Year : 2021

Investigating multisensory-motor mechanisms of bodily self-consciousness by combining psychophysical, electrophysiological and computational tools

Bertoni Tommaso Enrico

Bertoni Tommaso Enrico, 2021, Investigating multisensory-motor mechanisms of bodily self- consciousness by combining psychophysical, electrophysiological and computational tools

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_4A74700A5AC53

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Département des neurosciences cliniques

Investigating multisensory-motor mechanisms of bodily self-consciousness by combining psychophysical, electrophysiological and computational tools

Thèse de doctorat en Neurosciences

présentée à la

Faculté de Biologie et de Médecine
de l'Université de Lausanne

par

Tommaso Enrico Bertoni

Physicien diplômé de l'Université de Paris VI, France

Jury

Prof. Jean-Philippe Thiran, Président
P.D. Dr. Andrea Serino, Directeur
Prof. Micah Murray, Co-Directeur
Prof. Joseph Makin, Expert
Prof. Alessandro Farnè, Expert

Thèse n° 302

Lausanne 2021

Programme doctoral interuniversitaire en Neurosciences des Universités de Lausanne et Genève

LEMANIC NEURO SCIENCE DOCTORAL SCHOOL

UNIL | Université de Lausanne

UNIVERSITÉ DE GENÈVE

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Département des neurosciences cliniques

Investigating multisensory-motor mechanisms of bodily self-consciousness by combining psychophysical, electrophysiological and computational tools

Thèse de doctorat en Neurosciences

présentée à la

Faculté de Biologie et de Médecine
de l'Université de Lausanne

par

Tommaso Enrico Bertoni

Physicien diplômé de l'Université de Paris VI, France

Jury

Prof. Jean-Philippe Thiran, Président
P.D. Dr. Andrea Serino, Directeur
Prof. Micah Murray, Co-Directeur
Prof. Joseph Makin, Expert
Prof. Alessandro Farnè, Expert

Thèse n° 302

Lausanne 2021

Programme doctoral interuniversitaire en Neurosciences des Universités de Lausanne et Genève

LEMANIC NEURO SCIENCE DOCTORAL SCHOOL

UNIL | Université de Lausanne

UNIVERSITÉ DE GENÈVE

UNIL | Université de Lausanne

UNIVERSITÉ DE GENÈVE

Programme doctoral interuniversitaire en Neurosciences des Universités de Lausanne et Genève

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	Jean-Philippe	**Thiran**
Directeur·trice de thèse	Monsieur	Prof.	Andrea	**Serino**
Co-Directeur·trice de thèse	Monsieur	Prof.	Micah	**Murray**
Expert·e·s	Monsieur	Prof.	Alessandro	**Farné**
	Monsieur	Prof.	Joseph	**Makin**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Tommaso Enrico Bertoni

Master de l'Université Pierre et Marie Curie/Paris VI, France

intitulée

Investigating multisensory-motor mechanisms of bodily self-consciousness by combining psychophysical, electrophysiological and computational tools

Date de l'examen: 26 avril 2021

Date d'émission de l'Imprimatur: Lausanne, le 25 mai 2021

pour Le Doyen
de la Faculté de Biologie et de Médecine

Prof. Niko GELDNER
Directeur de l'Ecole Doctorale

Acknowledgements

This long journey would not have been possible without all the great companions that accompanied me, helped me, supported me, or asked me why I could not just get a real job.

First, thanks to Andrea for always finding the time and the patience to guide me, and for making all of this possible by believing that a physicist could turn into a neuroscientist. I would also like to thank the thesis committee, for accepting to evaluate whether this transformation was successful.

Thank you to all the MySpace lab for being friends and not only colleagues, it has been a pleasure to work with you. In particular, I would like to thank Giulio for the coffee breaks, the great science and the not so great chess, Michel, for always listening to me when I interrupted his work with inconclusive questions, and Silvia for the beers we shared at the Great Escape. I would also like to thank my friend and former MySpacer JP for the insightful discussions, the past collaborations and the many more to come.

Science is not only done in the lab, and all my friends played a role in this thesis in one way or another. I thank Nico for the great food and conversations about life, Matteo (and Giulio again!) for the epistemological conversations about science, Ale and Tobi for the epistemological conversations about, well…epistemology. Thank you Aurélie, Brunè and J for making P4 an happier place, and to all my other friends in Lausanne for making me feel at home here. Thank you also to Luca, Giulio, Bart, Fabio, Robi, for your friendship and for letting me bore you with my research, and to all the others back in Milan.

Special thanks go to my father, I owe my passion for science to you, and to my mother, for teaching me that science is not the only thing that matters.

Thank you, Elisa, for being besides me in this four years.

Last but not least, thanks to all of you who, maybe for a brief moment and without even knowing, helped me find the motivation and the enthusiasm to go forward.

Abstract - English

The goal of this research is to provide novel insights into the neurocomputational foundations of self-consciousness. Classically topic of philosophical investigation, self-consciousness has become an object of scientific study, in particular in the domain of neuroscience. Indeed, understanding the origin of self-consciousness is one of the most fundamental questions in our culture. Nevertheless, it is arguably as fascinating as it is elusive in quantitative scientific investigation, as it is traditionally regarded as an exquisitely experiential phenomenon. However, extensive research within the last two decades has demonstrated that so-called pre-reflexive, implicit components of self-consciousness can be explained in terms of bottom-up sensorimotor mechanisms, making them susceptible to scientific investigation. The minimal form of self-consciousness that arises from those components roughly translates to the subjective experience of existing within a physical body, with a precise location in the environment, and has been therefore termed bodily self-consciousness (BSC). In synthesis, it can be said that the two pillars of BSC are the experience of owning a body (body ownership), and of being in control of its actions, and through them, of events in the world (sense of agency). Both components of BSC depend of the integration of bodily sensorimotor signals. Body ownership emerges from the multisensory coherence of inputs within the peripersonal space, the space immediately surrounding the body (e.g.: seeing a hand being touched and simultaneously feeling touch on my hand can lead me to believe that what I am looking at is my hand). The sense of agency arises from sensorimotor congruencies between motor commands and observed actions (e.g.: sending a motor command to move my hand and seeing it moving with the same timing leads me to believe that it is me who is moving my hand). Here, we sought to uncover the key mechanisms of BSC by investigating behavioural, neurophysiological and computational properties of its two key components, body ownership and sense of agency. Four studies are presented, two focusing on sense of agency, and two on body ownership.

In the first study, we developed a neural-network model that learns the natural associations between visual, proprioceptive and tactile inputs to build a body-part centred representation of space. We then showed how such self-learned representation is sufficient to elicit a minimal form of body ownership, as shown by the ability of the model to reproduce some of its behavioural correlates.

In the second study, a pre-registered protocol, we tackle again the theme of body ownership, within the framework of Bayesian approximations of brain function. In this view, the brain performs nearly optimal statistical inference to estimate the probability that an object belongs to the body based on the congruency of multisensory inputs, leading to body ownership. This framework provides a quantitative support to bottom-up theories of self-consciousness, but empirical evidence for its applicability to body ownership and BSC in general is still rather scarce. Therefore, we proposed a set of tasks, and a paired modelling framework, to extend the evidence base for this hypothesis.

In the third study, we investigated how the congruency of visual and somatosensory feedback with motor commands affected the sense of agency, and neural signals in the motor cortex. The study was carried out in a tetraplegic participant using a brain-machine interface, decoding his motor commands and translating them into functional movements of his arm via a neuromuscular electrical stimulation system. We found that both visual and somatosensory incongruent feedback strongly reduced the sense of agency, and that the motor cortex encoded information about the congruency of sensory feedback. Moreover, incongruent somatosensory (but not visual) feedback led to a decrease in the decoding accuracy of motor commands, by partially overwriting the encoding of motor commands in the motor cortex.

In the fourth study, we used the same experimental model to investigate the relation between neural oscillations and sense of agency. We found that, regardless of sensory feedback, sense of agency was higher when movements occurred in a specific phase of mu waves before movement onset, so that the movement onset coincided with a negative trough of the oscillations. We then developed another paradigm to derive an implicit measure of sense of agency and applied the same analyses. We confirmed that the same oscillatory phase coinciding with high explicit agency judgements correlated with high implicit measures of agency. Since the pre-movement mu phase did not significantly affect local multiunit activity in the motor cortex, we speculate that its effect on agency may be mediated by influencing the connectivity between M1 and other brain areas involved in sense of agency.

