



Registered Report

The self and the Bayesian brain: Testing probabilistic models of body ownership through a self-localization task



Tommaso Bertoni ^{a,1}, Giulio Mastria ^{a,1}, Nikita Akulenko ^a, Henri Perrin ^b, Boris Zbinden ^a, Michela Bassolino ^c and Andrea Serino ^{a,*}

^a MySpace Lab, Department of Clinical Neurosciences, University Hospital of Lausanne, University of Lausanne, Lausanne, Switzerland

^b School of Medicine, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland

^c School of Health Sciences, HES-SO Valais-Wallis, Sion, Switzerland

ARTICLE INFO

Article history:

Protocol received 18 June 2021

Protocol approved 16 December 2021

Received 7 November 2022

Reviewed 18 January 2023

Revised 29 March 2023

Accepted 19 June 2023

Action editor Andrew Wilson

Published online 25 July 2023

Keywords:

Body ownership

Multisensory integration

Bayesian causal inference

Self

ABSTRACT

Simple multisensory manipulations can induce the illusory misattribution of external objects to one's own body, allowing to experimentally investigate body ownership. In this context, body ownership has been conceptualized as the result of the online Bayesian optimal estimation of the probability that one object belongs to the body from the congruence of multisensory inputs. This idea has been highly influential, as it provided a quantitative basis to bottom-up accounts of self-consciousness. However, empirical evidence fully supporting this view is scarce, as the optimality of the putative inference process has not been assessed rigorously. This pre-registered study aimed at filling this gap by testing a Bayesian model of hand ownership based on spatial and temporal visuo-proprioceptive congruences. Model predictions were compared to data from a virtual-reality reaching task, whereby reaching errors induced by a spatio-temporally mismatching virtual hand have been used as an implicit proxy of hand ownership. To rigorously test optimality, we compared the Bayesian model versus alternative non-Bayesian models of multisensory integration, and independently assess unisensory components and compare them to model estimates. We found that individually measured values of proprioceptive precision correlated with those fitted from our reaching task, providing compelling evidence that the underlying visuo-proprioceptive integration process approximates Bayesian optimality. Furthermore, reaching errors correlated with explicit ownership ratings at the single individual and trial level. Taken together, these results provide novel evidence that

* Corresponding author. MySpace Lab, Department of Clinical Neurosciences, University Hospital of Lausanne, University of Lausanne, Lausanne, Switzerland.

E-mail address: andrea.serino@unil.ch (A. Serino).

¹ Authors equally contributed.

<https://doi.org/10.1016/j.cortex.2023.06.019>

0010-9452/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

body ownership, a key component of self-consciousness, can be truly described as the bottom-up, behaviourally optimal processing of multisensory inputs.

© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Throughout our everyday experience, we are constantly accompanied by the pre-reflexive feeling of being “here and now”, experiencing the external world from the location and perspective of a body that we perceive as our own. The ensemble of these experiences has been termed bodily self-consciousness (BSC) and is considered the minimal building block of consciousness and self-awareness (Blanke et al., 2015).

BSC and the identification with our own body are so rooted in our everyday experience that they are easily given for granted as a part of our normal brain functioning. However, a rich body of experimental studies suggests that BSC is a continuously built, *in fieri* phenomenon, linked to specific neural mechanisms. The manipulation of various aspects of multisensory processing can alter key components of BSC, such as body ownership, inducing self-attribution of an external object (Botvinick & Cohen, 1998) or dis-embodiment of an actual body part (Longo et al., 2008; Della Gatta et al., 2016). These lines of evidence suggest that body ownership is the result of the multisensory integration of tactile, proprioceptive, and visual bodily stimuli in both the spatial and temporal domains. Accordingly, an influential theoretical view proposes that body ownership emerges when multisensory bodily stimuli are congruent with the normally experienced signals originating from one's own body and can be altered otherwise (Armell & Ramachandran, 2003). In line with Bayesian accounts of low-level multisensory integration (Ernst & Banks, 2002a), it has been proposed that such principle could be conceptualized as an inference-like process. Ramachandran (Armell & Ramachandran, 2003) proposed that the illusory ownership of a fake hand when stroked in synchrony with one's own hand (the rubber hand illusion, RHI), emerges from “Bayesian logic”. The idea was that, since the repeated co-occurrence of visual and tactile stimulation would be very hard to obtain by chance, the brain deems the hypothesis that the fake hand and the real hand are the same, i.e., one's own hand, as the most probable. Over the years, such inference-based accounts have also included the integration of mental states (Moutoussis et al., 2014) and interoceptive signals (Limanowski & Blankenburg, 2013; Seth, 2013; Seth & Friston, 2016; Seth & Tsakiris, 2018), linking inference on internal bodily and neural states to self-consciousness.

Arguably, one of the keys to the success of Bayesian approximations of brain function is that they constitute a normative framework, i.e.: they find a clear evolutionary motivation in the need to behave optimally in a noisy sensory environment. In its initial field of application to low-level multisensory integration processes, the validity of such approach has been rigorously proven experimentally by showing signatures of optimality in various setups and sensory modalities (Alais & Burr, 2004; Butler et al., 2010; Ernst &

Banks, 2002a; Thurman & Lu, 2014). Instead, although Bayesian descriptions of BSC have been popular for almost two decades, most accounts are purely conceptual (Apps & Tsakiris, 2014; Litwin, 2020; Noel et al., 2018; Seth, 2013; Seth & Friston, 2016) or mathematical (Kilteni et al., 2015). Experimental studies in their support, where modelling is paired with *ad hoc* behavioural assessments, are still rather scarce (Chancel et al., 2022; Fang et al., 2019; Samad et al., 2015) and do not provide conclusive proofs of the optimality of behaviour, weakening the motivation for the use of a normative model (see below for a detailed discussion).

Here, we aimed at extending the evidence base for Bayesian theories of BSC. To do so, we focused on body ownership as its key and arguably easiest component to be quantified. We tested a Bayesian model of hand ownership by means of a virtual reality-based reaching task, recently introduced to manipulate visual and proprioceptive disparity and derive an implicit measure of hand misattribution (Fang et al., 2019) (see below). We validated this model by rigorously assessing optimality through two alternative strategies: by testing the Bayesian model against appropriate competing models, and by separately measuring the unisensory visual and proprioceptive components.

1.1. Body ownership as the result of Bayesian causal inference

As previous Bayesian approaches to body ownership, the proposed model has its roots in classical models of multisensory integration (Ernst & Banks, 2002a). In such models, cues are weighted according to the inverse of their precision, under the assumption that they originate from the same physical source (forced fusion models). However, in the real world, stimuli occur simultaneously at multiple locations, and the brain needs to figure out which ones come from the same source and therefore have to be integrated (Shams & Beierholm, 2010). It has been suggested that this problem may also be solved in a probabilistic framework, i.e., Bayesian Causal Inference (Bayesian CI), whereby the brain infers the likelihood that two unisensory stimuli originate from the same cause, based on their spatial and temporal congruencies (Körding et al., 2007). This approach can be applied to the processing of unisensory bodily stimuli to explain how the feeling of owning a body as one's own could emerge from their integration. Based on a model of the expected mutual relations between the sensory stimuli normally originating from the body, the existence of the body itself would be inferred as their common physical cause. Then, the feeling of such a body as one's own would emerge by identifying with that “same old body always there” (James, 1890). This general principle has been translated into different mathematical formulations to model the relevant sensory variables (tactile, proprioceptive, visual cues etc.). As of now, such models have been experimentally tested only in two

empirical studies with different experimental setups (Fang et al., 2019; Samad et al., 2015).

Samad and colleagues (Samad et al., 2015) used a Bayesian CI model to account for the RHI, whereby the estimated probability of common cause (P_{com}) of visual and tactile inputs, as a function of the congruency between tactile, real hand stimulation and visual, rubber hand stimulation, is taken as a measurable estimation of ownership for the rubber hand. According to the model, P_{com} varies as a function of the spatial and temporal disparity between visual and proprioceptive cues about touch location and timing. Samad was the first to empirically test a specific prediction of a Bayesian model of body ownership. Indeed, as predicted by the model, the visual presentation of the rubber hand (in a position congruent with the participant's hand and within the hand peripersonal space), even in the absence of any tactile stimulation, was found to be sufficient to induce the illusion, although to a lesser extent.

More recently, Fang and colleagues (Fang et al., 2019) modelled ownership of a virtual hand as a function of visuo-proprioceptive disparity during a reaching task, based on an adaptation of classic paradigms of visuo-motor rotation (Krakauer, 2009). In their experiment, macaques and human participants had to reach targets with their real (proprioceptive) hand, hidden from view and replaced by a virtual hand, presented with various degrees of visuo-proprioceptive disparity. The error in the final reaching position induced by the virtual hand increased as a function of its displacement within a given range of visuo-proprioceptive disparity but decreased for larger levels of disparity. Such behaviour was well modelled by the predictions of a Bayesian CI model, inferring the probability of visual information from a virtual hand and somatosensory information from the real hand originating from the same physical cause (P_{com}). Since large visuo-proprioceptive disparities are unlikely to be generated by the same physical object (hand), P_{com} (and therefore, the probability that the virtual hand belongs to the subject) decreases proportionally to disparity, leading to attribute a lower weight to vision. Explicit ownership ratings in humans covaried with P_{com} , suggesting that this parameter might provide an implicit measure of subjective ownership probability at a trial-by-trial level.

1.2. Looking for a signature of optimality – limitations of previous studies

The works from Samad and Fang indeed constituted a major advancement. They first demonstrated that the Bayesian CI can be applied to model some aspects of body ownership in experimental setting. Furthermore, they developed the paradigms that made this possible. In particular, through Fang's virtual reality (VR) paradigm it is possible to parametrize the degree of visuo-proprioceptive disparity and to easily collect unprecedented amounts of data for modelling. However, despite such important achievements, these previous works do not provide conclusive evidence on whether body ownership can be truly described as the result of the optimal integration of multisensory signals for the following reasons.

As already anticipated, Bayesian models assume optimality as a normative constraint to brain function. Typically, optimality is defined as the behaviour minimizing squared errors on multisensory estimates (in this case: position estimates)

depending on a set of (unknown) free parameters (in this case: unisensory precisions). In the two previous studies, the match between behaviour and model predictions was used as a proof of optimal integration. Predictions from Bayesian CI are in line with classical principles of multisensory integration, stating that stimuli are integrated across modalities only if they are presented close in space and time (Cuppini et al., 2018; Stein et al., 2014; Stein & Stanford, 2008), i.e.: the “binding window” is spatio-temporally constrained. In order to show that those principles truly arise from Bayesian CI, the model should be tested against concurrent models implementing such spatio-temporal constraints outside the Bayesian framework. Until then, the pattern of decreasing visuo-proprioceptive integration with increasing spatio-temporal disparity observed by Fang and colleagues cannot be considered as an exclusive signature of optimality. Previous studies, instead, performed no model comparison, or only tested Bayesian CI against a forced fusion model, which also gives the Bayesian hypothesis for granted. Moreover, such model does not constitute a satisfactory competing model. Indeed, regardless of the Bayesian or non-Bayesian nature of the phenomenon, it is obviously wrong at large disparities due to the just mentioned spatio-temporal constraints on multisensory integration.

An alternative way to show that multisensory integration principles arise from Bayesian CI would be to directly link multisensory behaviour and its unisensory components. To do this, it is necessary either to manipulate the unisensory parameters in a controlled manner, or to directly measure them (Rohde et al., 2016). Compelling evidence in favour of optimal Bayesian CI would then be, for example, that the size of the spatial and temporal windows for integration vary with the amount of sensory noise consistently with model predictions. This was the strategy used in a very recent preprint by Chancel and colleagues (Chancel et al., 2022), who manipulated the amount of visual noise during the RHI. The authors found evidence in favour of a BCI model compared to a model that did not take into account the sensory manipulation. Alternatively, the windows for the integration of multisensory stimuli could be predicted from (independently measured) unisensory precisions. Such comparison was not possible in previous studies, where the unisensory parameters were either taken to be fixed values from the literature (Samad et al., 2015), or fitted from behavioural data (Fang et al., 2019).

In sum, we argue that rigorously testing the optimality assumption is crucial to support the normative justification of Bayesian models for BSC. This can be done either through model selection against appropriate alternative models, or by independently measuring/manipulating unisensory components.

1.3. Our approach

In the present work, we aimed at providing empirical evidence for the hypothesis that body ownership emerges from a Bayesian inference process, as a key component of quantitative bottom-up accounts of self-consciousness. To do this, we extended and revised the existing models and behavioural validations, focussing on spatial and temporal features of visual and proprioceptive inputs. We introduced a virtual reality adaptation of the reaching task used by Fang and colleagues – i.e., visuo-proprioceptive disparity task (VPD, Fig. 1a) – as a base

for the assessment of the Bayesian CI model predictions. Note that, since the VPD task is based on an active reaching movement, both visuo-proprioceptive and visuo-motor congruencies are manipulated. However, for simplicity, we will refer only to visuo-proprioceptive congruency throughout the text. Compared to the most used multisensory illusions, such as the RHI, this approach has two main advantages. First, the quantitative nature of the variables at play allows collecting granular data which is especially suitable for modelling. Second, the use of an active motor task (as opposed to the RHI or to a passive proprioceptive drift measure), allows to collect the large number of trials that is needed for the rigorous validation of a computational model.

In our approach, the link between subjective ownership and the Bayesian CI model can be only assessed indirectly, by correlating proprioceptive drift measures and subjective ratings. The link between proprioceptive drift and body ownership is an object of debate, as a correlation is not observed in all experimental setups (Abdulkarim & Ehrsson, 2016; Rohde et al.,

2011). The lack of a perfect overlap between body ownership ratings and proprioceptive drift may reflect the multicomponent nature of bodily experience, which includes both the feeling of a body part as belonging to one's own body as well as occupying a specific location in space (and in time). These are considered different, yet related components of bodily self-consciousness (Blanke et al., 2015; Blanke & Metzinger, 2009; Serino et al., 2013). In line with the approach introduced by Fang and colleagues, we chose to validate our model on proprioceptive drift as it is an implicit and objective measure that does not suffer the well-known limitations of explicit ratings and questionnaires. We tried to generalize such computational framework to body ownership in a second, separate step, by testing the correlation between proprioceptive drift and explicit ownership ratings. We argue that the advantages in terms of easiness of data collection and data granularity outweigh the lack of a direct model validation on ownership ratings.

In order to investigate different features of the Bayesian CI integration process, in addition to the spatial manipulation of

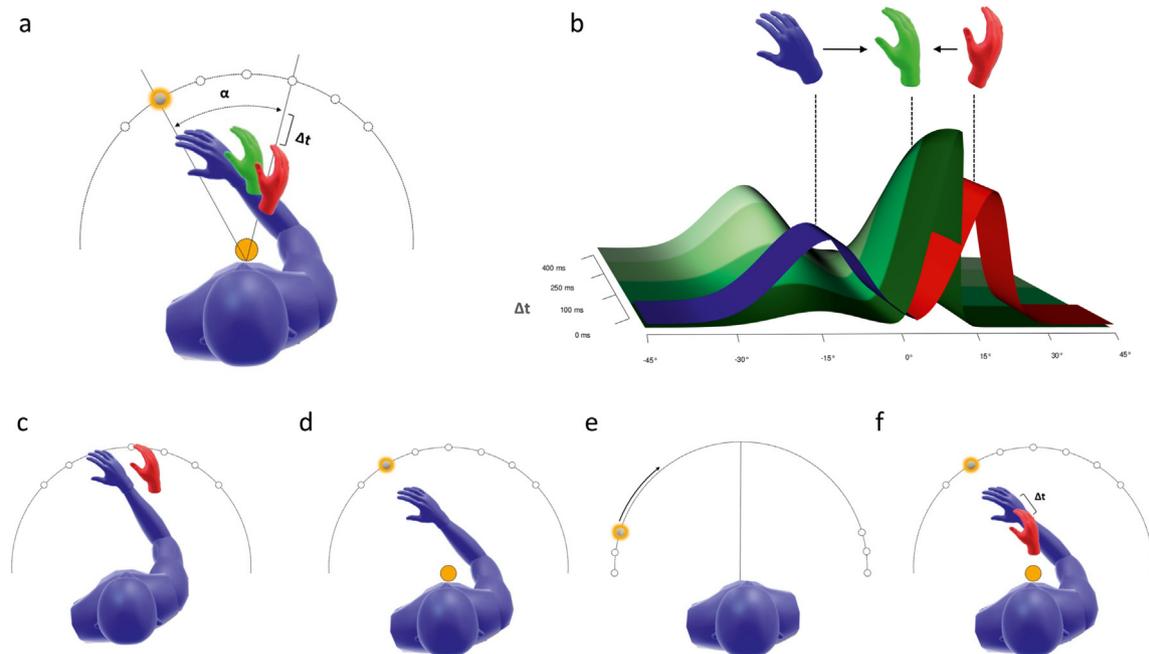


Fig. 1 – Experimental tasks and model rationale. In the visual-proprioceptive disparity task (multisensory task, a), a variable angle disparity α and temporal delay Δt are introduced between the real and the virtual hand during reaching movements towards a set of visual targets. The red hand represents the visual feedback from the virtual hand, the blue hand the proprioceptive feedback from the real hand, while the green hand is the final estimate resulting from visuo-proprioceptive integration. The relative weight attributed to the visual and the proprioceptive feedback determines the amount of error in the final position of the reaching movement. Hand position estimates (b) according to the Bayesian CI model (green) as a function of spatial and temporal disparities between the visual-virtual (blue) and proprioceptive-real (red) hands. The bottom row summarizes the set of tasks assessing each unisensory component. Proprioceptive precision: in the proprioceptive judgement task (PJ) (c) a virtual hand (red) is displayed in virtual reality at the left or the right of participants' real hand (blue), the perceived position of the hand has been determined using a two-alternative forced-choice converging algorithm. In the open-loop (OL) reaching participants reach targets without visual feedback of the hand (d). Visual precision (e): in the midline judgement task (MJ), participants have to report when they feel that a visual cue, moving across their visual field, is at their body midline. Temporal precision (f): in the simultaneity judgement task (SJ), participants evaluate the synchrony between the onset of their voluntary reaching movements and the displacement of the hand presented in virtual reality with a variable delay in the visual feedback. A visual morphing task (VM, not in the figure) has been used to measure the accuracy in the encoding of the visual features of ones' own hand, as part of the prior.

visuo-proprioceptive disparity as in Fang and colleagues, we also modulated the temporal disparity, by adding different levels of delay between the participant's movement and virtual reality visual feedback. The addition of a temporal modulation strengthens the interpretation of the behaviour observed in the VPD task and the underlying model. Indeed, ownership (or disownership) has to be a perceptually unitary phenomenon, regardless of whether it arises mainly from spatial or temporal cues. However, in a purely spatial task, reaching bias can be explained as the result of visuo-proprioceptive integration, with ownership being a mere epiphenomenon. If our temporal manipulation also induced the same reaching bias as the spatial manipulation, this simplistic explanation would be ruled out, and support the idea that ownership can be truly measured from reaching errors as the hidden variable linking spatial and temporal biases.

