ±45°, paired *t*-test, $t_{65} = 4.73$, $p < 10^{-4}$).

Supplementary Figure. 9

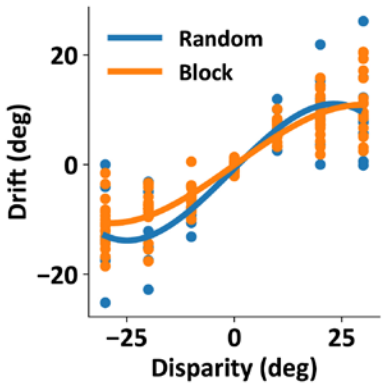


Figure S9. Spatial drift in random design (human Experiment 2, n=8) and block design (human Experiment 1, n=17) human experiments. There was no significant difference in the drift between the two experiments (two-way *ANOVA*, *Design* × *Disparity*, the main effect of *Design*, $F(1,25) = 0.36$, $p = 0.556$). Note that Data from both experiments were submitted to the models respectively (Table S2).

Supplementary Figure. 10

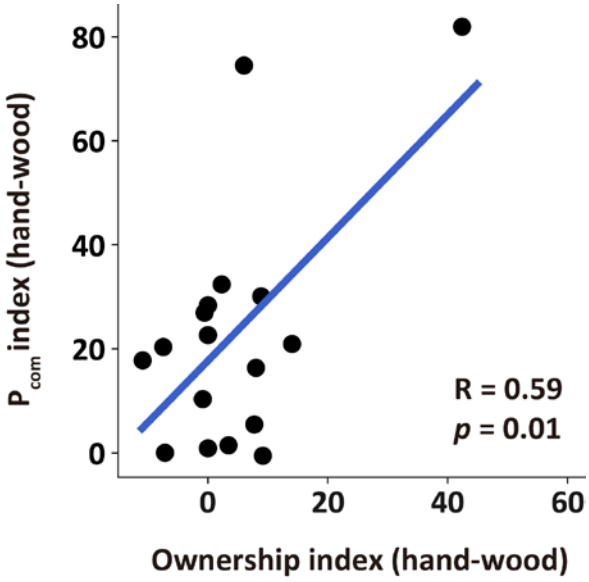


Figure S10. The correlation of the difference in ownership index and P_{com} between arm and wood conditions (Pearson correlation, R = 0.59, p = 0.01).

Supplementary Figure. 11

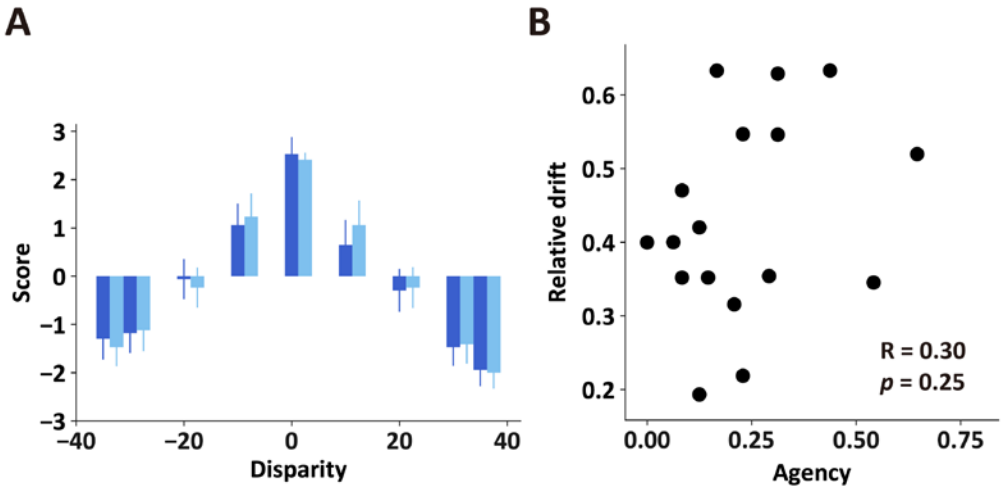


Figure S11. Agency rating results in arm (blue bars) and wood (light blue bars) conditions. (**A**) A two-way ANOVA (*Condition* X *Disparity*) analysis of agency ratings showed no main effect of Condition ($F(1,16) = 0.01$, $p=0.92$), but a significant main effect of *Disparity* ($F(8, 16) = 27.5$, $p<0.01$) was observed. (**B**) There was no significant correlation between the agency ratings and relative drift (Pearson correlation, $R=0.30$, $p = 0.25$).

Table S1. Questionnaire Statements

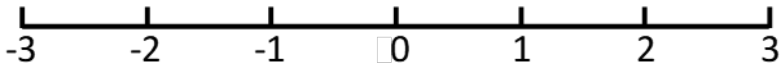
Ownership

Q1. I felt as if the hand/wood in the video were my own hand.

Ownership control

Q2. It felt as if I had no longer a right hand, as if my right hand had disappeared.

Ratings: 7-point Likert scale



Strongly disagree Uncertain Strongly agree

Table S2. Model parameters and fitting evaluations of two models for humans and monkeys.

Experiment	Causal Inference (model averaging)						Forced fusion				
	$relBIC_{Group}$	EP	R^2	σ_P	σ_V	*Prior*	$relBIC_{Group}$	EP	R^2	σ_P	σ_V
Human Experiment 1	0	>0.999	93.9±0.52	5.56±1.02	5.23±1.08	0.74±3.20	2837.7	7.63×10^{-6}	78.4±4.17	9.60±0.77	9.54±0.86
Human Experiment 2	0	0.95	92.8±1.00	7.48±3.36	4.01±0.97	0.44±0.34	367.3	0.046	88.1±2.63	9.24±2.66	6.95±2.33
Monkey H	0	>0.999	94.3±3.11	7.75±1.34	6.05±0.97	0.99±0.02	22703.9	3.62×10^{-13}	90.8±5.59	9.68±1.75	8.80±1.63
Monkey N	0	>0.999	90.6±0.23	8.99±0.83	4.99±0.71	0.79±0.20	15592.5	1.82×10^{-12}	82.5±0.84	10.81±0.61	9.90±0.09
Monkey B	0	>0.999	88.7±0.95	5.18±0.88	2.69±0.66	0.48±0.27	3479.3	1.53×10^{-5}	70.1±2.70	9.61±0.54	9.55±0.51
Monkey S	0	>0.999	90.5±1.12	9.51±2.59	6.81±1.35	0.73±0.22	2.714	1.22×10^{-4}	82.3±1.44	11.01±0.88	9.89±0.06

The model parameters and R^2 were averaged across all subjects (or across days for monkeys); data are presented as the mean ± the standard error of the mean. The $relBIC_{group}$ was the summation of all subjects' BIC (all days' BIC for monkeys).

Abbreviations: σ_P, standard deviation of the proprioception likelihood; σ_V, standard deviation of the vision likelihood; *Prior*, prior probability of common source; $relBIC_{group}$, Bayesian information criterion at the group level; EP, exceedance probability; R^2, coefficient of determination.