In this thesis, we tackled the vast topic of self-consciousness by focusing on its bodily, bottom-up component, BSC. We provided novel behavioural, computational and neurophysiological insights on how the congruence of multisensory and motor signals contributes to two key aspects of BSC: body ownership and sense of agency. We argue that Bayesian models can be a powerful tool to account for the emergence of BSC from sensorimotor congruencies within a unifying framework. However, they can be employed effectively only if their theoretical implementation is constantly complemented and refined by rigorous empirical testing.

Abstract - Francais

L'objectif de cette recherche est de fournir de nouvelles perspectives sur les fondements neurocomputationnels de la conscience de soi. Classiquement sujet d'investigation en philosophie, la conscience de soi est désormais devenue un objet d'étude scientifique, en particulier dans le domaine des neurosciences. En effet, comprendre l'origine de la conscience de soi est l'une des questions les plus fondamentales de notre culture. Néanmoins, la conscience de soi est sans doute aussi fascinante qu'insaisissable à l'investigation scientifique quantitative, car elle est traditionnellement considérée comme un phénomène exclusivement expérientiel. Cependant, des recherches approfondies menées au cours des deux dernières décennies ont démontré que les composantes implicites, dites préréflexives, de la conscience de soi peuvent être expliquées en termes de mécanismes sensorimoteurs ascendants, ce qui les rend susceptibles d'être étudiées scientifiquement. Cette forme minimale de conscience de soi se traduit grossièrement par l'expérience subjective d'exister dans un corps physique, avec une localisation précise dans l'environnement, et a donc été appelée conscience de soi corporelle (BSC). En synthèse, on peut dire que les deux piliers de la BSC sont l'expérience de la possession d'un corps (appropriation du corps), et le contrôle de ses actions, et à travers elles, des événements dans le monde (sens de l'agentivité). Les deux composantes de la BSC dépendent de l'intégration des signaux sensorimoteurs corporels. L'appropriation du corps émerge de la cohérence des signaux multisensorielle dans l'espace péripersonnel, l'espace qui entoure immédiatement le corps (par exemple : voir une main qui est touchée et sentir simultanément un toucher sur ma main peut me faire croire que ce que je regarde est ma main). Le sens de l'agentivité découle des congruences sensorimotrices entre les commandes motrices et les actions observées (par exemple, envoyer une commande motrice pour bouger ma main et la voir bouger au même moment me fait croire que c'est moi qui bouge ma main). Ici, nous avons cherché à découvrir les mécanismes clés de la BSC en étudiant les propriétés comportementales, neurophysiologiques et computationnelles de ses deux composants clés, à savoir : la possession du corps et le sentiment d'agentivité. Quatre études sont présentées, deux portant sur le sens de l'agentivité et deux sur l'appropriation du corps.

Dans la première étude, nous avons développé un modèle de réseau neuronal qui apprend les associations naturelles entre les entrées visuelles, proprioceptives et tactiles pour construire une représentation de l'espace centrée sur les parties du corps. Nous avons ensuite montré comment cette représentation auto-acquise est suffisante pour susciter une forme minimale d'appropriation du corps, comme le montre la capacité du modèle à reproduire certains de ces corrélats comportementaux.

Dans la seconde étude, un protocole pré-enregistré, nous abordons à nouveau le thème de l'appropriation du corps, dans le cadre des approximations bayésiennes du fonctionnement du cerveau. Dans cette théorie, le cerveau effectue une inférence statistique quasi optimale pour estimer la probabilité qu'un objet appartienne au corps en fonction de la congruence des signaux multisensorielles, ce qui conduit à l'appropriation du corps. Ce cadre fournit un soutien quantitatif

aux théories ascendantes de la conscience de soi, mais les preuves empiriques de son applicabilité à l'appropriation du corps et à la BSC en général sont encore assez rares. C'est pourquoi nous avons proposé une batterie de tâches à utiliser en synergie avec un cadre de modélisation, afin d'élargir la base de preuves de cette hypothèse.

Dans la troisième étude, nous avons examiné comment la congruence du feedback visuel et somatosensoriel avec les commandes motrices affectait le sentiment d'agentivité et les signaux neuronaux dans le cortex moteur. L'étude a été réalisée chez un participant tétraplégique à l'aide d'une interface cerveau-machine décodant ses commandes motrices et les traduisant en mouvements fonctionnels de son bras via un système de stimulation électrique neuromusculaire. Nous avons constaté qu'un feedback visuel et somatosensoriel incongruent réduisait fortement le sentiment d'agentivité, et que le cortex moteur encodait des informations sur la congruence du feedback sensoriel. De plus, un feedback somatosensoriel incongruent (mais pas visuel) entraîne une diminution de la précision du décodage des commandes motrices, en remplaçant partiellement l'encodage des commandes motrices par l'encodage du feedback dans le cortex moteur.

Dans la quatrième étude, en utilisant le même modèle expérimental, nous avons étudié la relation entre les oscillations neuronales et le sentiment d'agentivité. Nous avons constaté que, indépendamment du feedback sensoriel, le sentiment d'agentivité était plus élevé lorsque les mouvements se produisaient dans une phase spécifique des ondes mu avant le début du mouvement, de sorte que le début du mouvement coïncidait avec un creux négatif des oscillations. Nous avons ensuite développé un autre paradigme pour dériver une mesure implicite du sentiment d'agentivité, et appliqué les mêmes analyses basées sur les oscillations de phase. Nous avons confirmé que la même phase oscillatoire coïncidant avec des jugements d'agentivité explicites élevés était corrélée avec des mesures d'agentivité implicites élevées. Puisque la phase de 8 Hz avant le mouvement n'a pas affecté de manière significative l'activité neuronale locale dans le cortex moteur, nous spéculons que son effet sur l'agentivité peut être médié par des effets sur la connectivité entre M1 et d'autres zones du cerveau impliquées dans le sens de l'agentivité.

Dans cette thèse, nous avons abordé le vaste sujet de la conscience de soi en nous concentrant sur sa composante corporelle, ascendante, la BSC. Nous avons fourni de nouveaux aperçus comportementaux, computationnels et neurophysiologiques sur la façon dont la congruence des signaux multisensoriels et moteurs contribue à deux aspects clés de la conscience de soi : la possession du corps et le sentiment d'agentivité. Nous soutenons que les modèles bayésiens peuvent être un outil puissant pour rendre compte de l'émergence du BSC à partir des congruences sensorimotrices dans un cadre unifié. Cependant, ils ne peuvent être utilisés efficacement que si leur mise en œuvre théorique est constamment complétée et affinée par des tests empiriques rigoureux.

List of abbreviations

BSC	Bodily self-consciousness
MPS	Minimal phenomenal selfhood
M1	Primary motor cortex
BMI	Brain-machine interface
RHI	Rubber hand illusion
IHI	Invisible hand illusion
VIP	Ventral intraparietal cortex
PPS	Peripersonal space
CCT	Crossmodal congruency task
TPJ	Temporo-parietal junction
SPL	Superior parietal lobule
IPS	Intraparietal sulcus
fMRI	Functional magnetic resonance imaging
PET	Positron emission tomography
SMA	Supplementary motor area
tDCS	Transcranial direct current stimulation
TMS	Transcranial magnetic stimulation
DLPFC	Dorso-lateral prefrontal cortex
MEG	Magnetoencephalography
EEG	Electroencephalography
LFP	Local field potential
NMES	Neuromuscular electrical stimulation