1.4. Main hypotheses of the study

In order to investigate whether body ownership can be modelled as a bottom-up process, whereby sensory information is integrated in a behaviourally optimal way, we have:

- I) Validated a Bayesian CI model of an implicit measure of body ownership, as the reaching bias induced by the temporal and spatial manipulation in the multisensory reaching task (Hypothesis I).
- II) Assessed if the implicit measure modelled by the Bayesian CI model also accounts for the subjective feeling of ownership, as assessed through explicit questions (Hypothesis II).

Hypothesis I. We aimed at performing a rigorous model validation by tackling the two main limitations of previous studies. That is, we compared our Bayesian CI model to a set of truly concurrent models (model selection) and compared model estimates to independently assessed unisensory parameters (unisensory correlations).

The first approach requires that a Bayesian CI model, taking into account both the spatial and temporal (BCI_{ST}) manipulation, should outperform an appropriate set of concurrent alternative models (Hypothesis Ia). First, we tested it against the forced fusion model (FF) (replicating Fang et al., 2019) and a model including only spatial disparity (BCI_S). Crucially, we further tested our data against two heuristic, non-Bayesian models, designed to well describe experimental data within a non-Bayesian framework (Heuristic and semi-FF, see methods for a detailed description). These models simply describe a continuous transition from integration to segregation, in line with basic multisensory integration principles but outside a Bayesian framework.

The second approach to validate the BCI_{ST} model is to independently investigate unisensory components to assess optimality. As mentioned, this can be done by either manipulating or measuring independently unisensory precisions. Here we chose the latter method due to the difficulty of accurately manipulating unisensory noise for both vision and proprioception in healthy humans during the VPD, multisensory task. Indeed, while manipulating only visual precision

would be possible in our setup, manipulating proprioception with the same reliability is problematic. Here, we aimed at mapping the contribution of the different unisensory modalities to the multisensory process with comparable resolution, and we therefore chose to measure, rather than manipulating, unisensory precisions. This approach is particularly relevant in light of future applications of the proposed methodology to clinical populations suffering deficits that can affect any sensory modality. The Bayesian CI model provides predictions of the reaching bias as a function of both spatial and temporal disparity, depending on four free parameters: σ_v , the unisensory visual precision, σ_p , the unisensory proprioceptive precision, the temporal precision σ_t and a global prior about ownership of the virtual hand P_π . The free parameters have been fitted from experimental data at the individual level by finding the parameter set that maximizes the match between model predictions and reaching bias. These parameters have been then independently extracted from a set of dedicated tasks (Fig. 1) and correlated with those fitted from the multisensory task, thus assessing if reaching errors, and therefore the size of the binding/ownership window, are coherently predicted by unisensory precisions (Hypothesis Ib).

Both model selection (Hypothesis Ia) and unisensory correlations (Hypothesis Ib) can be considered as compelling signatures of optimality. Therefore, we chose to consider model validation as achieved in case at least one of these two methods proves to be successful.

Hypothesis II. In order to establish a full link between Bayesian integration and subjective body ownership, we hypothesize that explicit evaluation of the ownership feeling during the VPD task should match the probability of multisensory integration (P_{com}), estimated by the Bayesian CI model both in the spatial and temporal domain. To test this hypothesis, subjective ratings have been collected and put in relation with reaching bias to ascertain its link with subjective ownership.

In summary, this study provides evidence to validate or reject the hypothesis that body ownership arises from a Bayesian CI process, supporting the quantitative approach to the investigation of BSC (see Table 1).

2. Materials and methods

2.1. Transparency statement

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.2. Bayesian CI model

Following several successful approaches to model multisensory integration in probabilistic terms (Fang et al., 2019; Körding et al., 2007; Rohe & Noppeney, 2015; Samad et al., 2015), we modelled the process of visuo-proprioceptive integration in a Bayesian Causal Inference (Bayesian CI)

Table 1 – Summary of preregistered hypotheses, sampling and analysis plan, and outcome interpretations.

Question	Does a Bayesian CI process combining visual and proprioceptive information in the temporal and spatial domain account for implicit body ownership?		Does subjective body ownership, as assessed through explicit questions, emerge from the same visuo-proprioceptive integration process?
Hypothesis	Hypothesis Ia The proposed spatio-temporal BCI _{ST} model, fitted to reaching errors in the VPD task, should outperform all the other alternative models (FF, BCI _S , Heuristic, Semi-FF).	Hypothesis Ib The parameters estimated from the unisensory tasks (σ_p , σ_v , σ_t , P_π) should correlate with the correspondent parameters extracted from the multisensory task through the BCI _{ST} model.	Hypothesis II The subjective ownership ratings are reflected by the estimated probability of a common source of visual and proprioceptive information about the hand (P_{com}) in both the spatial and temporal domain.
Sampling plan (e.g., power analysis)	Based on our Monte Carlo simulations, and Fang's previous study, a sample size of 22 subjects should be sufficient.	Based on Monte Carlo simulations, a sample of 35 participants should be sufficient to obtain a statistical power above 90%.	Assuming a true R value of .82, based on Fang's 2019 study, a 90% power with a significance threshold of .02 can be obtained with 10 participants.
Analysis Plan	We used BIC as an approximation for model evidence and compute the exceedance probability to test the BCI _{ST} model against all the alternative models.	We used the model fitting procedure described in the Analysis plan to extract unisensory parameters from the VPD task (σ_p , σ_v , σ_t , P_π). Then, we assessed the significant correlations of those parameters with the corresponding parameter extracted from unisensory tasks (OL, PJ, MJ, SJ, VM).	Pearson correlation scores between the standard deviation of the bivariate Gaussian fitted on ownership ratings and the width of P_{com} extracted from the VPD in both the spatial and temporal dimensions. The scores are expected to be significantly positive.
Interpretation given to different outcomes	Hypothesis Ia should be accepted if the BCI _{ST} model outperforms all the alternative models with exceedance probability >.95. Hypothesis I (and the underlying research question) should be accepted if either Hypothesis Ia or Hypothesis Ib holds. Then, the binding of visual and proprioceptive information leading to implicit ownership would show a clear signature of Bayesian processing. Otherwise, regardless of the Bayesian or non-Bayesian nature of the binding process, we conclude that the BCIST model cannot provide predictions that are refined enough to be informative for body ownership.	Hypothesis Ib should be accepted if any of the correlations with unisensory parameters is significant after Bonferroni correction ($p < .01$).	Hypothesis II should be accepted if both correlation tests yield the expected results. Otherwise, subjective ownership and reaching errors in the VPD task emerge from different phenomena. Then, the subjective feeling of ownership can be meaningfully modelled by modelling reaching errors in the VPD. If Hypothesis I also holds, this means that subjective body ownership is well modelled by the BCIST model.

framework. Early Bayesian models of multisensory integration, called forced fusion models, postulated that the brain estimates the position of a stimulus by simply combining unisensory estimates with a weight that is inversely proportional to their variance, quantified as mean squared error. Behaviour under such models is optimal (i.e.: it minimizes the mean squared error of position estimates) only under the condition that the sensory inputs considered in the different modalities always have the same physical source. Clearly, in real-life situations, where several stimuli are presented to different modalities simultaneously, this assumption is not granted. Therefore, before integrating unisensory estimates, the brain needs to infer whether and which stimuli need to be combined at all. Bayesian CI models account for this additional level of complexity by incorporating this inference in a probabilistic framework, in which the likelihood that two stimuli in different modalities have the same physical source is estimated from their features. In our case, this framework

was applied to the integration of visual and proprioceptive inputs about the hand, focussing on the specific factors that we expect to intervene in our multisensory task. As already extensively documented at the qualitative level by behavioural studies, the main factors contributing to hand ownership in a visuo-motor task are spatial and temporal visuo-proprioceptive congruencies. Therefore, to derive the equations of the BCI_{ST} model, we started from describing the generative model of the sensory stimuli underlying those congruencies, that is, the joint probability distribution of physical stimuli and their associated neural representation. In particular, in the case of our reaching task, the physical stimuli of interest are the hand position defined by visual and proprioceptive stimuli (s_v and s_p , expressed in degrees from the shoulder), and their relative timing with respect to the reaching movement (τ_v and τ_p , expressed in seconds). First of all, the visual and proprioceptive inputs may have one ($C = 1$) or two ($C = 2$) causes, that is, whether the virtual hand is or is

not the participant's hand. C is drawn from a Bernoulli distribution with probability P_π

$$P(C = 1) = P_\pi \quad (1)$$

Then, we model the joint probability distribution of visual and proprioceptive inputs in the spatial and temporal domain, conditional on whether $C = 1$ or $C = 2$. Given the radial nature of our task, we use the angle from target origin as the most natural coordinate for positions. If $C = 1$, then the visual and proprioceptive position of the hand is the same $s_v = s_p = s$ and is drawn from a uniform distribution on the $-90/90^\circ$ range, approximating the set of reachable angles. Previous works (Fang et al., 2019; Samad et al., 2015) used a Gaussian centred in 0 and with very large standard deviation ($\sigma = 10,000$) to approximate a uniform distribution. While simpler to treat analytically, this choice is problematic since the value chosen for the width of the positional prior influences the fitted value of the common cause prior P_π . This is because, while when the Gaussian is large enough it can always be approximated to a uniform distribution, the exact value of its standard deviation still influences model predictions through the normalization constant (see [Supplementary material](#) for the detailed calculation). In our case, by explicitly choosing a uniform distribution, the value of the normalization constant is naturally constrained by the reachable range and is thus less arbitrary. Similarly, the timing of visual and proprioceptive inputs related to the movement is the same, $\tau_v = \tau_p = \tau$, and is drawn from a uniform distribution. The range of the distribution was fixed at 0–30 sec, as a plausible value of the average interval between different movements. If $C = 2$, s_v, s_p are drawn independently from the same uniform distributions. As routinely done in Bayesian modelling, in order to simulate variability in sensory inputs, we assume that the true positions and timings of the sensory inputs are corrupted by unbiased Gaussian noise to generate their internal representations $x_v = s_v + N(0, \sigma_v)$, $x_p = s_p + N(0, \sigma_p)$, $t_v = \tau_v + N(0, \sigma_{t_v})$, $t_p = \tau_p + N(0, \sigma_{t_p})$. The first two variables refer to visual and proprioceptive-motor positions, and the other two to visual and proprioceptive-motor timing of movements, respectively. Then, the Bayes theorem allows to compute the posterior distributions for the positions of the stimuli and the number of underlying causes that an ideal observer would compute, provided that she/he knows the distribution of internal representations conditioned on the true positions and number of causes. Starting from the number of causes of the observed stimuli, we have:

$$P(C = 1 | x_p, x_v, t_p, t_v) = \frac{P(x_p, x_v, t_p, t_v | C = 1)P(C = 1)}{P(x_p, x_v, t_p, t_v | C = 1)P(C = 1) + P(x_p, x_v, t_p, t_v | C = 2)P(C = 2)} \quad (2)$$

Excluding for simplicity the region outside the $-90/90$ and $0/30$ sec range, where the contribution to the integral is negligible, the likelihood functions defined by our generative model are:

$$P(x_p, x_v, t_p, t_v | C = 1) = \frac{1}{\alpha} \exp \left[-\frac{1}{2} \frac{(x_v - x_p)^2}{\sigma_p^2 + \sigma_v^2} - \frac{1}{2} \frac{(t_v - t_p)^2}{\sigma_{pt}^2 + \sigma_{vt}^2} \right] \frac{1}{\alpha} e^{-\frac{\delta_s^2}{2\sigma_s^2} - \frac{\delta_t^2}{2\sigma_t^2}} \quad (3)$$

$$\alpha = \int_{-90}^{90} \int_0^{30} P(x_p, x_v, t_p, t_v | C = 1) \approx 2\pi\sigma_s\sigma_t \bullet 30 \bullet 180 \quad (4)$$

where α is the normalization constant, δ_s and δ_t denote spatial and temporal disparities respectively, and σ_s^2 and σ_t^2 are short forms for the sum of spatial and temporal variances. See [Supplementary material](#) for details about the approximation in equation (4). When there are two separate causes, we simply have:

$$P(x_p, x_v, t_p, t_v | C = 2) = \frac{1}{30^2 \bullet 180^2} \quad (5)$$

Therefore, the probability of common cause is given by:

$$P_{\text{com}} = P(C = 1 | x_p, x_v, t_p, t_v) = \frac{P_\pi \frac{1}{\alpha} e^{-\frac{\delta_s^2}{2\sigma_s^2} - \frac{\delta_t^2}{2\sigma_t^2}}}{P_\pi \frac{1}{\alpha} e^{-\frac{\delta_s^2}{2\sigma_s^2} - \frac{\delta_t^2}{2\sigma_t^2}} + (1 - P_\pi) \frac{1}{30^2 \bullet 180^2}} \quad (6)$$

The final estimate of hand position is obtained by combining the forced fusion and proprioceptive estimates, weighting them by the probability of common and separate causes, respectively:

$$\widehat{s}_p = P(C = 1 | x_p, x_v, t_p, t_v) \frac{\sigma_p^2 x_v + \sigma_v^2 x_p}{\sigma_p^2 + \sigma_v^2} + P(C = 2 | x_p, x_v, t_p, t_v) x_p \quad (7)$$

Then, we explored BCI_{ST} model predictions about sense of ownership (P_{com}) in a spatio-temporal disparity setup through numerical simulations. To illustrate model predictions, shown in [Fig. 2](#), we selected plausible parameters for the unisensory precisions and the prior, and a set “ground truth” temporal and spatial disparities, representing the actual physical spatial and temporal disparity between visual and proprioceptive inputs. Then, we added Gaussian noise of variance σ_s and σ_t respectively, in order to obtain samples from the noisy internal representation of the stimuli. As noted in [Körding et al. \(2007\)](#), this procedure is the only correct mean of simulating behavioural experiments within Bayesian models of brain function. The process was repeated 1000 times and the probability of common cause P_{com} was extracted following equation (6) and averaged across the 1000 trials. We show how P_{com} varies as a function of spatial and temporal disparity in [Fig. 2a](#). Coherently

with expectations and qualitative findings from behavioural studies, the analysis resulted in a region of very high ownership probability when spatio-temporal incongruences are below a certain threshold. This can be seen as the

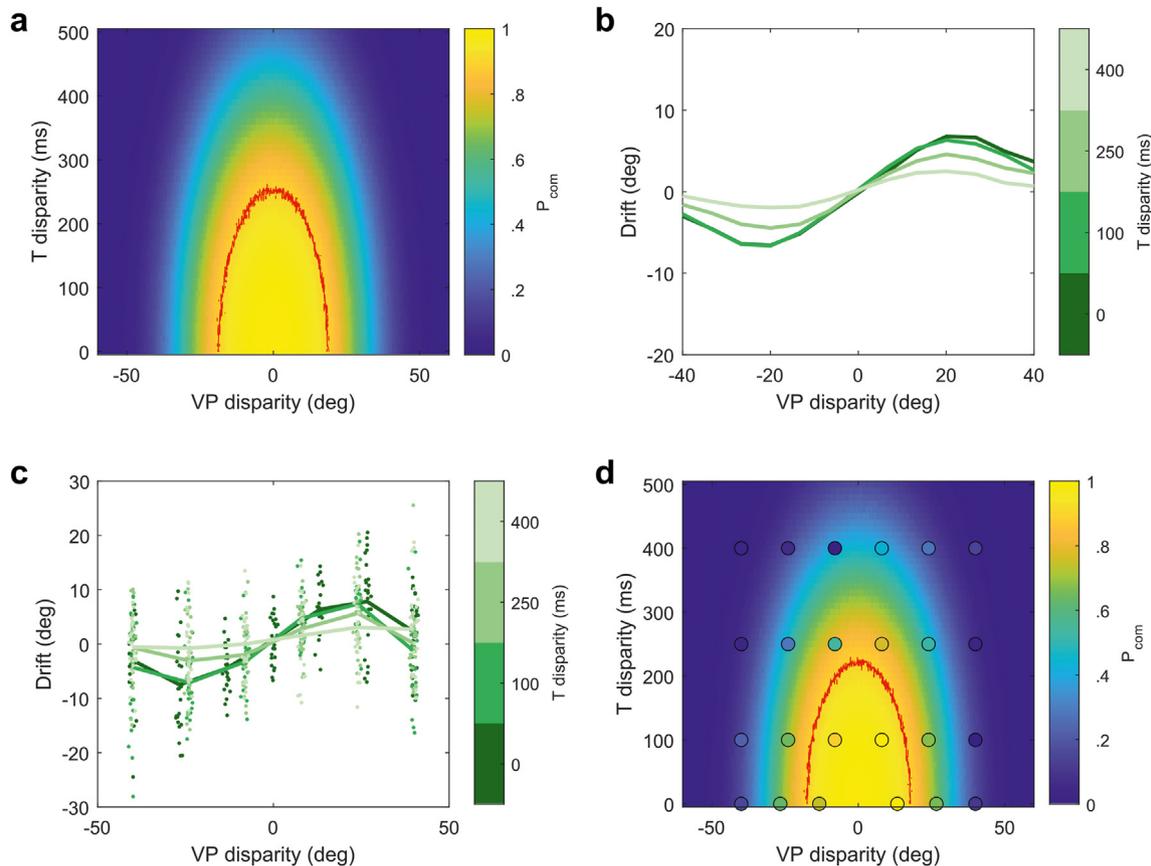


Fig. 2 – Model predictions and simulated experimental results for typical values of σ_v , σ_p , σ_t , P_π . Panel (a) shows the average probability of common cause as a function of spatial and temporal disparity, obtained after simulating BCI_{ST} model predictions for 1000 trials (in order to obtain a virtually noiseless prediction). The red half-circle denotes the area where the probability of common cause is above 95% and can be seen as the region corresponding to a subjective experience of complete ownership. Only positive values of temporal disparity are plotted, as they are the only that can be achieved in our experimental setup, but model predictions are symmetrical with respect to time. Panel (b) shows the (averaged, approximately noiseless) simulated results for a participant with the same typical parameters in the visual-proprioceptive disparity task. The y axis indicates the proprioceptive drift in the reaching movement, so that movements completely based on proprioception would have drift equal to 0, and movements completely based on vision would have drift equal to the shown disparity. Different degrees of green denote different values of temporal delay, increasing from dark to light green. Panel (c) is the result of the same simulation as panel (b), run with spatio-temporal disparities and number of trials matching the experimental design of our VPD task to obtain data from one surrogated participant. The 2D heat map in (d) shows how model predictions based on parameters fitted from noisy simulated experimental data match results based on the ground truth parameters used in panels (a), (b), (c). Values of σ_v , σ_p , σ_t and P_π were obtained by fitting the BCI_{ST} model on the simulated data shown in panel (c) and used to recover the expected probability of common cause as done for panel (a). The overall shape of P_{com} as a function of spatial and temporal disparity is very similar to the one obtained from ground truth parameters. Inside the black circles we show “empirical” P_{com} values, defined as the ratio between simulated drift and the forced fusion estimate, so a drift coinciding with the forced fusion estimate would correspond to $P_{com} = 1$, and no drift with $P_{com} = 0$. This analysis was only performed for visualization purposes and is not used to extract model parameters, as they are recovered more robustly by directly fitting reaching errors from the VPD task.

mathematical counterpart of the empirical notion that in the case of little or no disparity, as in normal conditions, the feeling of ownership is granted and constant. We then simulated the expected results from our VPD task, for the same set of parameters. For each spatio-temporal disparity, we simulated 1000 trials (a number large enough to render sampling noise negligible) by adding Gaussian noise to the real positions

and timings. Then, we extracted the hand position estimate according to BCI_{ST} model and computed the average reaching bias as a function of the spatial and temporal disparity. Finally, to illustrate how model predictions can be recovered from noisy behavioural data at the single participant level, we performed the same simulation, with the limited subset of spatio-temporal disparities and the number of trials of our actual

experiment. Fig. 2d shows the ownership probability, as extracted from our simulated behavioural experiment and Fig. 2c shows the reaching bias associated with the spatio-temporal disparities explored in our setup.

2.3. Alternative models

As a first model validation, we tested the spatio-temporal Bayesian CI (BCI_{ST}) model against the forced fusion model (FF), to replicate the analysis presented by Fang and colleagues. Using the same notation as in the previous paragraph, the forced fusion estimate is simply:

$$x_{\text{FF}} = \frac{\sigma_p^2 x_v + \sigma_v^2 x_p}{\sigma_p^2 + \sigma_v^2} \quad (8)$$

Additionally, we tested it against a Bayesian CI model including only the effect of spatial disparity in the estimate of $P_{\text{com}}(\text{BCI}_s)$. This is very similar to the original model proposed by Fang:

$$P_{\text{com}} = P(C = 1 | x_p, x_v, t_p, t_v) = \frac{P_{\pi\alpha} \frac{1}{\alpha} e^{-\frac{x_p^2}{2\alpha^2}}}{P_{\pi\alpha} \frac{1}{\alpha} e^{-\frac{x_p^2}{2\alpha^2}} + (1 - P_{\pi\alpha}) \frac{1}{180^\circ}} \quad (9)$$

As stated in the introduction, we also aim at challenging the BCI_{ST} model through alternative models that, unlike the FF model, are specifically designed to well fit experimental data and are outside the Bayesian framework. Only if the BCI_{ST} model also outperforms them, it can yield predictions that are subtle enough to advance our understanding of the link between body ownership and multisensory integration. Practically, we challenged the BCI_{ST} model with two models that heuristically implement the combination of visual and proprioceptive information through the notion of a spatial and temporal binding window in multisensory integration, as known from classical literature (Cuppini et al., 2018; Stein et al., 2014; Stein & Stanford, 2008).

The first model is completely non-Bayesian, and we will hence refer to it as the Heuristic model. We assume that visual and proprioceptive information are combined according to weights that do not depend directly on unisensory precisions. We then implement the spatial rule of multisensory integration by imposing that the visual weight decreases as a function of spatial and temporal disparity following a bivariate Gaussian, peaking at zero disparity. A Gaussian random error is added to the estimate to model neural noise.

$$\widehat{s}_p = \frac{s_p + s_v W_v e^{-\frac{x_p^2}{2\alpha_s^2} - \frac{t_p^2}{2\alpha_t^2}}}{1 + s_v e^{-\frac{x_p^2}{2\alpha_s^2} - \frac{t_p^2}{2\alpha_t^2}}} + N(0, \sigma_e) \quad (10)$$

here W_v denotes the weight of visual information at zero disparity, and the proprioceptive weight is set to 1 by definition, as the final estimate only depends on their ratio. α_s and α_t respectively indicate the spatial and temporal width of the “binding window”. Note that the estimate is based on the true position of the physical stimuli (s_p and s_v) instead of their noisy neural representation (x_p and x_v), since the model is purely heuristic, and the stochasticity of neural processing is rendered by the Gaussian error term added to the estimate.

The second model is based on the forced fusion model, and it therefore employs noisy internal representations as a starting point. However, unlike in the Bayesian CI model, the transition between the integration and the segregation regime is not governed by optimal inference, but again by the heuristic principle of a spatial and temporal binding window which is not optimised on unisensory precisions. In other words, the forced fusion estimate is obtained according to classical optimal integration, and it is then combined with the purely proprioceptive estimate with a weight that decreases as a Gaussian function of spatial and temporal disparity. Crucially, the spatial and temporal width of such binding window do not depend on unisensory precisions. We will therefore refer to this model as the semi-forced fusion model (Semi-FF).

$$\widehat{s}_p = e^{-\frac{x_p^2}{2\alpha_s^2} - \frac{t_p^2}{2\alpha_t^2}} x_{\text{FF}} + \left(1 - e^{-\frac{x_p^2}{2\alpha_s^2} - \frac{t_p^2}{2\alpha_t^2}}\right) x_p \quad (11)$$

As shown in Fig. 3, the predictions of both the Heuristic and semi-FF models on the average error as a function of disparity are very similar to those of the BCI_{ST}. The key difference is that, unlike in the BCI_{ST} model, the parameters governing the average disparity–error relation are partially or completely independent of the parameters governing the dispersion of individual reaching errors around the average. Therefore, both the Heuristic and Semi-FF models yield subtly but consistently different predictions, allowing disentangling the different underlying functions by model comparison on behavioural data (see [Sampling plan](#)) to assess the presence of a Bayesian signature in the integration process.

2.4. Materials and general procedure

The battery of tasks is administered via the Oculus Rift S Virtual Reality system, comprising an Oculus Rift S head-mounted display (HMD) and two Oculus Touch, or motion controllers. visual morphing (VM) is implemented in Python, while all the other tasks are implemented in Unity and are compatible with other virtual reality systems allowing 6 axis hand and head tracking (e.g., HTC Vive). A virtual reality implementation of our setup was preferred to a physical implementation for several reasons. First of all, the tracking accuracy achieved by modern HMDs, albeit still inferior to dedicated kinematics recording systems (e.g., Vicon optical camera system), allows to record kinematics with a precision that should be largely sufficient for our tasks (<1 cm) (Jost et al., 2019; Spitzley & Karduna, 2019). Second, the usage of a commercial, readily available apparatus, allows a quick and standardized replication of the tasks for large-scale data collection and sharing. Third, the use of an immersive environment allows to fully control and standardize visual inputs, while enhancing the vividness of the task experience.

The behavioural experiment consisted of a main task (VPD) and five complementary tasks. The VPD task yields an estimate of ownership of a virtual hand as the tendency to integrate its visual position with proprioception while estimating one's hand position. It consists of the repetition of a reaching movement while a variable angular disparity or temporal delay is introduced between the movement of participants' real hand and a virtual hand displayed in immersive VR. The

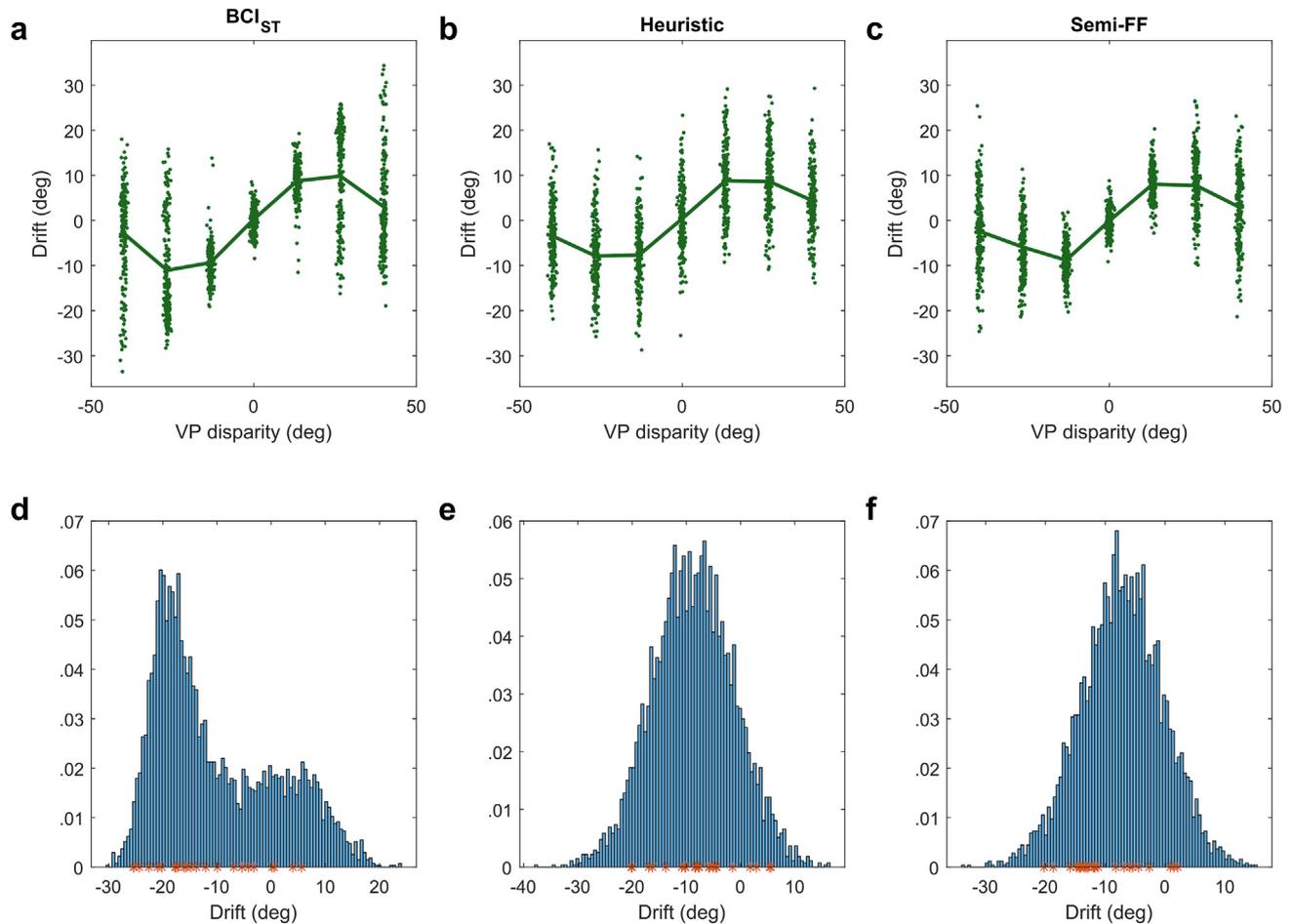


Fig. 3 – Comparison of BCI_{ST}, Heuristic and Semi-FF model predictions. Panels (a–c) show the induced drift in the VPD task for a simulated subject in the three models respectively. Model parameters were set to plausible values, and the data was simulated first for the BCI_{ST} model. Then, such data was fitted with the Heuristic and Semi-FF model to obtain the set of corresponding parameters able to generate the most similar results in the two alternative models. Those parameters were then used for panels (b–c). This way, the three models can be compared in the case where their predictions are the most similar. For simplicity, only the zero temporal lag was included in the figure. The mean drift (solid line) is rather similar across the three models, but the dispersion around it is different at all disparities between the BCI_{ST} and Heuristic models, while the difference is present mainly at intermediate disparities between the BCI_{ST} and the Semi-FF models. As shown in our Sampling plan, this difference should be enough to disentangle the different models with the planned sample size. Panels (d–e) show the full distribution of simulated drifts for the three models at -26.6° degrees of disparity, where the difference between the BCI_{ST} and the other models is most evident, in that only the BCI_{ST} model predicts a skewed distribution. The red asterisks show 24 simulated reaching errors (the same number of trials of our actual task), to qualitatively demonstrate that the difference can be appreciated in our experiment.

other tasks assessed independently the relevant parameters included in the BCI_{ST} model (i.e., σ_p , σ_v , σ_t and P_π).

To assess the proprioceptive precision (σ_p) we considered several options. Position matching tasks are commonly used in literature to assess proprioception. However, these tasks present some difficulties, as they may require a memory component (Rincon-Gonzalez et al., 2011), or the computation of a symmetric target location compared to a reference position in the contralateral space [i.e., a mirror-matching task (Dukelow et al., 2010)]. Another important disadvantage in using a mirror-matching task would be that the proprioceptive precision of the active body part (in our case, the right hand) cannot be

disentangled from the precision of the target body part (in our case, the left hand). This issue is especially relevant since this model and the associated tasks may be applied in the future to the study of clinical populations with lateralized deficits (e.g., stroke patients suffering a contralesional upper limb impairment).