Table of contents

1. Introduction

1.1. Bodily self-consciousness

Throughout the centuries, self-consciousness has been a central theme in philosophy. When scientific thinking dared to approach the world of living organisms, it became clear that self-consciousness somehow emerged from extremely complex patterns of bioelectric activity of cells within the central nervous system. Indeed, the idea that the brain is the physical cause of self-consciousness constituted a major breakthrough, but the materialistic path has proven to be not less hard to follow than the metaphysical one. Arguably, the first difficulty with setting the bases for the scientific study of self-consciousness is that its rigorous definition is, to say the least, elusive. Traditional attempts to define self-consciousness in the 20th century are intimately related to linguistic and semantic aspects of self-reference, as they have essentially focused on the capacity to generate and master "I"-thoughts. The meaning of "I"-thoughts may seem intuitively clear, but they entail a subtle yet crucial difference from the mere use of the first person pronoun: immunity to error through misidentification. For example, it is theoretically possible for someone to be wrongfully referring to himself when saying "I am wearing a watch on my right wrist", as he may in principle be looking at someone else's right wrist, which looks and is placed exactly like his own. However, it is immediately clear that a misidentification error is strictly impossible when saying something like "I am feeling cold". In Wittgenstein's words, the pronoun "I" is respectively used as an object and as a subject in these examples (Wittgenstein, 1958). Only the latter use is effectively immune to misidentification errors, and therefore defines "I-thoughts". In this view, self-consciousness would to some extent reduce to the capacity to master the use of the pronoun "I" as a subject. This "deflationary" definition has been recently challenged on grounds of its unavoidable circularity (Bermúdez, 1998). The semantic mastery that is taken to be the crucial point in defining self-consciousness must unavoidably be employed "as it is" in order for such definition to make sense, as it cannot be reduced to non self-referential semantic concepts. Or, quoting Bermudez, "Any theory that tries to elucidate the capacity to think first-person thoughts through linguistic mastery of the first-person pronoun will be circular, because the explanandum is part of the explanans". In order to overcome the circularity entailed in narrative aspects of the self, it has been proposed to focus instead on more minimal aspects of selfhood (Gallagher, 2000), taking "non-conceptual first-person content" (Bermúdez, 1998) or pre-reflexive forms of subjective experience (Metzinger, 2003) as a starting point to investigate self-consciousness. Essentially, they consist in the self-specifying information that permeates perceptual experience (Gallagher, 2000), and allows to perceive oneself as the subject of such experience before cognitive reasoning (hence, pre-reflexively). For example, even when processing inputs that are not self-related, such as when visually perceiving an external object, we implicitly gather additional information about its location in space by putting it in relation with a spatial reference frame that is centred on our location in space. In other words, we not only see that object,

but we see it from a first person perspective, and, to some extent, a minimal sense of ourselves is therefore attained in all forms of perception, which had already been noted as early as in Aristotle's *Sense and Sensibilia*. The phenomenological manifestation arising from such basic set of self-specifying components of perception has been defined "minimal phenomenal selfhood" (MPS, Blanke & Metzinger, 2009), in order to distinguish it from higher level, narrative forms of selfhood. MPS does not require mastery of any specific semantic concept to be phenomenologically accessible, as it is intimately related to our physical body and its interaction with the external world. It has been argued that MPS defines a specific subset of self-consciousness, grounded in the "here and now" experience of our bodily selves. Such minimal and pre-reflexive form of self-consciousness has been therefore termed Bodily Self Consciousness (BSC, Blanke, 2012; Blanke et al., 2015). In synthesis, BSC rests on two key aspects of our subjective experience as physical selves: the sense of owning a body with certain characteristics and a precise location in space and time, and the sense of controlling that body and being able to act through it on the external world (Blanke & Metzinger, 2009; Gallagher, 2000). These key components of BSC, sense of ownership and sense of agency, are not only subjects of philosophical speculations, but they can be experimentally manipulated to investigate their behavioural and neural properties, making the study of self-consciousness scientifically accessible.

1.2. Multisensory mechanisms of bodily self-consciousness

As we mentioned, these conceptual advancements were fostered by a paradigmatic change in the scientific study of BSC, arising from empirical studies showing that it is possible to experimentally manipulate some of its key components. Amongst the first and best known is the study by Botvinick and Cohen (1998) about the "rubber hand illusion" (RHI). In a simple yet extremely insightful experiment, a rubber hand was in placed in front of naive participants, anatomically congruent with their real hand, which was hidden from their view. The participant's hand and the rubber hand were simultaneously stroked with a paintbrush in the same location and fashion. Subjects had the impression to feel touch *on* the rubber hand, and most importantly, they reported feeling as if the rubber hand was their own hand, with the illusory ownership being reported spontaneously by eight out of ten subjects. An interesting additional effect of the RHI is the so-called proprioceptive drift. When subjects were asked to point at their own hand after synchronous stroking, they reported a position shifted towards the rubber hand, as if the illusion also interfered with their hand position sense. A series of subsequent studies showed that, through synchronous visuo-tactile stimulation, it is possible to alter the subjective experience of ownership over a variety of body parts (the face, Sforza et al., 2010, the belly, Normand et al., 2011, the feet, Crea et al., 2015). It was shown that even a blank spot on a table (Armel & Ramachandran, 2003), or a completely empty space (Guterstam et al., 2013) can be felt as belonging to the body, by stroking it synchronously with participants' hands (invisible hand illusion, IHI).

Crucially, the rubber hand illusion can also be applied to the whole body (Ehrsson, 2007; Lenggenhager et al., 2007), eliciting illusory identification with an alien avatar, along with shifts in perceived self-location. This is a decisive step as, albeit related and elicited through the same multisensory mechanisms, the illusory ownership for body parts and for the whole body are conceptually different. While the former only involves a temporary incorporation of alien objects in the body schema, the latter affects the identification of the self with a physical body, one of the cornerstone features of MPS (Blanke & Metzinger, 2009). This conceptual difference is well explained by a simple example: it would be nonsensical to think that someone missing a limb has a diminished level of bodily-self consciousness, but it would be reasonable for someone missing the perception of his whole body. Taken together, these studies showed that a key component of BSC, such as body ownership, is not immutable and stable in healthy individuals, but it can be altered, at least temporarily, through simple multisensory manipulations. Arguably, the fundamental common trait of those manipulations is that they reproduce the intersensory congruencies that are normally observed for body parts (or the whole body) on an external object (or an avatar), which gets embodied as a result. The sensory modalities involved are typically a combination of inherently self-related modalities (the perceptual equivalent of immunity to error through misidentification?), such as touch and proprioception, and external senses, such as vision and audition, through which the external object to be embodied is perceived. When the spatio-temporal relations between touch/proprioception and visual stimulation of an external entity match the ones normally experienced throughout life, that entity is perceived as part of our body. If this coherent visuo-tactile stimulation pattern is produced on the back of a body-shaped avatar, it is perceived as our own body itself (Ehrsson, 2007; Lenggenhager et al., 2007).

1.3. Multisensory integration in peripersonal space and body ownership

Multisensory stimulation seems to be the key of manipulations of body ownership. In particular, visuo-tactile stimulation in anatomically congruent postures played a central role in a large number of studies. In order to answer the question about which brain areas are responsible for the emergence of ownership from multisensory integration, it is therefore natural to turn our attention to brain areas where visuo-proprioceptive-tactile integration is known to occur. Fronto-parietal regions, mainly in the ventral premotor cortex and the ventral intraparietal sulcus (VIP) in the primate brain are known to exhibit such response properties. In the early 80s, Rizzolatti and colleagues (Rizzolatti et al., 1981) discovered a network of visuotactile (or audiotactile, Graziano et al., 1999) neurons, responding to touch over relatively broad body areas (e.g. a hand, a forearm, the face…) and to external stimuli close to that body area. Importantly, their response properties were body-part centred, independently of eye orientation or the body part's location in space respective to the trunk. It is thought that such neurons are part of a network of functional areas that represent the space

closely surrounding the body, the peripersonal space (PPS). For brevity, let us then refer to those neurons as PPS neurons. When stimulated, PPS neurons evoke a variety of functional movements, which resemble the defensive behaviour animals display when startled by sounds or puffs of air (Cooke et al., 2003). Therefore, this network of multisensory neurons is thought to be connected to predicting interactions with the environment with a defensive purpose. Furthermore, these areas are contiguous to areas involved in planning and executing reaching and grasping movements, and the PPS network may be involved in more general mechanisms of interaction between the body and its surrounding space. Nevertheless, here we will mainly focus on visuo-tactile integration properties, as the most relevant aspect of PPS neurons for body ownership. The visual receptive fields of PPS neurons, as we mentioned, are anchored in space to the body part where the tactile receptive fields are, and therefore necessarily involve a set of transformations from retinotopic to body part-centred reference frames. Work in macaques showed that such reference frame transformations are guided by visual and proprioceptive information about the position of the limb in space (Graziano, 1999, 2000). Interestingly, when the monkey's true arm is hidden, PPS neurons can anchor their receptive fields to a fake arm, but only if this is placed in an anatomically plausible posture (Graziano, 2000). Furthermore, a subset of these cells has been shown to start coding for stimuli in the reference frame of a fake hand only after prolonged synchronous visuo-tactile stimulation, closely resembling the one delivered to elicit the RHI (Graziano, 2000). In sum, the PPS network appears to be a good candidate for being involved in the detection and processing of visuo-tactile congruencies that can be used to manipulate BSC in humans.