For this reason, we searched the literature for previous studies in which the proprioception precision was assessed unilaterally (Desmurget et al., 2000; Haggard et al., 2000; Jones et al., 2010; Longo & Haggard, 2010). We devised a proprioceptive judgement task (PJ, Fig. 1c) in virtual reality in which the felt position of the hand is assessed through a forced choice converging algorithm. However, the parameter σ_p

estimated through the model from the multisensory task likely incorporates motor components that are not captured by a purely proprioceptive task. Therefore, in addition to the PJ, an open-loop reaching task (OL) was used to isolate the combination of motor and proprioceptive components which can determine the end-point precision in a reaching task (Fig. 1d). In this task, participants performed the same reaching movements towards virtual visual targets as in the multisensory task (VPD), but in the absence of visual feedback about hand position.

Concerning visual precision, it is worth noting that, despite visual acuity is extremely high compared to proprioceptive precision in humans (Jones et al., 2010; Kniestedt & Stamper, 2003), the final accuracy in visually determining the hand's position does not depend on visual acuity alone. Indeed, when coordinating vision and proprioception in a motor task, it is necessary to transform the extremely accurate retinotopic visual information in body-centred coordinates, through a set of computations involving gaze angle and head orientation (Buneo et al., 2002). In this view, we believe that participants' precision in visually determining their body midline, as assessed by a midline judgement (MJ) task (Fig. 1e), can be used as the closest approximation to isolate the contribution of visual information to position estimates in our task.

The temporal precision is the third main parameter of our model. This component was measured through a simultaneity judgement task (SJ), whereby participants evaluated the synchrony between the onset of a voluntary reaching movement and the displacement of the hand displayed in virtual reality (Fig. 1f).

The purely “bottom-up” Bayesian approach focussing on visuo-tactile congruencies, initially used to model body ownership, has been extended in conceptual models to incorporate also “top-down” cognitive constraints, given by the posture and the visual appearance of the body (Tsakiris, 2010; Tsakiris et al., 2010; Tsakiris & Haggard, 2005). These (Armel & Ramachandran, 2003; Ehrsson, 2005; Makin et al., 2008; Tsakiris, 2010; Tsakiris et al., 2010; Tsakiris & Haggard, 2005) have been described as the result of another inference process, comparing incoming visual features with an internal model of the body to estimate the probability that they originate from one's own body (Apps & Tsakiris, 2014; Kilteni et al., 2015; Tsakiris, 2017). Nevertheless, the role of visual appearance has not yet been studied quantitatively within a Bayesian framework. Although the problem is too complex and high dimensional to be explicitly modelled in a Bayesian CI model, its influence is reflected in the global “prior” as the marginal probability of all the factors that are not modulated during the experiment and included in the other parameters' computation. Based on previous accounts (Tsakiris, 2010) and as suggested by Fang's study (Fang et al., 2019), most of the variance in such “prior” is likely explained by high-level visual features of the stimulus. In the attempt to quantify their role, we implemented a VM task, based on the continuous morphing of pictures of the participant's real hand into other peoples' hands. Similarly to what was planned for the visual and proprioceptive precision in the spatial and temporal domain, we measured participants' accuracy in recognizing their own hand against other realistic hands. Note that we deliberately chose not to test their recognition

ability against virtual hand avatars (as the ones used in the VPD task), in order to make the task more challenging, according to internal pilots. Furthermore, this would allow to demonstrate that the visual recognition ability predicts behaviour in a general, rather than an idiosyncratic manner, allowing a more compelling validation of the Bayesian CI model. If the inter-individual differences in discriminating the visual appearance of one's own hand play a quantifiable, probabilistic role in determining ownership, we expect that the accuracy in the VM task should correlate with prior probability of cue combination fitted from the multisensory task.

2.4.1. Visual-proprioceptive disparity (VPD) task

Participants are requested to sit in front of a chest-height table, with their arm placed in front of them. Participants wear a head-mounted display (HMD) and hold a motion controller in their right hand. During the experiment, participants cannot see their real hand, but a realistic hand is displayed in virtual reality using the tracking of the motion controller. During the task, the spatial congruency between visual and proprioceptive information is manipulated as an angular disparity between the real (proprioceptive) and a virtual (visual) hand. Moreover, a delay is introduced between the onset of the real and the displayed movement in order to alter the temporal congruency of the stimuli (Fig. 1a).

Participants are asked to make reaching movements to targets in virtual reality (white spheres with 3 cm diameter) from a fixed starting position. The starting point is a sphere of 15 cm diameter, fixed 15 cm away from the participant's sternum. Target positions are arranged on an arc centred on the resting position. The arc radius is set according to each participant's maximum reaching distance, calibrated at the beginning of the experiment.

The task consists of three experimental blocks with slightly different designs. In the first three blocks, 7 targets (from T1 to T7) are equally spaced between -45 and 45° with respect to the participant's sternum. Across trials, the visual hand is randomly rotated with a given angular disparity from the participants' proprioceptive hand, with their sternum as the (vertical) rotation axis. Additionally, a temporal delay of 0, 100, 250 or 400 msec is added between the onset of the movement and the displacement of the virtual hand. For the 0 msec delay condition, 7 spatial disparities are used: 0° , $\pm 13.3^\circ$, $\pm 26.6^\circ$ or $\pm 40^\circ$ (+: clockwise, CW; -: counterclockwise, CCW). For 100, 250 and 400 msec delay conditions, 6 spatial disparities, uniformly distributed on the same range, are used: $\pm 8^\circ$, $\pm 24^\circ$ or $\pm 40^\circ$. This was done to increase the variability of the explored disparities, and to avoid collecting uninformative trials at zero spatial disparity. All the possible combinations between target position, temporal and spatial disparity [$7 \times (7 + 6 \times 3)$] are tested in randomized order for a total of 175 trials in each of the first three blocks. In the fourth block, in which subjective ratings of ownership are also collected (see below), only 3 targets are presented to keep the total duration constant, and again one trial is collected for each combination of target, disparity and delay (75 trials in total). Each block lasts approximately 15 min, and 600 trials were collected over approximately 75 min (including 5 min breaks between blocks).

Participants are requested to place their hand on the starting position to initiate a trial. At the beginning of the trial, the virtual hand is rotated by one of the possible disparity angles during 1 sec, and the target appears. This mismatch between the real (proprioceptive) and the virtual (visual) hand is maintained for 1.5 sec as the preparation period. In order to make apparent the temporal delay along with the angular disparity during the preparation period, participants are instructed to make a movement of prono-supination of the hand at a speed of approximately 1 Hz while fixating the virtual hand. After the preparation period, the target is turned green as a “go” signal. Movement of the hand outside the resting position at any time before the “go” cue automatically restarts the trial. Participants are instructed to reach the target with their real (proprioceptive) hand and return to the resting position within 1.5 sec, ending the trial. The spatial and temporal mismatch is maintained throughout the whole trial along with the hand movement.

Additionally, in the fourth block participants are requested to report their subjective feeling of ownership for the virtual hand through the virtual interface, evaluating their agreement with the statement [adapted to VR from Fang et al., 2019] “I felt as if the virtual hand was my hand” on a 1–10 Likert scale, by pointing to the answer with the motion controller.

2.4.2. Proprioceptive judgement (PJ) task

The set-up of this task is similar to the VPD (Fig. 1c). Participants' real hand is not displayed in virtual reality. The experimenter passively moves participants' real hand to one out of 7 possible target position (from T1 to T7) arranged at 0° , $\pm 15^\circ$, $\pm 30^\circ$, $\pm 45^\circ$ with respect to participants' sternum on an arc with radius equal to each participant maximum reaching distance. Target position is selected randomly trial by trial. A two-alternative forced-choice converging algorithm is used to find the position in which the participants perceive their hand. At the beginning of each trial, a virtual hand is displayed at $+30^\circ$ (right) or -30° (left) with respect to participants' real hand. The sign of the initial angle is randomized trial by trial. Participants then report whether they feel that the displayed hand is located to the left or right of their real, unseen hand. In the following step, the position of the virtual hand is moved halving the angle and mirroring it in the opposite direction with respect to participants' previous answer. In five steps, the algorithm converges towards a certain angle at which participants have an equal probability of reporting left or right. The proprioceptive-based estimation is computed as the intermediate hand position between the last displayed position and the next position that would have been displayed by the algorithm according to the participant's last answer. Each target position is tested 4 times in randomized order, for a total of 28 trials.

2.4.3. Open-loop reaching task (OL)

The set-up of this task is similar to the VPD task (Fig. 1d), besides the fact that the virtual hand is not displayed; therefore, participants do not receive any visual feedback about their hand for the entire duration of the experiment. From a fixed starting position, participants are asked to make a reaching movement to one out of seven visual targets (from T1 to T7), arranged at 0° , $\pm 15^\circ$, $\pm 30^\circ$, $\pm 45^\circ$ with respect to

participants' sternum. Target position is selected randomly trial by trial. Participants are required to place their hand on the starting position for 1.5 sec to initiate a trial. After the initiation period, one of the targets appears. The reaching target is then turned green as a “go” signal. Movement of the hand outside the resting position at any time during the initial resting period automatically restarts the trial. Participants have to reach the target with their real hand and come back to the resting position within 1.5 sec, ending the trial. Each participant is asked to complete 10 trials for each target ($10 \times 7 = 70$ trials).

2.4.4. Midline judgement task (MJ)

In this task, participants are asked to sit on a chair keeping their head and trunk aligned while wearing an HMD. On each trial, a white sphere with 3 cm diameter moves horizontally across participants' field of view at a speed of $10^\circ/\text{s}$, starting from $\pm 45^\circ$, $\pm 40^\circ$, $\pm 35^\circ$, and $\pm 30^\circ$ from the body midline, on an arc centred on participants' sternum, with a radius equal to their maximum reaching distance, as in the VPD (Fig. 1e). Participants have to report when they feel that the visual cue is aligned with the midline of their body by pressing a response button, with the possibility to subsequently ask the experimenter to manually adjust the judgement. The starting positions of the visual cue are randomized across trials. 24 trials in total are collected.

2.4.5. Simultaneity judgement task (SJ)

This task consists of a series of reaching movements in virtual reality towards 10 targets (from T1 to T10) equally spaced between -45° and $+45^\circ$ with respect to participant's sternum (Fig. 1f). On each trial, the displacement of the hand in virtual reality is delayed by a variable amount with respect to the onset of the movement of the real hand, spanning 8 values equally spaced between 0 and a maximum of 175 or 350 msec (see below). 10 trials are collected per temporal disparity, with each target repeated once. The order of targets and delays was randomized across the task. On each trial, participants are asked to report whether their movement and the displacement of the virtual hand occurred at the same time answering the question “did you notice a delay between the virtual hand and your hand?”. The answer is recorded as a binary variable by the experimenter.

A practice session was conducted to let participants familiarize with the task and to determine for each subject the maximum delay for the experiment. The practice session consisted in 15 trials: five for each of the 0, 175 and 350 msec delays. If the participant noticed the delay (when present) in at least 9 out of 10 trials, the maximum delay was set at 175 msec and 350 msec otherwise.

2.4.6. Visual morphing task (VM)

At the beginning of the task, 3 digital pictures of the participant's right hand are taken, with 3 different postures with varying distance between the fingers (narrow, medium and large). The pictures are converted to black and white, scaled to 300×400 pixels, and the background is removed. Each picture is morphed towards 10 target hands from a fixed database of hands (5 male and 5 female lab members). The morphing was performed using an automated feature mapping software

(Liao et al., 2017). Ten intermediate morphing steps are created for each target, each frame representing a 10% incremental change from participant's hand to a target hand or from 100% self to 0% self. A random rotation and translation (uniform on the $-10/10^\circ$ and $-10/10$ pixels range, respectively) are added to each image to prevent the participant from learning specific orientations-positions of the hands. The morphing of each image is checked by visual inspection prior to the task, and generates 100 morphing steps, uniformly distributed between 0 (i.e., 0% self) and 100 (100% self). For each of the 10 target images, 15 steps of morphing are selected, in such a way that sampling is more frequent for intermediate (harder to recognize) levels of morphing. In the end frames number 1, 17, 28, 36, 41, 45, 48, 50, 52, 55, 59, 64, 72, 83, and 100 are selected. One frame for each target image and level of morphing is selected based on the quality of the morphing (e.g., absence of deformations, discontinuity of the margins, etc.). Hence, a set of 100 images is created. Participants sit in front of a screen, with their hand occluded from view. At each trial, one of the 100 possible images is presented on a computer screen, while the question "is this a picture of your hand?" is displayed at the top of the image. Participants answer to the question by pressing a left or a right button, for negative and positive answers respectively, and their response and reaction times are recorded.

2.5. Sampling plan

2.5.1. Inclusion criteria and missing values

According to the power analysis described below, 40 right-handed participants were recruited. Inclusion criteria were: no history of neurological, vestibular or psychiatric disorder, normal or corrected to normal binocular vision for VR. Participants were informed about the inclusion criteria beforehand and asked to apply only if no criteria are violated.

There was no outlier removal in the collected data. If at any point technical issues arose during an experiment interfering with the experiment's procedure or data-logging, the participant was excluded from the experiment in which the issue emerged. Issues rated as interfering with the experiment's procedure include any type of freezing of the displayed virtual environment or any other faulty distortion of the presented virtual environment. Technical issues were recognized by the experimenter, who monitored the procedure of all experiments on a separate display. If a participant wanted to stop the experiment due to motion sickness or any other discomfort, he/she was excluded from the experiment. Each task was considered complete if the participant performed at least 80% of the trials. Participants who failed to complete more than one task were excluded from the experiment. In addition, any participant with less than 80% of the trials in the VPD task was excluded. All excluded participants due to the above reasons were replaced with another participant. These criteria were established before data collection.

2.5.2. Monte Carlo simulations approach

Due to the complexity of the planned analysis, and the scarcity of published empirical data using a similar setup, we could not perform a power analysis through standard techniques for most hypotheses of our study. Therefore, in order

to determine the optimal sample size, we relied on a custom method combining Monte-Carlo simulations and previous data to estimate the chances of observing the hypothesized effect. In short, we used model fits from Fang's (Fang et al., 2019) largest experiment to infer a plausible distribution of the main model parameters, and we simulated behavioural results from 500 surrogate participants, assuming that the BCI_{ST} model, described in Section 2.2, is correct. Then, our analysis pipeline was repeatedly run on several random samples of surrogate participants of different sizes, and the fraction of resamples yielding significant results were computed. This should provide an unbiased estimate of the probability of observing an effect, assuming that the BCI_{ST} model is correct.

2.5.3. Hypothesis Ia: model validation through model selection

Concerning the hypothesis that the BCI_{ST} model would outperform both the BCI_S model and FF model, since real data from a similar experiment is available, we based ourselves on the sample of the largest experiment performed by Fang on humans. A sample size of 22 participants was largely sufficient to select the best model with fewer trials per participant.

In order to assess the probability of the BCI_{ST} model to outperform the Heuristic and Semi-FF models, we fitted the surrogated participants (simulated with the BCI_{ST} model as ground truth) also with these two alternative models, and compared them to the BCI_{ST} model by computing the exceedance probability over resamples. We explored sample sizes ranging from 5 to 60 in steps of 5. For each sample size N , subsets of N participants were randomly selected 10,000 times (with replacement) from our pool of 500 simulated participants, and the fraction of resamples yielding significant results was computed. The models yielded a good fit compared to the FF model (R^2 BCI_{ST} : $.911 \pm .0016$; Heuristic: $.884 \pm .002$; Semi-FF: $.893 \pm .002$; FF: $.854 \pm .003$). Nevertheless, when tested by computing the exceedance probability, the BCI_{ST} model overwhelmingly outperformed both alternative models even with 5 participants (power = 100%). The average log-likelihood difference in favour of the BCI_{ST} model was very large: 65.3 ± 43.4 (SD) against the Heuristic model, and 47.66 ± 30.7 (SD) against the Semi-FF model. Indeed, the brain is expected to only approximately perform Bayesian CI, while our simulations are based on an exact implementation of the model. Nevertheless, our analysis shows that the predictions of the different models differ quite radically. If the brain approximates optimal Bayesian CI closely enough to render the Bayesian hypothesis meaningful within this context, our model comparison pipeline should be able to detect it with 40 participants.

2.5.4. Hypothesis Ib: model validation through unisensory correlations

The above-described Monte-Carlo approach was then used for assessing the correlations between model parameters extracted from VPD task (σ_p , σ_v , σ_t), and unisensory precisions extracted from the PJ and OL (σ_p), MJ (σ_v) and SJ tasks (σ_t).

First of all, ground truth values for the parameters σ_p and σ_v were drawn from a distribution modelled on the parameters extracted from our pilot for 500 surrogate participants.

Therefore, values of σ_p were drawn from a Gaussian of mean 5.56, while values of σ_v drawn from a Gaussian of mean 5.23. The means of the randomly generated parameters were the same as what was obtained in Fang's (Fang et al., 2019) VPD experiment with 22 participants, the largest similar dataset available. The standard deviation of the Gaussians was set to 2°. A lower and upper cut-off of 3 and 8° were set on both values to avoid unrealistic extreme values. For σ_t , since it was not possible to collect pilot data, we assumed a distribution based on literature about the perception of delay in a visuo-motor task. Fitting our model of the SJ on data extracted from figures in the only previous study that was well adapted to this goal (Farrer et al., 2008) indicates a value of σ_t of about 85 msec (see Supplementary material for details). To cover a plausible and wide range of temporal precisions, we assumed σ_t to be uniformly distributed in the range 50–120 msec, which is symmetrical around 85 msec. In order to be conservative, the power analysis was performed only with the longer delays (0–350 msec), which should guarantee that in the actual experiment the values of σ_t should be extracted with equal or greater precision than in our simulations. Finally, values of P_π were uniformly distributed between .4 and 1, symmetrical around the average reported value of .7. Then, the VPD, PJ, OL, MJ and SJ tasks were simulated for each participant.