1.3.1. Peripersonal space in humans, evidence from behavioural and neuropsychological studies

Evidence for the existence of a neural system dedicated to the multisensory representation of the space surrounding the body, in all primates including humans, comes from neuropsychological, psychophysical and neuroimaging studies. In certain brain-damaged patients with hemispatial neglect, the characteristic impairments of action and perception are specific for stimuli in the near space (Brain, 1941), leading clinical diagnoses to distinguish between peripersonal and extrapersonal neglect. Similar spatially selective impairments can be artificially induced with targeted lesions in the postarcuate cortex of macaques, the same premotor region in which PPS neurons were first observed (Rizzolatti et al., 1983). The animals displayed a lack of attention in the contralesional side both at the somatosensory and the visual level, and the visual deficit was limited to the near space. Interestingly, such deficit was accompanied by an inability to grasp food when presented in the contralesional portion of the visual space. Another pathological manifestation of such dedicated, multisensory-motor functional network is crossmodal visuotactile extinction, in which visual stimuli presented on the ipsilesional side can suppress the perception of tactile stimuli on the contralesional side, which would be normally perceived in the absence of an ipsilesional distractor

(Bender, 1952; Driver et al., 2004). Importantly, extinction has been noted to occur only when the distracting visual stimulation occurred spatially close to the ipsilesional hand, showing that such cross-modal inteferences can happen in a body-part centred reference frame (di Pellegrino et al., 1997).

This observation has arguably inspired the methods that were subsequently developed to investigate PPS behaviourally in healthy humans. One of the first amongst these was the so-called crossmodal congruency task (CCT) (Driver & Spence, 1998). In the CCT, participants are asked to perform speeded judgements of the vertical elevation of tactile stimuli delivered on the hand, while congruent (e.g. same elevation) or incongruent (opposite elevation) visual distractors are presented at various locations in space. Participants are generally faster at performing the discrimination when visual and tactile stimuli are congruent. Importantly, such effect is stronger when the visual distractors are presented within the peripersonal space (Maravita et al., 2003). Later on, other tasks were developed to investigate visuo-tactile interactions thought to be the hallmark of PPS representation, based on simplified versions of the CCT. Subjects are asked to respond as fast as possible to tactile stimuli presented on the hand, face or other body parts, while task-irrelevant visual or auditory stimuli are presented either near or far from the target body part (see e.g., Serino et al., 2007). Participants are faster to respond to touch when external stimuli are near, as opposed to far, from the haptically stimulated body part (Holmes et al., 2020). This speeding up of tactile reaction times by visual or auditory stimuli is used as a proxy of PPS, whose spatial extent is defined by the region where the speeding effect occurs. Following the observation that PPS neurons are especially sensitive to stimuli that approach the body (Graziano et al., 1997), this paradigm was later improved by using looming visual or auditory stimuli (Canzoneri et al., 2012). This latter behavioural measure of PPS representation will be the main focus of this thesis, as it directly tackles visuo-tactile interactions that are key to bodily self-consciousness.

Nevertheless, it may be worth briefly mentioning other measures of PPS representation that have been developed in the last two decades. Focusing specifically on defensive properties of PPS, Sambo and colleagues found that the eye blinking reflex elicited by painful stimuli delivered to the hand increased when the hand was near to the face (Sambo et al., 2012). They concluded that the intensity of the blinking reflex could be used as a proxy of the activation of the defensive PPS system (see also Bufacchi & Iannetti, 2018 for a review). In an older study by Longo and colleagues (Longo & Lourenco, 2006), a phenomenon known as pseudoneglect was used to investigate PPS representation. When asked to bisect a line presented in the near space, healthy humans show a consistent leftward shift, as if the left portion of space was somewhat over represented. This bias decreases and reverses to a rightward bias as the distance of the line increases, and this phenomenon has been interpreted as a proxy of the differentiation between peripersonal and

extrapersonal space. Finally, in a recent study, a method based on the bias induced by somatosensory stimuli on the perceived temporal order of visual stimuli was proposed (Spaccasassi & Maravita, 2020). When presenting a tactile stimulus on one hand, and visual stimuli at various level of temporal asynchrony on the two sides of the body, the visual stimulus on the same side of the tactile stimulus is perceived earlier in time, but only when the visual stimuli are in the near space. Thus, the spatial modulation of the temporal bias effect induced by visual stimuli has also been taken as proxy of the differentiation between PPS and extrapersonal space.

1.3.2. The link between BSC and PPS in behavioural and neuroimaging studies

Indeed, the characteristics of the PPS system suggest that it shares some of the neural principles that underlie BSC. The processing of visuo and audio tactile stimuli in a body and body-part centred reference frame is implicitly linked with self-location and body representation, and with mechanisms that underlie multisensory manipulations of body ownership. Importantly, the link between BSC and PPS representation has been confirmed behaviourally, by combining an audio-tactile measure of PPS representation and the full body illusion (Noel et al., 2015). After inducing the full-body illusion in healthy participants, the region of space in which auditory stimuli induce a reduction of reaction times shifted towards the embodied virtual avatar, following the shift in perceived self-location induced by the full-body illusion (Ehrsson, 2007; Lenggenhager et al., 2007). This effect has been interpreted as if the PPS was anchored to the perceived location of the self, and not to the physical body in space. Neuroimaging literature provides evidence that PPS representation is relevant for BSC not only behaviourally, but in terms of shared neural mechanisms. A recent meta-analysis (Grivaz et al., 2017) investigated which brain areas are consistently activated across different studies investigating the neural correlates of PPS processing and body ownership. PPS processing was found to be localized mainly at the temporo-parietal junction (TPJ), and at the intersections between dorsal and ventral premotor cortex, and between primary somatosensory cortex and the superior parietal lobule (SPL). For body ownership, consistent activations were found at a slightly lower location in the ventral premotor cortex, and between the superior parietal lobule (SPL) and the intraparietal sulcus (IPS). This latter region, in the left hemisphere, showed a consistent overlap between body ownership and PPS related processing. Importantly, connectivity analyses showed that regions of the PPS network and of the body ownership network were strongly interconnected between each other.

1.3.3. Neural-network models of PPS representation. Previous works and open questions.

The principles of bodily self-consciousness and their link with PPS representation have been extensively investigated conceptually, behaviourally and at the neuroimaging and neurophysiological level. However, from a mechanistic point of view, a connectionist explanation of how these properties

emerge from the interaction of individual neurons is still missing. In this direction, computational simulations of neural-network models are arguably an extremely valid investigation technique, but only a few significant attempts in this direction have been made. The first neural network model of PPS representation was proposed by Magosso and colleagues (Magosso et al., 2010). This biologically realistic network consisted of two populations of unisensory tactile and visual (or auditory) neurons, connected with feedforward and feedback synapses with a population of multisensory neurons. Tactile neurons coded for touch on the hand, and visual neurons for the distance from the hand of an external visual stimulus, with each neuron coding preferentially for a given location of the hand-centered visual space. The network did not directly model the reference frame transformations from retinotopic to hand-centred coordinates, and focused directly on the modelling of upstream neural computations. The weights of unisensory to multisensory synapses were tuned in order to reproduce in the artificial multisensory neurons the properties observed in neurophysiological studies in PPS neurons from the primate VIP and ventral premotor cortex. Multisensory neurons received strong excitatory projections from tactile neurons, while the strength of the synaptic input from visual neurons decreased with distance. This way, artificial PPS neurons responded strongly both to tactile inputs and visual stimuli presented close to the “hand”. Furthermore, when a tactile input was provided coupled with a nearby visual stimulus, the activity of PPS neurons raised faster than when the visual stimulus was in the far space. Under the reasonable assumption that the response time to a tactile input is related to the velocity of the response of PPS neurons, this result is in line with behavioural assessments of PPS representation, showing a decrease of tactile reaction times in the presence of near external stimuli. In a further study, a version of this network modified to include Hebbian learning reproduced behavioural and physiological findings about the plasticity of PPS representation (Serino, Canzoneri, et al., 2015). A seminal study by Iriki (Iriki et al., 1996) showed that, after training monkeys to use a tool to retrieve distant food, the visual receptive fields of hand-centred PPS neurons enlarged. Confirmations of tool-induced plasticity of PPS representation followed shortly afterwards in neuropsychological (Maravita et al., 2001) and behavioural (Canzoneri et al., 2013) studies. PPS as assessed through crossmodal extinction, or by measuring visuotactile interactions through tactile reaction times, was found to enlarge when using a tool or a prosthesis. Serino and colleagues further investigated PPS plasticity in a coupled neural-network and behavioural study, showing that tactile stimulation coupled with visual (or auditory) stimulation in the far space can lead to extend PPS representation beyond its usual boundaries, even in the absence of a tool (Serino, Canzoneri, et al., 2015). Crucially then, it seems that the key of PPS plasticity lies in bottom-up Hebbian associations in multisensory inputs, requiring no form of top-down attentional control. Nevertheless, in the neural network the learning effect was still implemented on a pre-defined synaptic connectivity.

In the following years, other neural-network models in the field of robotics investigated different aspects of PPS representation with an applicative rather than neuroscientific purpose. These include models of visuo-tactile interactions for safety and impact avoidance (Nguyen et al., 2018; Roncone et al., 2016; Straka & Hoffmann, 2017), reaching (Juett & Kuipers, 2019), the development of a body schema through visuo-tactile interactions (Pugach et al., 2019). Finally, a different, but closely related topic is the field of neural-network models of reference frame transformations. Pouget and colleagues (2002) modelled reference frame transformations amongst three general spatial coordinate sets, represented by one “unisensory” neural population each. For example, one population could code for the position of the hand in retinotopic coordinates, the second for its proprioceptive position, and the third for the gaze angle. For any pair of coordinates encoded in two populations, the offset between them was encoded in the third population. Depending on the relative weight of feedback and feedforward synapses, the network could either produce an estimate of the position in a given reference frame, given the position encoded in the other two populations, or combine information from the different modalities to produce more reliable estimates. Makin and colleagues (2013) addressed the same topic with a slightly different architecture, using three unisensory populations connected to a multisensory population and not directly between themselves. The crucial advance in this approach was that the weights of synaptic connections between neural populations were not hand-wired, but learned through Hebbian plasticity from the natural statistics of simulated sensory inputs. In other words, reference frames spontaneously aligned by synaptic tuning to accommodate the statistical relations between inputs in different modalities. Finally, a more recent work offers an interesting example of unisensory reference frame transformations, where a multi-layer neural network develops neurons tuned to the hand-centred coordinates of external objects, based on purely visual inputs (Born et al., 2017). Visuo-tactile interactions and reference frame transformations are key to PPS representation, and common mechanisms link them to body representation and body ownership. Neural-network modelling would allow investigating the mechanistic neural computations that underlie such link, but no previous study specifically addressed this question.

1.4. Sense of agency

The second cornerstone of BSC is the sense of agency, the feeling of being the cause and the ones in control of our actions, and consequently of events in the external world that such actions cause (Gallagher, 2000). Especially if restricting our attention to pre-reflexive mental phenomena, the awareness of our causal efficacy on our body, the external world, and arguably even on our mental states (Gallagher, 2000), is a fundamental aspect in recognizing and defining ourselves as independent entities, or, in other words, in shaping the self-other boundary.

In the most simple and general way, agency arises from the congruence between intended and executed actions (Jeannerod, 2003).The brain constantly monitors self-generated efferent signals

and afferent signals arising from the body and the surrounding environment. When an action is generated, afferent information from the body informs the brain about its sensory consequences, and when they match our intentions a sense of agency arises. In everyday experience, this match is practically always present for intentionally generated actions. However, if the available sensory feedback about an action is either spatially or temporally misaligned with our intentions and motor commands (e.g., by showing delayed or distorted visual feedback about hand movements), the experience of agency is diminished or abolished (Farrer et al., 2008). This simple principle is thought to be at the basis of the earliest signs of self-recognition in infants, who at 5 months of age are already consistently able to discriminate their own movements from those of other infants based on sensorimotor contingencies (Bahrick & Watson, 1985). A minimal capability to distinguish self from other generated movements seems to be even present from birth, as newborns were noted to produce more head movements when touch in the perioral region is provided by another as compared to when it is self-generated (Rochat & Hespos, 1997).

Indeed, the detection of visuo-motor congruencies can be used as a robust and effective tool to distinguish oneself from others. For example, congruency detection alone is sufficient to self-recognize in a mirror in unsupervised manner (for an interesting example in robots see Gold & Scassellati, 2009), as it requires no previous knowledge of one's visual appearance. In this view, congruency detection would act as the initial seed of self-recognition, allowing then to associate oneself with specific visual features. Furthermore, recent evidence suggests that relatively short exposure to visuomotor congruencies is sufficient to influence and update the visual representation of one's own face (Serino, Sforza, et al., 2015). Participants observed a virtual avatar's face moving either synchronously or asynchronously with their own face movements, and were later asked to evaluate whether they self-recognized in faces obtained by different levels of visual morphing between their own face and the avatar. After exposure to the synchronously moving face, participants were more likely to recognize as their own faces more similar to the virtual avatar.

Visuomotor congruencies are likely to play a crucial role in a classical test of self-awareness testing mirror self-recognition: the mirror mark test. In this simple and brilliant experiment, first performed by Gallup in chimpanzees (1970), animals are put in front of a mirror with a paint mark on their face. If they try to remove the mark from their own face, this is taken as a sign of self-consciousness, while if they show no interest in the mark or the reflection, or attempt to socialize with it as if it was another individual, they are deemed as non self-conscious. Remarkably, animals who are able to self-recognize in a mirror have been reported to intentionally exploit visuomotor congruencies, by performing repetitive movements to investigate whether the image in the mirror is indeed their own (Ari & D'Agostino, 2016). Nevertheless, only a small fraction of animals (essentially great apes, dolphins and elephants) is able to self-recognize in a mirror, and some species seem to even be

able to detect sensorimotor contingencies, and deliberately produce movements to assess them, without this leading to passing the mirror mark test (Ari & D'Agostino, 2016). Children, in turn, seem to be able to robustly pass the test starting at the age of 15 to 20 months (Bahrick, 1995). Intriguingly, a recent study showed that it is possible to train animals to pass the mirror mark test. Chang and colleagues (Chang et al., 2017) trained Rhesus monkeys (a species known not to be able to spontaneously self-recognize in a mirror) to touch spots of light produced by a laser beam on their face. This way, the animals could experience visuo-motor and visuo-proprioceptive-tactile congruences arising from self-directed behaviour in front of a mirror. After this training, animals could reliably pass the mirror mark test. Taken together, this suggests that the detection of sensorimotor contingencies is a necessary, but not sufficient condition for self-awareness, at least as assessed by the mirror mark test.

Furthermore, the mechanism of contingency detection that leads to sense of agency may also play a role in the development of causal reasoning that underlies higher levels of self-consciousness. In infancy, voluntary actions are routinely employed to learn about the causal structure of the world, and sense of agency may act as a guiding factor in such inference (Zaadnoordijk et al., 2015). Interestingly in this sense, disorders of the self such as schizophrenia, that are characterized by dysfunctional causality inferences, or delusions, are also accompanied by an impaired sense of agency for bodily actions. People with schizophrenia experience abnormal feeling of control over actions and thoughts, sometimes misattributing to others their own actions (Daprati et al., 1997; Hur et al., 2014; Mellor, 1970; Moore & Obhi, 2012). It has been suggested that these issues may not simply be a symptom of the disease, but they may reflect a more general impairment in causality inference that might act as a pathogenic factor (Fletcher & Frith, 2009). Therefore, it has been proposed that schizophrenia may essentially originate as a disorder of prediction, extending from low-level sensory events to higher level cognitive functions (Fletcher & Frith, 2009).

1.4.1. Predictive and postdictive accounts of agency

The fact that sensorimotor contingencies give rise to sense of agency in humans is widely accepted, but the nature of the process through which this happens is debated. Two main classes of theories have been proposed, known as predictive and postdictive accounts of agency. According to predictive accounts, first proposed by Frith and colleagues (2000), crucial for sense of agency would be internal predictions of the expected sensory consequences of motor commands, generated and internally stored before the actual movement takes place. Once the action occurs, these predictions are compared to the actual afferent information leading to sense of agency when a match with predictions is met. This account is often referred to as the "comparator model" of agency.

The postdictive account, first proposed by Wegner (2002), emphasizes the idea that the true underlying causes of motor acts would be unconscious processes, while only the act itself and the intention to do so would be cognitively accessible (Moore, 2016). In this view, intentions would be only the conscious manifestation of some underlying unconscious process, and therefore play no actual causal role in decision making and in the generation of actions. In Wegner's theory, not only there is no such thing as a *metaphysically free* will (the philosophical uncaused cause), but the idea that our conscious thoughts cause our actions is itself an illusion: in fact, actions simply "happen" to us. As a consequence, sense of agency would be based on a post-hoc inference on the cognitively accessible cues about intention, or the retrospective evaluation of whether what happened is compatible with what I had intended to do. These two accounts are not necessarily opposed or mutually exclusive, as the use of terms such as prediction and postdiction would seem to suggest. From the pure temporal perspective, predictive processes need to include post-movement cues in order for the comparison to be made. At the same time, a postdictive evaluation requires pre-movement intentions to be at least stored in memory until sensory feedback is available, even if predictions are not explicitly formulated. The distinction between the two accounts is then somewhat subtler. Predictive accounts put a strong accent on low-level sensorimotor aspects, and on the fine-grained integration of afferent and efferent information that occur pre-reflexively, on a fast timescale during and immediately after the movement. Postdictive accounts do not deny these aspects, but rather consider them of secondary importance, and claim that sense of agency arises from a process more akin to a higher level, cognitive "confabulation" about the cause of the movement, taking place mostly after the action occurs. The two positions are therefore complementary rather than opposed, and in more recent works it is generally accepted that both predictive and postdictive processes contribute to sense of agency (Synofzik et al., 2013).