In order to simulate the VPD task, we simply applied the same procedure used to simulate the trials at a given amount of visuo-proprioceptive disparity, used for model fitting and described in the previous section. In order to simulate the reached position in the OL task, or the judged position in the PJ task, we added Gaussian noise of standard deviation σ_p to the

target position for each trial. Similarly, for the MJ task, we generated 24 normally distributed values, with standard deviation σ_v . Finally, we used the same procedure developed for model fitting to simulate the SJ task.

After the data was generated, we extracted the unisensory parameters from the multisensory and unisensory tasks as described in the Analysis plan (Section 2.6.5). Then, we evaluated the probability of observing significant effects as a function of sample size through repeated resampling, as described previously. As a final outcome, we chose the statistical significance of the Pearson correlation between the values of σ_p , σ_v , σ_t obtained by fitting the BCI_{ST} model on the multisensory task, and their unisensory counterparts from the other five tasks, with a threshold of $p = .004$.

This procedure allowed us to identify the major sources of uncertainty in model fitting, especially for the multisensory task, where all parameters are estimated at once, and optimize the experimental design accordingly. The final combination of spatial and temporal disparities presented in the previous sections was selected between several possible designs, as the one maximizing power while keeping the expected duration of the multisensory task below 90 min. Similarly, the number of repetitions in the unisensory tasks was defined by heuristically optimizing the trade-off between task duration and statistical power, based on the expected precision in measuring the relative model parameter as seen in Fig. 4c. As summarized in Fig. 4a, the analysis shows that a sample of 35 participants should be sufficient to obtain a power above 90% for all the three correlations with a significance threshold of .004 (92.7% for the MJ task with 15

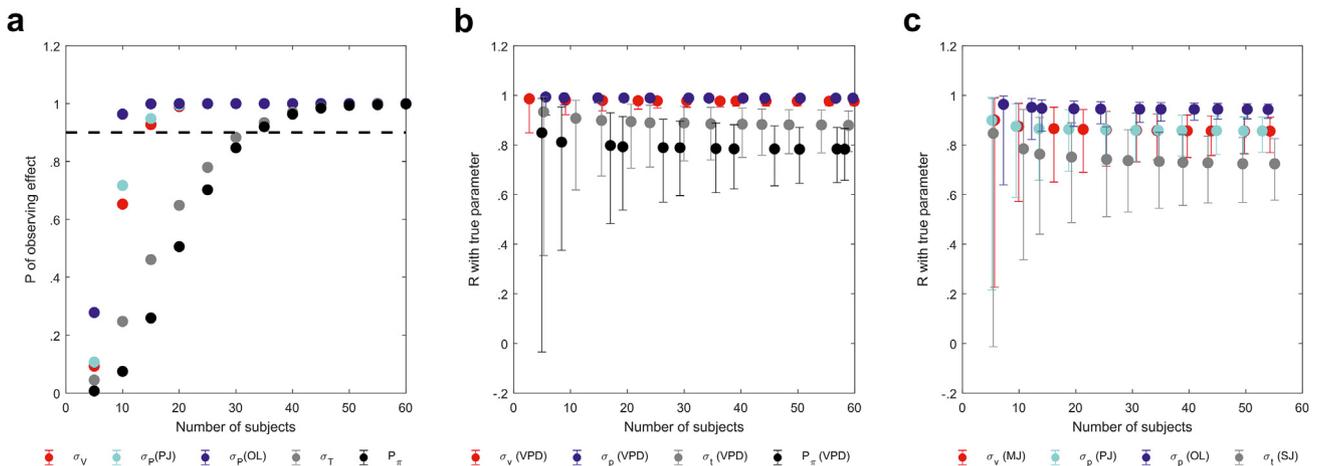


Fig. 4 – Power analyses. Panel (a) shows the results of the power analysis conducted by Monte-Carlo simulation for Hypothesis 1b. We plot the probability of observing a significant effect after Bonferroni correction ($p < .004$) as a function of the number of participants. The red, cyan, blue and grey dots denote respectively the correlations between the values of σ_v , σ_p , σ_t as extracted from the multisensory task (VPD) and its unisensory correspondents. The black dots indicate the estimated probability of observing a significant correlation between the slope of the VM task and the value of P_π extracted by fitting in the VPD task. Probabilities for each sample size were computed over 10,000 random draws, with error bars estimating the 95% c.i. due to the sampling procedure. Panel (b) shows the average Pearson correlation value between true parameters used in simulations and fitted values, over 10,000 random draws of simulated participants. The colour-coding of different parameters is the same as in panel (a). Panel (c) shows the same average correlations between parameters extracted from unisensory tasks and true parameters used in simulations. Error bars in panels (b) and (c) denote 95% confidence intervals obtained by applying the Fisher transformation to correlation coefficients and then applying the inverse transformation to the confidence intervals on the Z scores obtained in such a way.

participants, 94.8 with 15 participants for PJ, 96.4% with 10 participants for OL and 93.4% with 35 participants for SJ).

Finally, as no simple mathematical model can be used to describe the expected relation between the VM task and the prior on common cause in the visuo-proprioceptive disparity task (P_{π}), we relied on more conventional methods, based on the assumed value of the correlation. Our simulations indicate that the expected correlation between the true prior and the value fitted from the VPD task should be $R \sim .777$. We assumed the morphing slope value would have the same correlation with the true value of P_{π} . For each sample size N , we selected 10,000 times N surrogated participants and generated n random vectors with an imposed correlation coefficient of $.777$ with respect to the true value of P_{π} . The obtained values were then correlated with the fitted value of P_{π} , and a significance test was performed with $\alpha = .004$. After computing the fraction of significant correlations, we found that a sample size of 35 would allow to reach a power of 92%.

2.5.5. Hypothesis II: link with subjective ownership

Concerning the hypothesis that subjective ratings should correlate with reaching biases, we based sample size estimation on the previous work by Fang (Fang et al., 2019), who performed a very similar analysis with fewer trials per participant. Assuming a true R value of this correlation to be equal to $.82$, a 90% power can be obtained with 10 participants.

In summary, based on the hypothesis that requires the largest sample size, we estimate that 35 participants should be sufficient for meeting our statistical power target of 90%. To be conservative, we oversampled to 40 participants.

2.6. Analysis plan

Overall, the main hypothesis of our study is the validity of the Bayesian CI model in both the spatial and temporal domains (BCI_{ST}) as a potential mechanism for body ownership. This was done in two steps: model validation (Hypothesis I) and linking the model predictions to subjective ownership (Hypothesis II).

2.6.1. Hypothesis Ia: model validation through model selection

We fitted the BCI_{ST} , FF, BCI_S , Heuristic and semi-FF on our multisensory reaching task (VPD), and then test the BCI_{ST} model against the concurrent models by evaluating the exceedance probability (original code available from: <https://github.com/sjgershm/mfit>).

For all models, the fitting approach is similar to the one of Fang and colleagues (Fang et al., 2019). For each spatial and temporal disparity, we simulated 5000 trials, and maximize the likelihood of the data given the simulated model predictions, with respect to the set of fitted parameters (e.g., σ_v , σ_p , σ_t , P_{π} , for the Bayesian CI model). The fitting was performed through the BADS Matlab optimization tool (<https://github.com/lacerbi/bads>) (Acerbi & Ma, 2017). To avoid convergence problems or poor optimization, the fitting procedure was repeated 5 times with different randomly selected starting parameters, and the fit with the highest log-likelihood was selected. No pre-processing step was performed on the data, except for the removal of systematic biases in reaching

(possibly due to tracking or VR calibration). This was done by subtracting, for each participant, the mean reaching bias at zero spatial and temporal disparity.

2.6.2. Hypothesis Ib: model validation through unisensory correlations

As an alternative model validation, we proposed to compare individual parameters extracted from the multisensory task through the BCI_{ST} model with their unisensory correspondents measured through independent tasks. Such tasks were designed to match the VPD setup, to capture the relevant unisensory components as accurately as possible. However, this independent measure necessarily implies some variations, and a strict 1 to 1 correspondence between the measured and the fitted parameters cannot be guaranteed. Still, a correlation between the measured unisensory and the fitted parameters is to be expected if the multisensory task relies on the proposed Bayesian CI process. Therefore, we assessed whether the parameters extracted from the unisensory tasks positively correlate with the ones extracted from the multisensory task by performing significance tests on Pearson correlation scores. Regarding proprioception, the correlation of the σ_p fitted by the model with the proprioceptive precision extracted from either of the two proprioceptive tasks (PJ and OL) was considered valid. The detailed extraction of unisensory parameters is described below (Section 2.6.5).

2.6.3. Hypothesis II: link with subjective ownership

Finally, we compared model predictions about P_{com} with subjective ratings about ownership. We fitted a bivariate Gaussian to the average subjective ratings as a function of spatial and temporal disparity, in order to extract the values of the tolerated spatial and temporal disparities as the standard deviation of the fitted Gaussian. Those values were correlated with the values obtained by performing the same fit on P_{com} values extracted from model fitting on the reaching task. To obtain P_{com} values for each subject, we simulated 5000 trials for each spatio-temporal disparity, using the parameters fitted from the VPD task, and take the average value of P_{com} , similarly to what done in Fig. 2d. Since, obviously, negative temporal disparities cannot be sampled, we mirrored the ratings and P_{com} values symmetrically with respect to the spatial axis, so to be able to perform the fits. We then performed significance tests on Pearson correlation scores. The Gaussian fit has six free parameters: the spatial and temporal standard deviation (our parameters of interest), the spatial and temporal means, a normalization constant and a global offset.

2.6.4. Interpretation of possible outcomes

Model validation (Hypothesis I) was considered achieved if we could demonstrate the presence of optimality by confirming Hypothesis Ia or Hypothesis Ib. Model selection (Hypothesis Ia) was considered successful if the BCI_{ST} outperformed all the alternative models (exceedance probability $>.95$ against FF, BCI_S , Heuristic, semi-FF). Regarding the unisensory correlations (Hypothesis Ib), a significant correlation, after Bonferroni correction (5 parameters, $p < .01$), between at least one of the unisensory estimates and the corresponding model parameter was considered sufficient. Indeed, despite our careful planning, it is possible that one or

more unisensory tasks do not measure the appropriate unisensory function involved in the putative Bayesian CI process taking place in the multisensory task. On the contrary, if the BCI_{ST} is not valid, the chances of a significant correlation between the independent estimates of a unisensory precision with the corresponding model parameter is extremely low. In any case, even if all of the unisensory tasks fail to capture the relevant unisensory component, model selection should still be able to demonstrate the validity of the proposed model, as it is performed purely on the multisensory task.

Hypothesis II was accepted if the correlation described in Section 2.6.3 was significant.

2.6.5. Extraction of unisensory parameters

Regarding OL task, where the source of variability in reaching movements is proprioceptive-motor noise, we expect:

$$x_r = x_t + N(0, \sigma_p) \quad (12)$$

where x_r denotes the reached position, and x_t the target position.

Then, in order to extract σ_p , we simply need to fit $x_t \sim x_r$, and extract the root mean square error of the fit. The same reasoning holds for the PJ task, by replacing reached positions with individual proprioceptive judgements. In order to control for possible active movements of the hand during the task, trials in which participants' hand show a displacement >5 mm [i.e., a value which takes into account the reported precision of the motion tracking system (Jost et al., 2019)] were discarded from the analysis.

Similarly, in our MJ task, we expect:

$$x_j = x_m + N(0, \sigma_v) \quad (13)$$

where x_j denotes the judged midline position, and x_m the true midline position. Then again, σ_v can be simply extracted as the root mean square error of midline judgements.

The analyses are slightly more complicated for the extraction of σ_t . Participants do not directly report judgements about timing, as this would be hard to do practically and possibly introduce cognitive biases, but instead they express judgements about simultaneity. In a Bayesian framework, this is best described as another causal inference process. The causal inference equations are very similar to the ones described for inferring P_{com} , the main difference being that they extend only to the temporal domain. P_{sim} , the inferred probability that the motor command and the observed movement are simultaneous given a perceived amount of delay is

Assuming that participants report stimuli to be simultaneous when P_{sim} is larger than a given threshold β , the expression can be used to predict the shape of the psychometric curve obtained in our simultaneity judgement. Then the value of σ_t can be recovered by fitting through a process similar to the one used in our multisensory task. It is important to note that, as verified by simulations, the fitted value for σ_t depends only on the slope of the psychometric curve and is not affected by the values of the prior on simultaneity or the response criterion (see [Supplementary material](#)).

Finally, since the VM task cannot be connected mathematically to the Bayesian CI model in a straightforward manner, we chose an empirical criterion for assessing its impact. We performed a logistic fit on the judgements expressed as a function of the percentage of morphing according to the following model.

$$\ln\left(\frac{p_y}{1-p_y}\right) = \beta_0 + \beta_1 x + \varepsilon \quad (15)$$

where p_y denotes the probability of replying “yes” in the VM task, x denotes the morphing percentage (with 0 meaning 0% self, and 100 meaning 100% self), β_0 and β_1 are fit parameters and ε is the error term. Then, the subjective equivalence point is given by $-\beta_0/\beta_1$, and the slope at that point by $\beta_1/4$.

We used such value of the slope as a proxy of accuracy in visually discriminating one's own hand. In principle, participants are expected to be stricter in embodying the virtual hand, which does not look like their own. We therefore expect a significant positive correlation between values of P_{π} and slopes of the psychometric function.

2.7. Timeline

Pending limitations deriving from the COVID-19 pandemic, data collection can be initiated immediately upon Stage 1 acceptance in principle, and the study can be completed within 4 months since then. Approximately one month will be needed for data collection, two months for data analysis and one month for writing and editing.

2.8. Pilot data

To confirm the possibility of translating the task developed by Fang and colleagues to immersive virtual reality, and test our model fitting procedure, we collected pilot data from 10 healthy participants (4 females, aged 24.1 ± 2.4 years, age range 21–29). The task was a close replication of Fang's task (including the spatial, but not the temporal manipulation of

$$P_{\text{sim}} = P(\tau_p = \tau_v | t_p, t_v) = \frac{P(t_p, t_v | \tau_p = \tau_v)P(\tau_p = \tau_v)}{P(t_p, t_v | \tau_p = \tau_v)P(\tau_p = \tau_v) + P(x_p, x_v, t_p, t_v | \tau_p \neq \tau_v)P(\tau_p \neq \tau_v)} = \frac{P_{\sigma_a} \frac{1}{\sigma_a} e^{-\frac{t^2}{2\sigma_a^2}}}{P_{\sigma_a} \frac{1}{\sigma_a} e^{-\frac{t^2}{2\sigma_a^2}} + (1 - P_{\sigma_a}) \frac{1}{3\sigma^2}} \quad (14)$$

visuo-proprioceptive disparity). Besides the experimental design in terms of spatial (angular) disparities, the task was the same as described in Section 2.4.1. For each participant, three experimental blocks with slightly different designs were collected. In the first two blocks, we collected a total of 49 trials with 7 disparities (at 0° , $\pm 13.3^\circ$, $\pm 26.6^\circ$ or $\pm 40^\circ$). Each disparity was presented 7 times, each time with a different target out of 7 equally spaced targets between -45 and 45° . Only 5 repetitions per disparity were collected in the third block, with 5 targets equally spaced on the same range. Target positions and spatial disparities were randomized within each block.

The results of the reaching task, shown in Fig. 5a, are qualitatively in line with what reported by Fang, confirming that the experimental setup can be successfully exported to an immersive virtual reality environment. We then tested whether our data were well modelled by a Bayesian CI model (BCI_S) by applying a fitting procedure very similar to the one proposed by Fang and colleagues to extract model parameters (see Analysis plan for details). As done in Fang (Fang et al., 2019), we then compared BCI_S model to the FF model predictions (fixed weights to vision and proprioception at all disparities). We used model Bayesian information criterion as an approximation of model evidence, and computed the model's exceedance probability (Wozny et al., 2010) (original code available from: <https://github.com/sjgshrm/mfit>). We found this analysis to favour the BCI_S model with an exceedance probability = .894, in line with the value reported by Fang (Fang et al., 2019) in his second experiment with 8 human participants (.954), showing that data from our VPD task is also quantitatively in line with this previous study. We then used the distribution of the extracted parameters as a basis for our power analysis.