A further attempt to reconcile predictive and postdictive accounts is based on the general theory of cue integration, stating that when different sources of information about the same phenomenon are available, the brain combines them according to their reliability to obtain more accurate estimates. Applying this theory to the sense of agency, predictive and postdictive cues would be then combined to infer the causal relations underlying our actions, and give rise to agency. An interesting consequence of this theory is that the relative weight of predictive and postdictive mechanisms would depend on the respective reliability in a specific situation, and therefore be context dependent. For example, consider the cognitive, postdictive cue constituted by the presence of other agents who might potentially have caused an event. If I am alone in a room and someone drops a pen next to me through a concealed remote control, I may be tricked into thinking retrospectively that I must have caused the pen to fall. This is because the postdictive cue about the fact that I am the only possible cause of the event is expected to be very reliable in this context. If many people instead

surrounded me, the cognitive cue about the potential author of an event would become irrelevant, and I would have to rely more upon sensorimotor mechanisms to determine its cause.

1.4.2. Behavioural and neural correlates of sense of agency

It is now worth spending a few words on the behavioural paradigms used to assess sense of agency, and on its neural correlates that have been identified in previous studies. The most straightforward way to assess agency is through explicit judgements, in which participants are simply asked whether (and/or how much) they felt like they generated an action or effect. Nevertheless, explicit agency judgements have been shown to be cognitively biased, as humans tend to consistently misattribute to themselves events that are not self-caused, especially when their outcomes have a positive valence. For this reason, researchers have developed implicit methods to assess agency, whose most widely used example is the intentional binding paradigm (Haggard et al., 2002). Subjects are asked to perform an action typically associated with a given predictable outcome (e.g. pressing a button that leads to a beeping sound), and to report the position of a clock at the time of the action or of the external event, in order to measure its perceived timing. It was found that subjects tend to perceive actions as occurring later, and their effects earlier, and that this “attractive pull” of actions towards their consequences is present only when voluntary movements are performed. Involuntary movements induce no or reversed effects (Moore et al., 2009; Yoshie & Haggard, 2013). These biases in time perception are therefore specifically linked to intentionality, and therefore they have been widely used as an implicit measurement of the sense of agency. Other indirect measures of sense of agency focus on sensory attenuation (Dewey & Knoblich, 2014; Garrido-Vásquez & Rock, 2020), the reduction in the perceived intensity of self-generated events (Blakemore et al., 1998, 2000). The key idea is that sensory attenuation would arise from expectations generated by a forward predictive model (Bays, 2006), the same that leads to sense of agency in the comparator model. Indeed, in the case of self-touch, introducing an artificial delay between the action and its sensory consequences leads to reduced attenuation and perceived self-causation (Bays et al., 2005; Blakemore et al., 1999). Furthermore, implicit self-attribution of external events can be assessed by monitoring automatic error compensation when provided with partially incongruent feedback, which can take place without conscious awareness of the error (Fourneret & Jeannerod, 1998; Grünbaum & Christensen, 2020; Kannape & Blanke, 2012). Sense of agency is strictly linked to the concept of intention (see Haggard, 2017), and thus it is important to mention here the cornerstone study about intentionality performed by Libet and colleagues in 1983 (Libet et al., 1983). They instructed participants to freely press a button whenever they wanted, while monitoring the position of a rotating clock hand to report the onset time of the subjective will to act. They found that the electrical signal that habitually precedes self-generated movements (the readiness potential, Kornhuber & Deecke, 1965, 2016) also preceded conscious awareness of the will to act by about 350 ms. This was interpreted as the first evidence that the chain of neural process that leads to intentional movements can actually commence unconsciously.

Overall, the conceptual and behavioural bases of the sense of agency are thus relatively well understood. At the neural level, several studies have also investigated the correlates of the sense of agency through neuroimaging and perturbative techniques, and the results are arguably more varied and less straightforward to interpret. The majority of studies aiming at identifying the key brain regions used functional magnetic resonance imaging (fMRI), yielding a plurality of areas linked to sense of agency. Two meta-analytic studies attempting to draw more general conclusions still found sparse clusters of activity. The first meta-analysis, considering 15 fMRI or PET (positron emission tomography) studies (Sperduti et al., 2011), evidenced the role of regions including the TPJ, pre-supplementary motor area (pre-SMA), precuneus, and dorsomedial prefrontal cortex. As pointed out in the second, more recent meta-analysis (Zito et al., 2020), part of this variability may be due to the wide range of manipulations, paradigms and definitions of the sense of agency used in such neuroimaging literature. One first confounding factor is the use of different contrasts between high and low agency conditions. These include, for example, comparing trials with spatially congruent or incongruent visual feedback, with delayed as opposed to synchronous feedback, or high versus low subjective ratings of agency. Clearly, this can introduce biases, such as activations of brain areas processing the specific sources of incongruence targeted in a given study, rather than “pure” sense of agency. Another important difference concerns the used definition of the sense of agency, depending on whether it was referred strictly to bodily motor control or to a more general feeling of causation over events in the external world. Finally, Zito and colleagues’ meta-analysis (2020) suggests that different regions are involved in positive as opposed to negative sense of agency, with the latter being possibly related to more general error signals in the brain. The study, which focused specifically on 22 studies on motor control, found no significant activation for positive agency, and it confirmed the role of the TPJ in negative agency. Overall, neuroimaging studies yielded largely variable and not conclusive evidence about the network of regions involved in the sense of agency.

Possibly, a clearer picture can be drawn by also considering studies using different techniques, and introducing a broad differentiation between frontal and parietal areas, which we will start reviewing from frontal areas. A meta-analysis on 7 transcranial direct current stimulation (TDCs) studies, using intentional binding as an implicit measure of agency (Khalighinejad et al., 2016), found that stimulating the dorso-lateral prefrontal cortex (DLPFC) increases the intentional binding between actions and outcomes, but only when participants spontaneously select the actions to perform. This is in line with the idea that prefrontal regions may be involved in the sense of agency, as they contribute to action planning and selection. Conversely, TDCs (Cavazzana et al., 2015) and inhibitory repetitive transcranial magnetic stimulation (TMS) (Moore et al., 2010) over the pre-supplementary motor area has been reported to reduce intentional binding. This suggests that a set of different regions is causally involved in sense of agency, and confirms the differentiation between positive and negative feeling of agency. As shown by an interesting neurophysiological work on epileptic

patients with implanted electrodes, relatively small populations of single neurons within the SMA are sufficient to predict the onset of volitional processes, preceding the subjects' awareness of the incoming intention to move (Fried et al., 2011). On the other hand, a recent study using electrical stimulation in awake surgery patients found that both stimulation of the premotor and somatosensory cortex led to the arrest of motor execution, but only in the first case subjects are unaware of it (Fornia et al., 2020). Similarly, in a previous study, premotor stimulation evoked overt movements, but lead to no awareness of those movements being performed (Desmurget et al., 2009). Studies on patients with premotor damage also seem to indicate that lesions to this area lead to diminished awareness of one's own motor deficit (Berti et al., 2005). Moving to parietal areas, it was shown that a TMS pulse over the angular girus (amongst the "non-agency" regions highlighted in fMRI meta-analyses) can decrease the effect of incongruent feedback about action outcome (Chambon et al., 2015). Interestingly, another fMRI study had previously shown that the connectivity between premotor areas and the angular gyrus is reduced in incongruent trials (Eimer & Schlaghecken, 2003). Therefore, this parietal region may be responsible for the comparison between sensory feedback and predictions, and for the detection of mismatches. It has been proposed that both frontal and parietal regions are involved in sense of agency, but possibly with a different role (Haggard, 2017). Frontal regions are responsible of volitional processes and action selection, and are therefore linked to intentions that are necessary for sense of agency. Parietal regions, instead, may host the processes where the comparison between intentions (or predictions of their sensory consequences) and afferent information is performed. Still, some of the studies reviewed here seem to suggest that action monitoring also takes place in premotor areas (Berti et al., 2005; Fornia et al., 2020). Conversely, Desmurget's 2009 study found that parietal stimulation could lead to a vivid "urge to move", as well as to the illusory sensation of having moved at higher intensities. An interesting opinion on this matter has been expressed by Desmurget and colleagues (2012), who suggested that parietal regions may represent the desired final state of actions to be performed.