We also performed a smaller pilot study on two healthy participants (2 males, aged 24 and 28 years) to test the

practical feasibility of combining both spatial and temporal disparities during the task, as this was never done before (Fig. 5b and c). The experimental design was exactly as described in the methods, except that, for temporal delays larger than zero, 6 disparities have been tested, uniformly distributed between -33.3° and 33.3° instead of $-40^\circ/40^\circ$. The design was changed as further simulations showed that this proposed design to be slightly better in terms of statistical power. Both participants were able to execute the task correctly, and the effect of temporal delay is present in both participants in line with our expectations. We used these pilot data to test our analysis pipeline and parameter extraction via model fitting. The fit converged for both participants and accurately modelled the data ($R^2 = .940$ and $.905$ respectively), yielding values of the parameters in line with our expectations (S01: $\sigma_v = 7.1$, $\sigma_p = 4.63$, $\sigma_t = .118$, $P_\pi = .938$, S02: $\sigma_v = 9.64$, $\sigma_p = 5.98$, $\sigma_t = .101$, $P_\pi = .826$).

3. Results

As planned, data from 40 healthy participants was collected for our study. Due to technical issues during data collection (failure of the morphing algorithm), data for the VM task was not collected for two participants. According to the pre-registered criteria, these participants were still included in the analyses involving all the other tasks.

3.1. Hypothesis I

Hypothesis I consisted in the validation of a Bayesian CI model describing participants' behaviour in a multisensory reaching task (VPD) under visuo-proprioceptive disparity as the result of the optimal integration of visual and proprioceptive

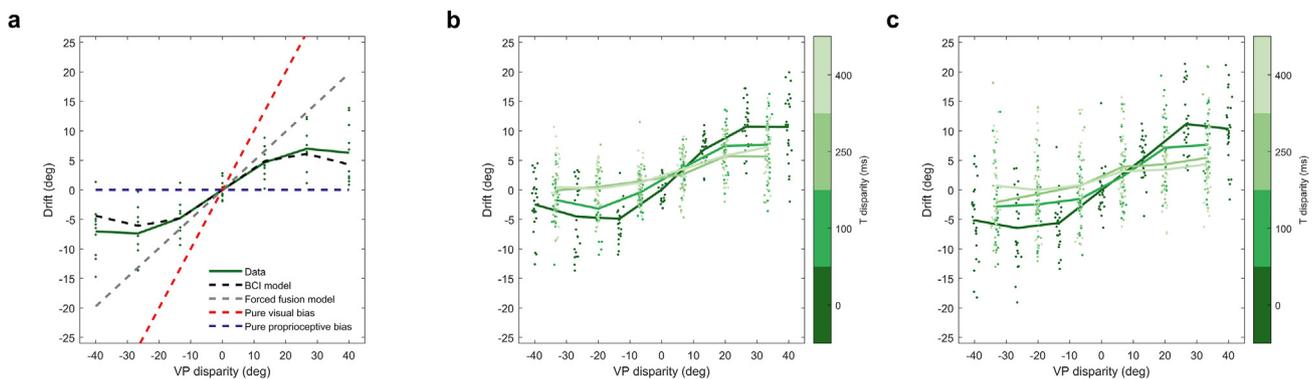


Fig. 5 – Pilot data for a purely spatial (a) and spatio-temporal (b, c) disparity setup. Panel (a) shows the results from a larger pilot study on 10 participants with the VPD task, with no temporal delay (as in Fang et al., 2019). The x axis indicates visuo-proprioceptive disparities, defined as the virtual hand angle minus the real hand angle (positive angles being on the right). The y axis indicates the proprioceptive drift defined as the target's angle minus the real hand's angle (a participant reaching left of a target experiences a proprioceptive drift towards the right, and vice versa). The blue and red dashed lines represent the expected drift in the case of a purely proprioceptive or visual dominance, respectively. The grey dashed line represents the predicted drift from a FF model of visual-proprioceptive integration, while the black dashed line shows the averaged predictions of the BCI_S model, in close agreement with averaged experimental results represented by the green solid line. Panels (b) and (c) show results for two pilot participants from the new spatio-temporal disparity setup. Solid lines represent conditional means, and the colours code represents the different temporal delays tested (T disparity; as in Fig. 1b). As expected, drift values increased at increasing temporal delays.

information in both the spatial and temporal domain (BCI_{ST}). We aimed at demonstrating our hypothesis through two approaches.

- (i) model selection, i.e., by comparing the BCI_{ST} model with four competing models (*Hypothesis 1a*).
- (ii) by showing that BCI_{ST} model predictions about unisensory precisions matched those measured through a set of dedicated tasks (*Hypothesis 1b*).

All models yielded good fits for reaching errors (R^2 BCI_{ST} : $.901 \pm .049$ SD; Heuristic: $.9 \pm .053$ SD; Semi-FF: $.902 \pm .052$ SD; FF: $.895 \pm .053$ SD). However, when computing the exceedance

probability (EP), the BCI_{ST} outperformed the FF ($BCI_{ST} = .999$, $FF = .001$), BCI_S ($BCI_{ST} = .984$, $BCI_S = .016$), and heuristic model ($BCI_{ST} = .984$, $Heuristic = .016$), but not the Semi-FF, which instead slightly outperformed the BCI_{ST} model ($BCI_{ST} = .04$, $Semi-FF = .96$). At the single subject level, amongst our 40 subjects, the BCI_{ST} model outperformed the FF in 35 participants, the BCI_S in 26 participants, the Heuristic in 26 participants and the Semi-FF in 16 participants. We show the average fits for all models, as well as for two exemplary participants, in Fig. 6.

Regarding the unisensory correlations (Fig. 7b), the fitted proprioceptive precision (σ_p) significantly correlated with the proprioceptive precision, as measured by both the PJ task ($r = .43$, $p = .005$, $p = .027$ after Bonferroni correction) and the

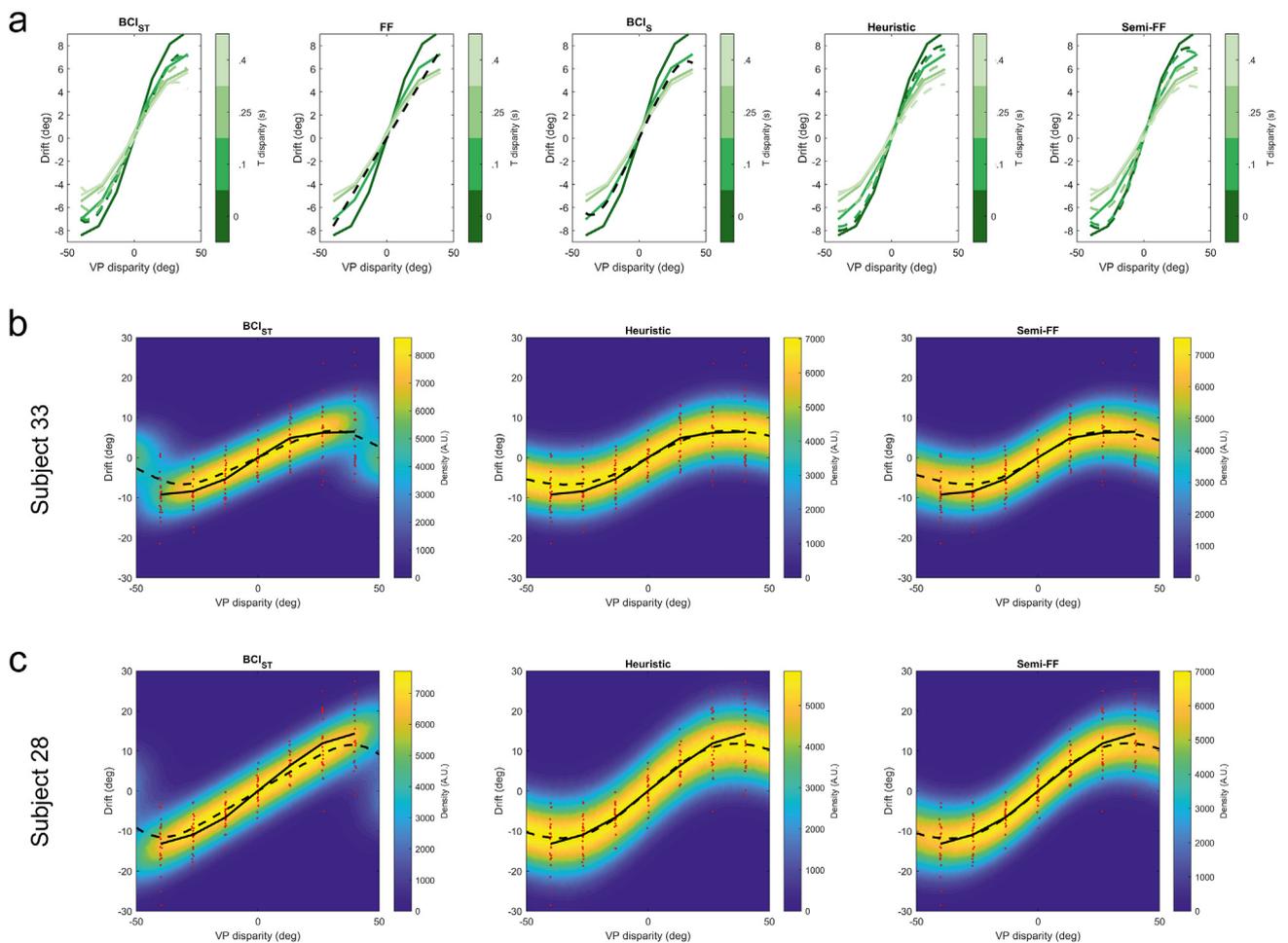


Fig. 6 – Model comparison. Panel (a) shows the average reaching errors in solid lines, and the corresponding model fits in dashed lines, for all the 5 models we compared. Note that the fitting is based on the full data distribution, taking into account not only the mean reaching error but the dispersion of individual trials around the mean. Hence, the best fits may not be the ones best approximating the mean, probably leading to small discrepancies between fitted and actual data. Panels (b) and (c) show the full data and model distribution for the three key models (BCI_{ST} , Heuristic, Semi-FF), for two exemplary subjects at zero temporal disparity. In the first (b), the BCI_{ST} outperformed all alternative models, in the second (c), it outperformed the heuristic but not the Semi-FF. The fitted data density is encoded in the colour, and its average in the dashed black line. The solid black line represents the data average and red dots individual trials. As visible in the plots, the BCI_{ST} and Semi-FF model mainly differ in the transition between low-disparity visual dominance and high-disparity proprioceptive dominance, which is much sharper in the and BCI_{ST} and leads to an almost bimodal distribution (see also Fig. 3d). In participants in which these subtle differences are not observable, the Semi-FF model ends up outperforming the BCI_{ST} due to the greater flexibility granted by its fully independent parameters.

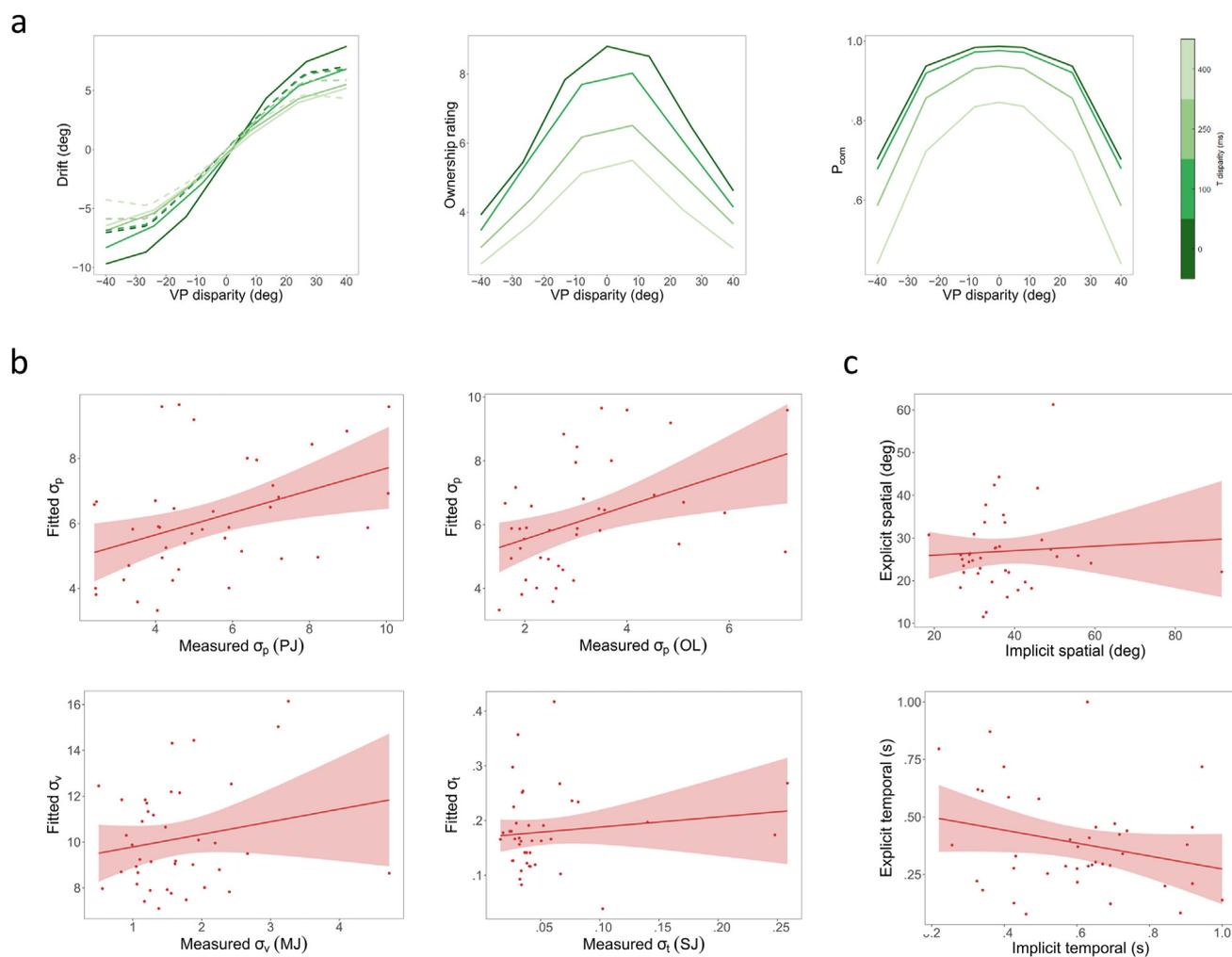


Fig. 7 – Pre-registered analyses. (a) Shows the average proprioceptive drift (left plot), subjective ownership ratings (centre plot), and estimated P_{com} (left plot) for the VPD task in 40 participants. The colour code represents the different temporal delays tested. Dashed lines represent model fits. **(b)** Shows the correlation between the unisensory precisions fitted by the BCI_{ST} model from reaching errors at the multisensory task (VPD) and those measured by the unisensory tasks, with each plot representing a different unisensory task. **(c)** Shows the correlation between the implicit (as extracted from reaching errors) and explicit (as extracted from ownership ratings) window for ownership. The top plot represents the spatial window, the bottom plot the temporal window.

OL ($r = .43$, $p = .006$, $p = .031$ after Bonferroni correction); in contrast, neither the fitted visual (σ_v) nor the fitted temporal (σ_t) precisions correlated with the corresponding unisensory precisions measured by the MJ ($r = .26$, $p = .1$ uncorrected) and SJ task ($r = .18$, $p = .26$ uncorrected) respectively. No significant correlation was found between P_π and the ability to discriminate the higher order visual features of the hand, as measured by the VM task ($r = .041$, $p = .8$ uncorrected). These results were replicated with a different choice of the flat priors in the BCI_{ST} model (see [Supplementary material](#)).

In summary, [Hypothesis 1b](#) can be considered satisfied, and, according to the pre-registered analysis plan, [Hypothesis 1](#) can be therefore accepted despite [Hypothesis 1a](#) not being fully satisfied.

3.2. Hypothesis II

With [Hypothesis II](#), we aimed at demonstrating that reaching errors in the VPD task, as modelled by our BCI model, truly reflect the subjective experience of body ownership. To do so, we extracted from reaching errors the probability of ownership as a function of both spatial and temporal disparity and computed the spatio-temporal extent of the tolerated window for implicit ownership by fitting a Gaussian function to P_{com} values. We hypothesized that the size of such implicit ownership window would positively correlate with the size of the explicit ownership window, extracted through the same method from explicit ownership ratings.

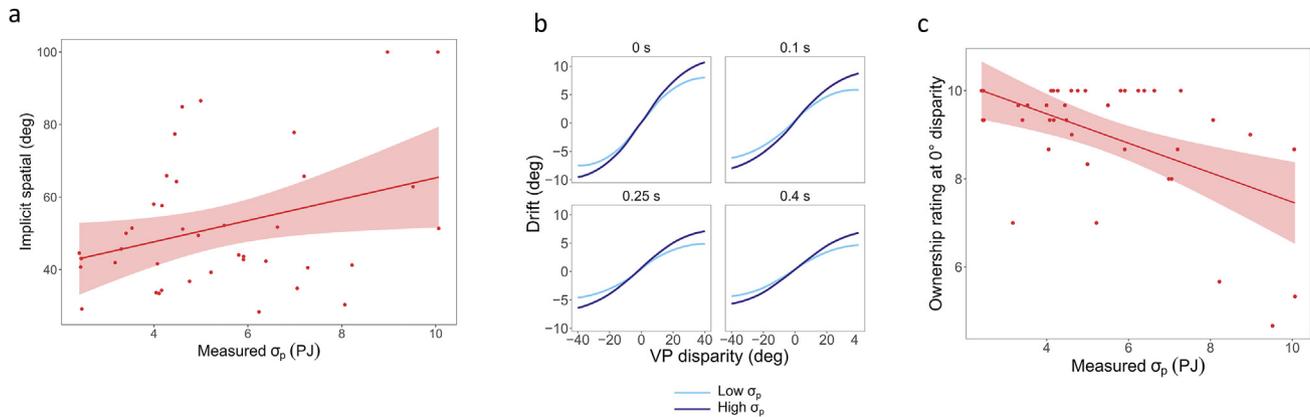


Fig. 8 – Proprioceptive precision and body ownership. Panel (a) shows the correlation between the implicit window for ownership in the spatial domain and the unisensory proprioceptive precision extracted from the PJ task. The plots in panel (b) show the average proprioceptive drift in each of the tested delays for lower (dark blue) and higher (light blue) proprioceptive precision measured by the PJ task. Panel (c) shows the correlation between average ownership ratings at zero disparity and proprioceptive precision (PJ).

We found that neither the size of the spatial, nor the one of the temporal windows for implicit ownership were correlated with their explicit counterpart (respectively: $R = .07$, $p = .67$; $R = -.26$, $p = .098$, Fig. 7c). Therefore, Hypothesis II cannot be considered to be satisfied according to the pre-registered analysis plan.

However, when investigating in detail the different steps of the analysis with empirical data, we found a computational issue in the parameters used for Gaussian fits on P_{com} values, which yielded noisy fitted values (see 6.3.2 for details). We hypothesized that such fitting instability might explain the observed lack of the correlation predicted by Hypothesis II (see Supplementary material for details). Therefore, we performed the conceptual replication of Hypothesis II after resolving such technical issue, as detailed in the exploratory analyses.

3.3. Exploratory analyses

3.3.1. Relationship between unisensory precisions and body ownership

Model selection showed that the of BCI_{ST} performed worse than the Semi-FF (see Section 3.1). Conceptually, the main difference between these two models is that the first assumes the spatial and temporal binding windows to be optimised on unisensory precisions, while in the latter these are considered independent. To assess more directly the existence of a relationship between unisensory precisions and the ownership window, we investigated the correlation between the width of P_{com} extracted by the BCI_{ST} in both the spatial and temporal dimensions and unisensory precisions extracted from the unisensory tasks. We found a positive correlation between σ_p measured by the PJ task and $\sigma_{p_{com}}$ in the spatial domain ($R = .33$, $p = .035$, Fig. 8a). This is directly visible by splitting reaching errors as a function of disparity for between high and low σ_p subjects (Fig. 8b). In contrast, $\sigma_{p_{com}}$ did not correlate with

the σ_p measured by the OL in the spatial domain ($R = .13$, $p = .43$) nor with σ_t measured by the SJ in the temporal domain ($R = .043$, $p = .79$).

Another interesting model prediction is that natural hand ownership, which in our setup is simulated by ratings at zero spatio-temporal disparity, should negatively correlate with proprioceptive precision. This may be especially relevant to the study of ownership alterations in patients with somatosensory deficits. Indeed, a less precise proprioception leads to higher visuo-proprioceptive incongruences in neural encoding, and less certainty that the hand is one's own even when the underlying physical stimuli are aligned. This was indeed the case in our data, where a significant negative correlation between σ_p (PJ) and ownership ratings at zero disparity was observed ($R = .52$, $p = .0006$, Fig. 8c).

3.3.2. Hypothesis II analyses with improved fit quality

As mentioned in Section 3.2, we identified a technical issue leading to poor quality of the Gaussian fits on P_{com} values, which is a crucial step for the validation of Hypothesis II. Essentially, the issue derived from the presence of three unnecessary free parameters in the fit of P_{com} values: the spatial and temporal location of the peak of the Gaussian, and a constant offset added to the Gaussian function. These parameters are necessary to fit explicit ratings, which could in principle peak at non-zero spatio-temporal disparity and may not reach zero even at very large disparities due to inter-individual differences. For simplicity, they were left free also when fitting P_{com} values, although the model used to compute the P_{com} values in the first place imposes *per se* these parameters to be zero. As detailed in the Supplementary material, this led to extremely poor quality of P_{com} fits, as in several subjects the fitted values largely varied when repeating the fits with different starting parameters and could be therefore considered almost meaningless. The issue is completely resolved by fixing the unnecessary free parameters to zero. In

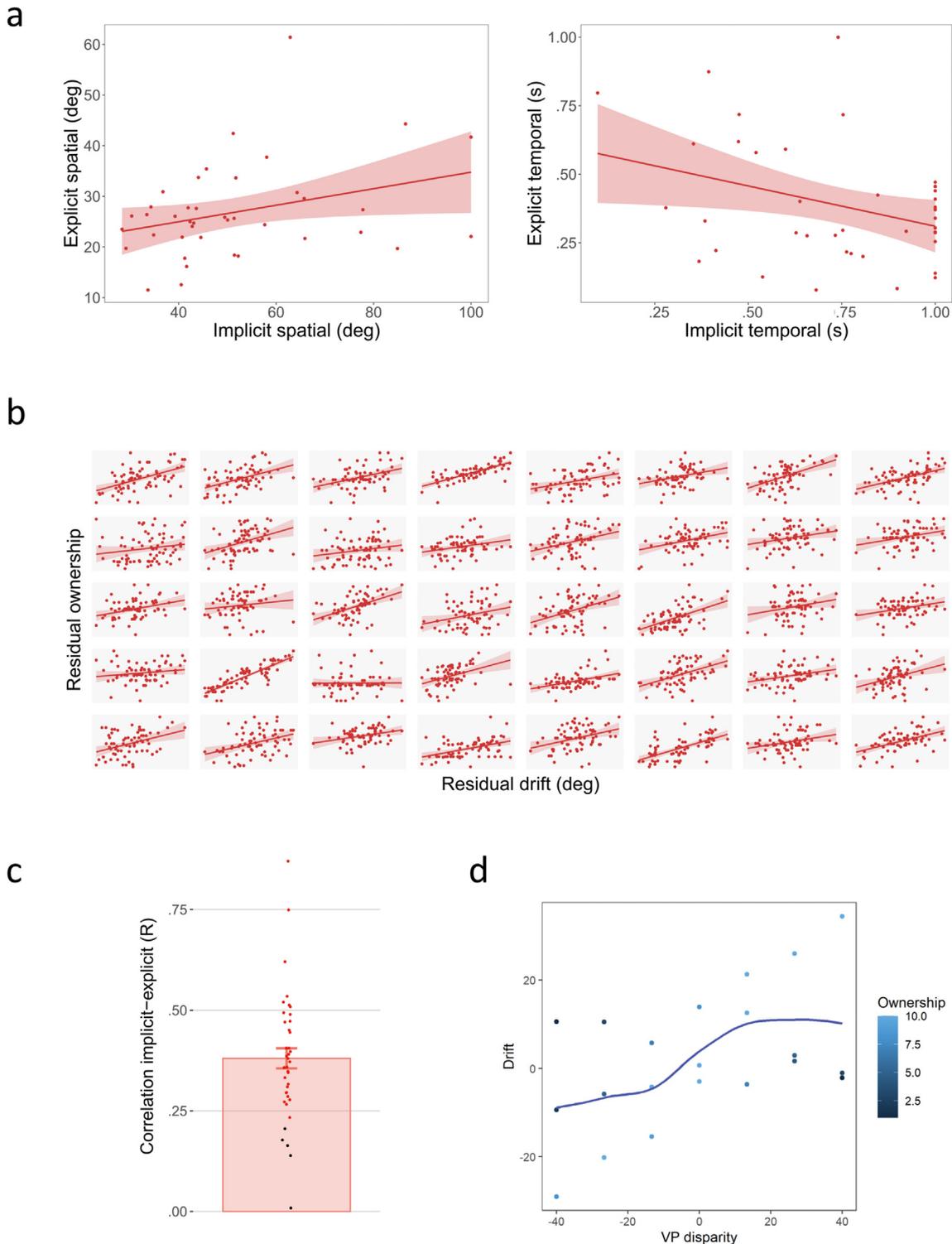


Fig. 9 – Hypothesis II revised. Panel (a) shows the correlation between the implicit and explicit window for ownership, after resolving fitting issues. The left plot represents the spatial window, the right plot the temporal window. Panel (b) shows the correlation between residual drift and residual ownership measures, with each subplot corresponding to one of the 40 participants. Panel (c) shows individual Pearson correlation coefficients between implicit and explicit ownership ratings. Subjects in red show a significant correlation at $p < .05$. Panel (d) illustrates the result on raw data in a paradigmatic subject. The solid blue line indicates the average drift for each disparity and dots the individual reaching errors. Therefore, trials between the line and zero have negative residual drift (more proprioceptive weight), and vice versa. The colour of the dots indicates the ownership rating (1–10) associated to that trial. Higher ownership ratings are associated with more weight attributed to the virtual hand.

addition, we discovered that the fitting quality could be further improved by increasing the iterations of the BCI_{ST} fit to 100,000 (see [Supplementary material](#)).

We performed further analyses after implementing such improvements in the fitting procedures. First, we confirmed that all results from [Hypothesis I](#) are left virtually unchanged by the modifications (see [Supplementary material](#)). Then, we tested whether the failure to confirm [Hypothesis II](#) could be determined mainly by such computational issue, by replicating the key correlation analysis, using the improved fits as an input. Indeed, since the modifications to the analysis procedure only improve the numerical quality of the fit, the conceptual and statistical meaning of the analysis is left unchanged.

We found that, in the spatial domain, there was a significant ($R = .32$, $p = .041$, [Fig. 9a](#), left) and positive correlation between the explicit and implicit window for ownership, as expected according to [Hypothesis I](#). Therefore, participants who integrated the virtual hand in their reaching movement up to larger spatial disparities also reported higher levels of ownership up to larger spatial disparities. Surprisingly, we found a significant negative correlation ($R = -.35$, $p = .029$, [Fig. 9a](#), right) in the temporal domain. This result is indeed puzzling and counterintuitive, and may depend on computational artifacts in the fitting, or other uncontrollable factors. Indeed, the overall effect of the temporal manipulation on reaching errors is small compared to the one of the spatial manipulation, leading to less reliable estimates of the dependence of P_{com} on temporal disparities. This is reflected by the large and unrealistic fitted values for the temporal window, as well as by the lack of correlation between fitted and measured σ_t . Therefore, interpretation of the negative correlation should be approached with caution.

3.3.3. Trial-by-trial correlation between ownership and reaching errors

Since both pre-registered and exploratory analyses addressing [Hypothesis II](#) yielded mixed results, we decided to further explore the link between subjective ratings at the individual trial level, which was not possible with the experimental design used by Fang et al., who collected explicit ownership ratings only by blocks. This way, we aimed to test the conceptual question of [Hypothesis II](#) in a way that is less affected by computational issues of numerical stability and can benefit from the higher statistical power of conducting analyses at the individual trial level. If reaching errors truly reflect the subjective ownership feeling at the implicit level, we would expect that trials with higher explicit ownership would be associated with a higher weight attributed to the virtual hand even at fixed spatio-temporal disparity. To test this hypothesis, we computed “residual” explicit ownership ratings at the individual level, by subtracting to each rating the average ownership rating at the same spatio-temporal disparity. Similarly, residual drift (i.e.: putative implicit ownership) values were obtained by subtracting to the reaching error its average value for each disparity. In order for positive residual drift values to always

indicate a larger visual weight and vice versa, residual drift values were multiplied by the sign of the spatial disparity. Zero disparity values were excluded as they yield no meaningful information in this analysis. Then, we tested whether residual drift and ownership values were correlated by means of a linear mixed model. Regardless of the structure of random effects used, we found a strong effect of residual drift values on residual ownership values ($p < .001$). Indeed, a correlation between residual ownership and drift values can be observed in almost all subjects individually, with 31 out of 40 participants showing a significant correlation at $p < .05$ (see [Fig. 9b–d](#)). The average correlation coefficient was $.381 \pm .032$ S.E., significantly larger than 0 at $p < .001$. This result strongly suggests that individual reaching errors reflect subjective ownership for the virtual hand as reported by the participant within the same trial.

4. Discussion

In this work, we aimed at rigorously assessing whether the emergence of body ownership from spatial and temporal visuo-proprioceptive congruency carries the signature of optimal Bayesian inference. To do so, we measured the proprioceptive drift induced in reaching movements by a spatio-temporally incongruent virtual hand and compared it to predictions of a Bayesian Causal Inference (BCI_{ST}) model. The validation of the model developed over two pre-registered hypotheses. With [Hypothesis I](#), we aimed at demonstrating that the model not only well fits experimental data, but that the data shows compelling evidence of Bayesian processing. This was done by comparing the model with appropriate non-Bayesian alternative models and by directly comparing the unisensory precisions fitted from the multisensory reaching task (VPD) to independent measures of unisensory precisions. With [Hypothesis II](#), we aimed at demonstrating that the reaching errors induced by the virtual hand can be considered as an index of implicit ownership and reflect subjective ratings of ownership (explicit ownership), by assessing the relationship between these two measures.

[Hypothesis I](#) was accepted according to the pre-registered analysis plan. A significant correlation was observed between the proprioceptive precision fitted from the VPD task and measured through unisensory static (PJ) and dynamic (OL) task. While there is a significant overlap between the VPD and OL task (both involving active reaching movements), the static PJ task is orthogonal to the VPD task and provides a more compelling signature of Bayesian optimality. No significant correlation was found between measured and fitted temporal precision (σ_t), visual precision (σ_v) and common cause prior P_π . Model comparison also favoured the BCI_{ST} model, which outperformed all but one of the alternative models, as we will discuss in more detail later.

[Hypothesis II](#), instead, was not validated according to our pre-registered analysis plan, as no significant correlation was observed between sizes of the spatial or temporal windows for implicit and explicit ownership. Nevertheless, detailed investigation of the fitting procedures involved in the analysis

with the full sample of empirical data revealed a purely computational issue which affected the quality of the results. After resolving it, a significant positive correlation emerged between the sizes of the spatial windows for implicit and explicit ownership, confirming the results by Fang and colleagues (Fang et al., 2019). Surprisingly, a negative and significant correlation emerged in the temporal domain. Since the temporal manipulation was only partially successful and led to noisy estimates of σ_t , we believe this result may be due to numerical artifacts and should be taken with caution. To further investigate the question raised in Hypothesis II, we performed an alternative exploratory analysis at the single trial level, revealing a strong correlation between variability in ownership ratings and reaching errors. Trials in which the virtual hand biased the reaching movement more strongly than on average were associated with ownership ratings higher than the average. Such correlation was significant not only at the group level, but also at the single subject level for 31 out of 40 (77.5%) participants.

In sum, our results suggest that visuo-proprioceptive integration and the sense of body ownership both carry the signature of Bayesian, uncertainty-based processing, and can be well described by a Bayesian Causal Inference model.

It is worth noticing that the analysis plan was thought to detect any signature of Bayesian processing, with different alternative outcomes being considered sufficient for validating Hypothesis I. This is already an important advancement with respect to most previous studies, whose experimental design did not allow to rigorously demonstrate such feature. Still, not all experimental results were in line with the initial hypotheses, strictly following all the pre-registered analyses. Since unsuccessful analyses can be as informative as successful ones, we will discuss these limitations extensively, in line with the philosophy of pre-registered reports.

First, three out of the five expected correlations between the fitted and measured parameters were not significant. The simplest explanation is that the unisensory tasks do not capture model parameters or do so with insufficient precision to reach the necessary statistical power. In our view, this is the most likely explanation for the absence of correlation between the fitted and measured σ_v and P_π . Indeed, although both parameters routinely appear in similar studies (Fang et al., 2019; Körding et al., 2007; Samad et al., 2015), they are conceptually difficult to measure independently, which is arguably a common limitation in Bayesian models of brain function. In contrast, the definition of temporal precision within the context of our experimental task seems less ambiguous, and the absence of correlation is more surprising. Indeed, the comparison with the purely spatial BCI model shows that the temporal manipulation does affect behaviour. However, such modulation is weaker than expected (likely also affecting the stability of the fit of the underlying parameter, see Fig. S6) and not coherent with model predictions. Possibly, this is due to the fact that the brain may optimise behaviour in a task-dependent way (Limanowski & Friston, 2020), an aspect which was not included in our model. Since the focus of the task is to maximise spatial accuracy, while attentional and computational resources are limited, spatial parameters may be optimised more accurately than temporal ones.