Although the studies described until now provide useful insights into which brain regions are involved in sense of agency, they provide little insight into the dynamics of the underlying neural processes, as the techniques used do not possess the temporal resolution needed to investigate such aspects. In this sense, the main sources of information are even scarcer, as they mainly consist of a few studies using EEG and magnetoencephalography (MEG). In an EEG study by Kang and colleagues (2015), it was shown that higher levels of agency, as induced by coherent virtual feedback during a hand movement task, are associated with lower alpha (8-12 Hz) power in temporal and parietal electrodes, and with a decreased coherence of alpha oscillations in frontal areas. Alpha band oscillations were also found to be involved in the process of detection of sensorimotor contingencies during mirror self-recognition (Serino, Sforza, et al., 2015). Participants exposed to a virtual face moving synchronously with their own showed a greater motor evoked suppression of mu oscillations

over sensorimotor areas, compared to when it moved asynchronously. Another interesting MEG study used instead a cognitive priming to bias participants' sense of agency at otherwise equal sensory feedback congruence during a hand tapping task (Buchholz et al., 2019). It was found that higher beliefs of agency were accompanied by a stronger beta-band connectivity between the primary motor cortex (M1) and both the inferior parietal lobe and right middle temporal gyrus, a region compatible with the putative seat of the sensorimotor comparator.

1.4.3. Brain machine interfaces, a tool to investigate sense of agency

Arguably, one of the main difficulties in the investigation of sense of agency, especially for normal bodily movements, is that the congruence between motor commands and actions is usually perfect, and extremely hard to manipulate experimentally. Indeed, until recently, performing a bodily action was ultimately the only mean of producing an effect in the external world. Nowadays, Brain machine interfaces (BMIs) provide a context where it possible to dissociate intentions, motor commands and external events. BMIs are systems that access and decode ongoing neural activity, typically based on signals from scalp or chronically implanted EEG, and use them to control an external device. They have high potential for clinical applications in neuroprosthetics, and, in the longer term, in the enhancement of human capabilities. More importantly for our field of research, they provide an alternative pathway for the brain to interact with the world, which can be manipulated and flexibly controlled. It is therefore evident that they constitute an exciting tool to investigate sense of agency. Furthermore, not only can BMIs inform research on the sense of agency, but advancing our understanding of sense of agency may help develop more efficient ad ergonomic BMIs. Indeed, since it accompanies all our spontaneous actions, sense of agency (or a lack of it) may affect the neural signals that are used by brain machine interfaces and influence their functioning. Despite the potential importance of the sense of agency for BMI control, it has not been investigated systematically in the BMI field.

1.5. Bayesian approximations of brain function

Body ownership and the sense of agency are two components of BSC that give rise to a distinct phenomenology, rely on different brain structures and are behaviourally dissociable (Kalckert & Ehrsson, 2012). For example, if a muscular twitch is artificially induced in my hand, I will still recognize that it is indeed my hand that is moving, but I will not experience agency for that movement. Conversely, when playing a videogame, my sense of agency for the character I am controlling will not typically lead to fully embodying it, so that I can easily tolerate if my character dies in the game. Nevertheless, empirical evidence shows that they share common multisensory principles. By showing participants a rubber hand that moved synchronously with their real hand, Dummer and colleagues (2009) could induce the same illusory ownership as in the classic RHI. Both active and

passive movements were effective in inducing the illusion, and, as for the RHI, temporal asynchrony abolished it. Additionally, the authors show that active movements induced a stronger feeling of ownership, although other studies could not reproduce this result (Kalckert & Ehrsson, 2014; Walsh et al., 2011). Furthermore, it is worth noting that even the 2012 study by Kalckert, which argued for the dissociability of agency and ownership, found that despite ownership can be induced without agency, and vice versa, subjective ratings are positively correlated, but only when both scores are high. Their conclusion was that ownership more strongly depends on anatomical congruence, while agency relies more strongly on volitional and sensorimotor cues, but they mutually reinforce each other when they are both present. Intuitively, the conceptual link between the mechanisms of agency and ownership lies in the striking importance of bottom-up congruencies in sensory and/or motor evidence about the body, and its relationship with the external world. For example, what made the RHI so remarkable was the fact that cognitive constrains and years of prior knowledge about our hand could not overcome the effect simple spatio-temporal contingencies in visuotactile inputs, at least at the pre-reflexive level. Clearly, participants *know* that the rubber hand is not their own, but this does not prevent them from *feeling* as if it was indeed part of their body. A few years after the first RHI paper, Ramachandran (2003) had the intuition to compare this phenomenon to what he called “Bayesian logic”, meaning that the process giving rise to illusory ownership would resemble statistical inference. Since in everyday experience it would be extremely unlikely to feel touch on the hand every time that an alien object is stroked, and extremely likely when it is my hand that is stroked, the brain “infers” that the rubber hand must be *my* hand. Indeed, this is nothing but a qualitative formulation of Bayes theorem for statistical inference.

1.5.1. Bayesian approaches to BSC

The idea that the brain’s functioning may be in general described in terms of statistical inference is amongst one of the oldest general neuroscientific principles, dating back even to Helmholtz and his “unconscious inferences”, and was subsequently present throughout decades of research in neuroscience. Barlow (1961) described sensory processing as a form of redundancy reduction, in the statistical framework of information theory. Later on, these ideas have inspired advances in machine learning and artificial intelligence, for example the brain-inspired artificial neural network for statistical inference named, not by chance, a Helmholtz machine (Dayan et al., 1995). However, it was not until recently that a precise mathematical formulation of this principle was empirically tested.

In 2002, Ernst and Banks studied how human subjects combined visual and tactile information when estimating the size of a hand-held object in different visibility conditions, and in the presence of conflicting visual and tactile information (Ernst & Banks, 2002). They found that not only the final estimate was a combination of the actual visually and haptically perceived sizes, but that their relative

weight in the combination was proportional to the squared inverse of the uncertainty on those modalities (measured as the mean squared error in a purely unisensory task). When visual information was made less reliable, its relative weight in the final size estimate decreased (and the mean error on size judgements increased) compatibly with model predictions. This cue combination strategy matches that of an ideal Bayesian observer aiming to minimize the error on the final size estimate given intrinsically noisy sensory information, and has been therefore termed optimal integration. Optimal integration is not limited to visual and tactile modalities, but has been shown to apply to a wide range of senses (Alais & Burr, 2004; Butler et al., 2010; Körding et al., 2007) or even when combining different cues within one sensory modality (Thurman & Lu, 2014).

Their flexibility, wide range of applications, and the presence of a clear evolutionary motivation (as they provide an "optimal policy"), led Bayesian approximations of brain function to become an extremely influential theoretical framework for human behaviour, which received increasing interest also in the field of BSC. Not only Ramachandran's intuition about body ownership, but also Synofzik's more recent (2013) attempts to reconcile predictive and postdictive accounts of sense of agency fall within this interpretative framework, proposing that sense of agency arises from the optimal integration of different cues about intention and action. Indeed, Bayesian approximations of brain function provide a computationally rigorous and general background to treat the detection of sensory and motor contingencies that seems to be key for both body ownership and sense of agency. Therefore, they may constitute a promising path towards a unifying account for these key components of BSC, which has been extensively explored at least conceptually. Apps and Tsakiris, for example, (2014) proposed that self-recognition would also result from the process of statistically inferring what is more likely to be caused by *me*. Indeed, inference in the brain may not be limited to sensorimotor processes, but extend to mental states, and humans would entail a model of themselves as the most likely causes of their thoughts (Friston, 2011; Limanowski & Blankenburg, 2013). Furthermore, the same idea has been applied to connect interoception and emotional processing (Seth, 2013; Seth & Friston, 2016), and to explain mental disorders such as autism (Palmer et al., 2017) and schizophrenia (Fletcher & Frith, 2009). In sum, the literature addressing the link between Bayesian inference and various aspects of BSC came to constitute an extremely rich body of studies in the last decade, amongst which we have reviewed here only a few representative ones. Remarkably, the vast majority of those studies are conceptual or purely mathematical. Works aiming to empirically test the predictions of a rigorous mathematical formulation of Bayesian inference for BSC are strikingly scarce, and they focus essentially on the two key components of BSC highlighted in this introduction: body ownership and sense of agency. As of now, two studies (Fang et al., 2019; Samad et al., 2015) proposed and empirically tested a Bayesian model of body ownership, and only one recent study (Legaspi & Toyoizumi, 2019) proposed an explicit formulation of a Bayesian model for agency, which was only tested on already existing data.