Another key aspect in Hypothesis I is that model selection showed that the BCI_{ST} model is outperformed by the Semi-FF model, in which unisensory precisions determine the relative weight of vision and proprioception, but not the size of the spatial and temporal windows of integration. This may seem surprising, and contradictory to the presence of correlations between measured and fitted unisensory precisions. Nevertheless, it is important to notice that the Semi-FF model was constructed post-hoc with the sole purpose of providing the best fit of already known empirical data. This way, it benefits from the absence of any theoretical constraint, but for the same reason it lacks any deeper insight into the underlying mechanism. More conceptually, Bayesian approaches to brain function start from the hypothesis that the brain approximates optimal inference, and arguably share the limitation that they do not provide predictions about how good the approximation will be. Whether an ad-hoc model can outperform a Bayesian model is therefore more an empirical question than a conceptual one. Thus, while this does not hinder the Bayesian brain hypothesis conceptually, it may limit its practical utility. On the one hand, the Semi-FF model cannot account for the observed correlations between unisensory and fitted parameters, suggesting that it does not provide the most exhaustive description of brain function. On the other hand, the fact that a non-Bayesian model can outperform a reasonably sophisticated Bayesian model confirms the difficulty of accurately reproducing subtle aspects of behaviour within the Bayesian framework (as those visible in Figs. 3d and 6b–c). This confirms the need of refining current models through rigorously paired theoretical and experimental research.

Coming back to Hypothesis II, taken together our results suggest that reaching errors induced by visuo-proprioceptive disparity do reflect subjective ownership feelings. In this sense, the last exploratory analysis (shown in Fig. 9) is especially interesting, as it shows a strong connection between random fluctuations of proprioceptive drift and subjective ownership. This is perfectly in line with a Bayesian account of body ownership, in which noisy and constantly fluctuating sensory information leads to subjective ownership through an online inference process. Such analysis is also relevant for the recent debate about demand characteristics of tasks used to assess body ownership (Lush et al., 2020). While it is possible to imagine that subjects may provide lower ownership ratings in the presence of incongruence, due to demand characteristics, our finding directly link the residual reaching errors and ownership ratings, after removing the average main effect of disparity. Such quantities require a complex mathematical computation to be obtained and are therefore cognitively inaccessible and arguably immune to demand characteristics. Furthermore, compared to explicit paradigms such as the RHI, our VR adaptation of Fang's task allows to easily collect a greater amount of quantitative data points while parametrically exploring several modulating factors.

Our paradigm does not directly apply the causal inference framework to subjective ownership ratings. This is instead the approach used by a recent work (Chancel et al., 2022). Their main finding is that trial-by-trial variations of noise in visual information are taken into account in rating ownership in a rubber hand setup, since the tolerance to temporal delays in visuo-tactile stimulation increases with levels of sensory

noise rather than being constant. This is a strong hint that uncertainty-based inference is taking place. However, these results cannot be taken as a quantitative demonstration of causal inference in body ownership, since the effect of sensory noise on unisensory precision is not independently quantified [as e.g., in Ernst & Banks (2002b)] and could not be compared to model fits.

In sum, we provided novel evidence that body ownership arises from visuo-proprioceptive integration through a process that can be modelled in a quantitative framework, whose main ingredients are the relative precisions of unisensory inputs. Crucially, we developed methods which are effective in independently measuring these key parameters of the model. These tasks can be especially useful in explaining abnormalities in the sense of body ownership in clinical populations, such as patients with acquired brain lesions (Garbarini et al., 2013, 2014; Romano & Maravita, 2019), or within the autistic (Cascio et al., 2012; Paton et al., 2012) and schizophrenic spectrum (Peled et al., 2000). In this sense, the correlation between zero-disparity ownership, a proxy of ownership for the real hand, and proprioceptive precision, provides a hint that natural hand ownership and the quality of (potentially deteriorated) proprioceptive inputs may be related. Previous studies attributed body ownership alterations to different steps of the multisensory integration process, represented by distinct parameters of the causal inference model. Apparent inconsistencies in some of these results (Buma et al., 2010; Burin et al., 2015; Germine et al., 2013; Llorens et al., 2017; Shaqiri et al., 2018; White & Aimola Davies, 2017) may be better investigated in the light of our quantitative framework, where the independent assessment of model parameters can help disentangle the unisensory and multisensory contributions to alterations in body ownership. On the theoretical side, our results show how body ownership, a key component of bodily self-consciousness, emerges from the optimal Bayesian integration of unisensory inputs, linking a higher order cognitive function to a well-established quantitative framework. Therefore, they provide a precious insight in how self-awareness may emerge from the evolutionary need to optimally combine sensory inputs relating the external world and our physical body.

Data availability plan

Compiled code for the experimental paradigms, custom analysis code, model fit outputs and all the data collected for this study is available from: https://osf.io/azh8p/?view_only=abdd1b1d33c24794b048847432374720. For further information on the code contact the first author (TB). The stage 1 accepted protocol can be found at: <https://osf.io/bzyge>.

Ethical approval plan

Participants were asked to sign an informed consent form prior to starting the experiment. All experimental procedures have been approved by the Ethical Committee of Human Research of the Vaud canton (CER-VD, project identifier: 2017-01588), Switzerland, and was run in accordance with the

ethical guidelines of the ethical committee and the Declaration of Helsinki. Participants were recruited using the online platform SonaSystem of the University of Lausanne (<https://epflunil.sona-systems.com>) and compensated 20 Swiss Francs per hour for their time.

Credit statement

Tommaso Bertoni: Conceptualization, Methodology, Formal analysis, Investigation, Writing – Original Draft, Visualization.

Giulio Matria: Conceptualization, Methodology, Formal analysis, Investigation, Writing – Original Draft, Visualization.

Nikita Akulenko: Investigation, Formal analysis, Visualization.

Henri Perrin: Investigation.

Boris Zbinden: Software, Resources.

Michela Bassolino: Writing – Review & Editing.

Andrea Serino: Conceptualization, Supervision, Writing – Review & Editing, Funding acquisition.

Open practices

The study in this article earned Open Data, Open Material and Preregistered badges for transparent practices. The data and material used for this study are available at: https://osf.io/azh8p/?view_only=abdd1b1d33c24794b048847432374720 and preregistered study available at: <https://osf.io/bzyge>.

Acknowledgements

This work was supported by a Swiss National Science Foundation Professorship grant, grant number 163951, <http://www.snf.ch>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cortex.2023.06.019>.

REFERENCES

- Abdulkarim, Z., & Ehrsson, H. H. (2016). No causal link between changes in hand position sense and feeling of limb ownership in the rubber hand illusion. *Attention, Perception & Psychophysics*, 78.
- Acerbi, L., & Ma, W. J. (2017). Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. *Advances in Neural Information Processing Systems*, 30, 1834–1844.
- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, 14, 257–262.

- Apps, M. A. J., & Tsakiris, M. (2014). The free-energy self: A predictive coding account of self-recognition. *Neuroscience and Biobehavioral Reviews*, 41, 85–97.
- Armel, K. C., & Ramachandran, V. S. (2003). Projecting sensations to external objects: Evidence from skin conductance response. *Proceedings of the Royal Society B: Biological Sciences*, 270, 1499–1506.
- Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, 13, 7–13.
- Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron*, 88, 145–166.
- Botvinick, M., & Cohen, J. (1998). Rubber hands ‘feel’ touch that eyes see. *Nature*, 391, 756–756.
- Buma, F. E., Lindeman, E., Ramsey, N. F., & Kwakkel, G. (2010). Functional neuroimaging studies of early upper limb recovery after stroke: A systematic review of the literature. *Neurorehabilitation and Neural Repair*, 24, 589–608.
- Buneo, C. A., Jarvis, M. R., Batista, A. P., & Andersen, R. A. (2002). Direct visuomotor transformations for reaching. *Nature*, 416, 632–636.
- Burin, D., Livelli, A., Garbarini, F., Fossataro, C., Folegatti, A., Gindri, P., & Pia, L. (2015). Are Movements Necessary for the Sense of Body Ownership? Evidence from the Rubber Hand Illusion in Pure Hemiplegic Patients. *PLOS ONE*, 10(3), e0117155.
- Butler, J. S., Smith, S. T., Campos, J. L., & Bulthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *Journal of Visualization*, 10, 23–23.
- Cascio, C. J., Foss-Feig, J. H., Burnette, C. P., Heacock, J. L., & Cosby, A. A. (2012). The rubber hand illusion in children with autism spectrum disorders: Delayed influence of combined tactile and visual input on proprioception. *Autism: the International Journal of Research and Practice*, 16, 406–419.
- Chancel, M., Ehrsson, H. H., & Ma, W. J. (2022). Uncertainty-based inference of a common cause for body ownership. *eLife*, 11.
- Cuppini, C., Stein, B. E., & Rowland, B. A. (2018). Development of the mechanisms governing midbrain multisensory integration. *Journal of Neuroscience*, 38.
- Della Gatta, F., Garbarini, F., Puglisi, G., Leonetti, A., Berti, A., & Borroni, P. (2016). Decreased motor cortex excitability mirrors own hand disembodiment during the rubber hand illusion. *eLife*, 5.
- Desmurget, M., Vindras, P., Gréa, H., Viviani, P., & Grafton, S. T. (2000). Proprioception does not quickly drift during visual occlusion. *Experimental Brain Research*, 134, 363–377.
- Dukelow, S. P., Herter, T. M., Moore, K. D., Demers, M. J., Glasgow, J. I., Bagg, S. D., ... Scott, S. H. (2010). Quantitative assessment of limb position sense following stroke. *Neurorehab Neural Repair*, 24(2), 178–187.
- Ehrsson, H. H. (2005). Touching a rubber hand: Feeling of body ownership is associated with activity in multisensory brain areas. *Journal of Neuroscience*, 25, 10564–10573.
- Ernst, M. O., & Banks, M. S. (2002a). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433.
- Ernst, M. O., & Banks, M. S. (2002b). Humans integrate visual and haptic information in a. *Nature*, 415, 429–433.
- Fang, W., Li, J., Qi, G., Li, S., Sigman, M., & Wang, L. (2019). Statistical inference of body representation in the macaque brain. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 20151–20157.
- Farrer, C., Bouchereau, M., Jeannerod, M., & Franck, N. (2008). Effect of distorted visual feedback on the sense of agency. *Behavioural Neurology*, 19, 53–57.
- Garbarini, F., Fornia, L., Fossataro, C., Pia, L., Gindri, P., & Berti, A. (2014). Embodiment of others’ hands elicits arousal responses similar to one’s own hands. *Current Biology*, 24, R738–R739.
- Garbarini, F., Pia, L., Piedimonte, A., Rabuffetti, M., Gindri, P., & Berti, A. (2013). Embodiment of an alien hand interferes with intact-hand movements. *Current Biology*, 23, R57–R58.
- Germine, L., Benson, T. L., Cohen, F., & Hooker, C. I. (2013). Psychosis-proneness and the rubber hand illusion of body ownership. *Psychiatry Research*, 207, 45–52.
- Haggard, P., Newman, C., Blundell, J., & Andrew, H. (2000). The perceived position of the hand in space. *Percept Psychophys*, 62, 363–377.
- James, W. (1890). *The Principles of Psychology* (Vol. 1). New York: Henry Holt.
- Jones, S. A. H., Cressman, E. K., & Henriques, D. Y. P. (2010). Proprioceptive localization of the left and right hands. *Experimental Brain Research*, 204, 373–383.
- Jost, T. A., Nelson, B., & Rylander, J. (2019). Quantitative analysis of the Oculus Rift S in controlled movement. *Disability and Rehabilitation. Assistive Technology*, 1–5.
- Kilteni, K., Maselli, A., Kording, K. P., & Slater, M. (2015). Over my fake body: Body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in Human Neuroscience*, 9.
- Kniestedt, C., & Stamper, R. (2003). Visual acuity and its measurement. *Ophthalmology Clinics of North America*, 16, 155–170.
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS One*, 2, Article e943.
- Krakauer, J. W. (2009). *Motor learning and consolidation: The case of visuomotor rotation* (pp. 405–421).
- Liao, J., Yao, Y., Yuan, L., Hua, G., & Kang, S. B. (2017). *Visual attribute transfer through deep image analogy*.
- Limanowski, J., & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience*, 7, 547.
- Limanowski, J., & Friston, K. (2020). Attentional modulation of vision versus proprioception during action. *Cerebral Cortex*, 30.
- Litwin, P. (2020). Extending Bayesian models of the rubber hand illusion. *Multisensory Research*, 33, 127–160.
- Llorens, R., Borrego, A., Palomo, P., Cebolla, A., Noé, E. I., Badia, S. B., & Baños, R. (2017). Body schema plasticity after stroke: Subjective and neurophysiological correlates of the rubber hand illusion. *Neuropsychologia*, 96, 61–69.
- Longo, M. R., & Haggard, P. (2010). An implicit body representation underlying human position sense. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 11727–11732.
- Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition*, 107, 978–998.
- Lush, P., Botan, V., Scott, R. B., Seth, A. K., Ward, J., & Dienes, Z. (2020). Trait phenomenological control predicts experience of mirror synaesthesia and the rubber hand illusion. *Nature Communications*, 11, 4853.
- Makin, T. R., Holmes, N. P., & Ehrsson, H. H. (2008). On the other hand: Dummy hands and peripersonal space. *Behavioural Brain Research*, 191, 1–10.
- Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J., & Friston, K. J. (2014). Bayesian inferences about the self (and others): A review. *Consciousness and Cognition*, 25, 67–76.
- Noel, J. P., Blanke, O., & Serino, A. (2018). From multisensory integration in peripersonal space to bodily self-consciousness: From statistical regularities to statistical inference. *Annals of the New York Academy of Sciences*, 1426, 146–165.
- Paton, B., Hohwy, J., & Enticott, P. G. (2012). The rubber hand illusion reveals proprioceptive and sensorimotor differences in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 42, 1870–1883.

- Peled, A., Ritsner, M., Hirschmann, S., Geva, A. B., & Modai, I. (2000). Touch feel illusion in schizophrenic patients. *Biological Psychiatry*, 48, 1105–1108.
- Rincon-Gonzalez, L., Buneo, C. A., & Helms Tillery, S. I. (2011). The proprioceptive map of the arm is systematic and stable, but idiosyncratic. *PLoS One*, 6, Article e25214.
- Rohde, M., Di Luca, M., & Ernst, M. O. (2011). The rubber hand illusion: Feeling of ownership and proprioceptive drift do not go hand in hand. *PLoS One*, 6.
- Rohde, M., Van Dam, L. C. J., & Ernst, M. O. (2016). Statistically optimal multisensory cue integration: A practical tutorial. *Multisensory Research*, 29, 279–317.
- Rohe, T., & Noppeney, U. (2015). Cortical hierarchies perform Bayesian causal inference in multisensory perception. *PLoS Biol*, 13, Article e1002073.
- Romano, D., & Maravita, A. (2019). The dynamic nature of the sense of ownership after brain injury. Clues from asomatognosia and somatoparaphrenia. *Neuropsychologia*, 132, 107119.
- Samad, M., Chung, A. J., & Shams, L. (2015). Perception of body ownership is driven by Bayesian sensory inference. *PLoS One*, 10, 1–23.
- Serino, A., Alsmith, A., Costantini, M., Mandrigin, A., Tajadura-Jimenez, A., & Lopez, C. (2013). Bodily ownership and self-location: Components of bodily self-consciousness. *Consciousness and Cognition*, 22, 1239–1252.
- Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17, 565–573.
- Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371, 20160007.
- Seth, A. K., & Tsakiris, M. (2018). Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences*, 1–13.
- Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in Cognitive Sciences*, 14, 425–432.
- Shaqiri, A., Roinishvili, M., Kaliuzhna, M., Favrod, O., Chkonia, E., Herzog, M. H., ... Salomon, R. (2018). Rethinking body ownership in Schizophrenia: Experimental and meta-analytical approaches show no evidence for deficits. *Schizophrenia Bulletin*, 44, 643–652.
- Spitzley, K. A., & Karduna, A. R. (2019). Feasibility of using a fully immersive virtual reality system for kinematic data collection. *Journal of Biomechanics*, 87, 172–176.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 9.
- Stein, B. E., Stanford, T. R., & Rowland, B. A. (2014). Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, 15.
- Thurman, S. M., & Lu, H. (2014). Bayesian integration of position and orientation cues in perception of biological and non-biological forms. *Frontiers in Human Neuroscience*, 8, 1–13.
- Tsakiris, M. (2010). My body in the brain: A neurocognitive model of body-ownership. *Neuropsychologia*, 48, 703–712.
- Tsakiris, M. (2017). The multisensory basis of the self: From body to identity to others. *Quarterly Journal of Experimental Psychology*, 70, 597–609.
- Tsakiris, M., 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. *Experimental Brain Research*, 204, 343–352.
- Tsakiris, M., & Haggard, P. (2005). The Rubber Hand Illusion Revisited: Visuotactile Integration and Self-Attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1), 80–91.
- White, R. C., & Aimola Davies, A. M. (2017). Asynchrony in the rubber hand paradigm: Unexpected illusions following stroke. *Cortex*, 93, 224–226.
- Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Computational Biology*, 6, Article e1000871.