Due to the importance of Bayesian approaches to brain function for this thesis, we will review these papers in more detail in the following paragraph.

1.5.2. Bayesian Causal Inference models for BSC. A powerful conceptual framework with little empirical support?

All models proposed in these three studies belong to the family of Bayesian Causal Inference (Bayesian CI) models, which can be seen as an extension of Ernst and Banks' model of optimal multisensory integration. In order to be optimal, the rule of the squared inverse precision needs a subtle yet fundamental assumption to be verified: the two sources of information in different modalities need to originate from the same physical source. To make a simple example, audiovisual integration allows to more precisely estimate the source of a sound by combining visual and auditory information, but clearly this is only true if the object whose visual location is being integrated with auditory information *is* the true source of the sound. If, as in the classical ventriloquism effect, the source of the sound is not what I believe it to be, visual information is worsening instead of improving my estimate, and the optimal policy would be to rely on hearing alone (see Alais & Burr, 2004). For simplicity, this same source hypothesis was assumed to hold *a priori* in Ernst and Banks' study, which is certainly reasonable within a controlled experimental setup. In everyday life, multiple inputs in different sensory modalities are presented simultaneously, and it is therefore clear that, as a necessary step for successful integration, the brain needs to figure out which pairs of stimuli are to be integrated across sensory modalities. Bayesian Causal Inference (Bayesian CI) models address this binding problem and the subsequent integration of stimuli within the same probabilistic framework. The probability that two sensory events originate from the same physical cause is computed as a first step. Then, it is used to refine the perceptual estimate, by weighing information in different modalities not only by their reliability, but, critically, also by the probability that they are truly causing the event about which inference is being made. The relevance of these models for BSC becomes evident by applying this interpretative framework to body related information, for example visual and proprioceptive cues about hand location. In this case, the process of causal inference would compute whether proprioceptive information about my hand has the same origin as the hand-shaped object I am looking at, in order to estimate whether visual cues can be relied upon when estimating hand position. If the "same cause" hypothesis is statistically favoured, visual information is integrated with proprioceptive information, and it is reasonable to assume that, subjectively, the hand is perceived as one's own. In the opposite case, the hand is perceived as an external object, and proprioceptive information alone is used. In other words, an object perceived through external senses is embodied if sensory information about it is estimated to be statistically more likely to have the same physical origin as somatosensory information. In this view, self-identifying inference would pertain the visual modality only, implying that somatosensory information would be inherently perceived as self-related, or, to use Bermudez's words, immune to error through misidentification.

While this is certainly a sound assumption for adults, it seems also reasonable that this ability is acquired and not innate. In this sense, it is less clear when and how the ability to automatically self-refer somatosensory information, that makes it “special” for BSC, emerges in the developing brain.

Within the Bayesian CI framework, Samad and colleagues (2015) modelled body ownership based on statistical inference upon the spatial and temporal congruency of visual, tactile and proprioceptive information, and tested their model in a RHI-like setup. They predicted the typical range of visuo-proprioceptive disparities that would allow for the illusion to arise, and found it to be in line with existing literature. Furthermore, they predicted that the simple presentation of a rubber hand would elicit a shift in the position of the hand as perceived by proprioception, and confirmed this experimentally. The evidence presented in their work remains mainly qualitative, as the unisensory precisions (needed to compare the tolerated disparity range) were fixed for all subjects based on literature. Measuring these quantities would have instead allowed producing individualized predictions, and comparing them with empirical data. More recently, Fang and colleagues (2019) proposed another Bayesian CI model for hand ownership, applied this time to a setup similar to the moving RHI, and based on the manipulation of visuo-proprioceptive congruency in the spatial domain. Humans and macaques had to perform reaching movements with their real hand, concealed from their view, while a virtual hand was displayed moving in synchrony with their real hand, but visually shifted by variable amounts. The reaching error linearly increased with the amount of shift as long as disparity levels remained small, as if the perceived hand position was a weighted mean of visual and proprioceptive cues and in line with predictions of a forced fusion model of optimal integration. As disparity levels increased further, participants were increasingly guided by proprioception in performing their movement, and reaching errors decreased. This broke the forced fusion assumption and was instead in line with the predictions of the proposed Bayesian CI model. At small levels of disparity, the probability that the virtual hand is the same as the proprioceptive hand is close to one, yielding the same predictions as the forced fusion model. At larger disparities, the same cause probability decreased and the weight attributed to proprioception increased, towards a regime of pure proprioceptive estimation. As mentioned previously, the same cause probability can be seen as the mathematical counterpart of body ownership. The same cause probability (or putative ownership probability), which can be computed from reaching errors as a function of disparity, correlated with subjective ratings of ownership from human participants. This suggests that, indeed, same cause probability and subjective feeling of ownership are related. Moreover, neurophysiological recordings obtained from the premotor cortex of macaques performing the task revealed populations of neurons tuned either to the segregation or to the integration of visual and proprioceptive information. In trials in which integrating neurons were more active, the relative weight of proprioception was higher, and the ownership probability inferred from reaching errors was therefore higher, and vice versa. Arguably, the main limitation of this insightful study is the validation

of model predictions at the behavioural level. In order for cue integration to be Bayesian optimal, the weights attributed to different sensory modalities need to reflect the precision of the relative unisensory estimates. Therefore, the unisensory precisions need to be either experimentally manipulated (as in Erns and Banks, where different levels of noise were added to the visual input) or measured in separate unisensory tasks. In this study, instead, the unisensory precisions are left as free parameters to be fitted from behavioural data, weakening the claim that the behaviour resulting from the visuo-proprioceptive task is Bayesian optimal. In another recent study, Legaspi and Toyoizumi (2019) applied a similar framework to model sense of agency and its implications for the intentional binding task. In their model, the only relevant variables were the timing of an action, and of its putative sensory outcome (e.g. the beep in the classical intentional binding paradigm). The probability that the action and the outcome have the same cause is estimated based on the likelihood ratio that the action did or did not cause the outcome, given the temporal delay between the two events. When the same cause probability is higher than the different cause probability, sense of agency is elicited, and the temporal estimates for the action and its effect are attracted towards each other. This provides a mathematical model of sense of agency whose predictions are qualitatively in line with experimental data (mainly from Haggard's works on intentional binding). However, the model was only tested on pre-existing reports from literature, and no ad-hoc experiment was performed. Overall, experimental evidence suggests that Bayesian models of brain function may provide a powerful normative framework to describe the processes leading to body ownership and sense of agency. However, compared to the now widely accepted Bayesian models of multisensory integration, studies investigating the Bayesian nature of BSC experimentally are still scarce and only partially conclusive.

2. Thesis outline

This thesis work focuses on investigating how the two key components of BSC, body ownership and sense of agency, emerge from the multisensory integration of bodily and external signals, focusing on psychophysical, neurophysiological and computational aspects. Four main studies are presented and discussed extensively, two focusing on body ownership and two on the sense of agency. The first study presents a neural network model of multisensory integration in PPS, and its link with body ownership. The second study, also focusing on body ownership, is a pre-registered protocol. We propose a behavioural paradigm to rigorously test a Bayesian model of how body ownership emerges from multisensory congruencies, overcoming the limitations of previous studies. In the third study, we investigated the interplay between motor commands, sensory feedback, and neurophysiological signals in generating sense of agency in an intracortical brain machine interface. In the fourth study, we investigated the role of neural oscillations for the sense of agency within the setup used in the third study.

2.1. Study 1 - From statistical regularities in multisensory inputs to peripersonal space representation and body ownership: Insights from a neural network model.

Tommaso Bertoni, Elisa Magosso, Andrea Serino

The European Journal of Neuroscience.
2021 Jan; 53(2):611-636. DOI: 10.1111/ejn.14981.

Personal contribution: designed and implemented the neural network model, collected behavioural data, analysed the data and wrote the paper.

In Study 1, we used artificial neural networks to investigate how the integration of visual, proprioceptive and tactile inputs can lead to visuo-tactile integration in body-part centred coordinates (PPS representation), and how this property spontaneously leads to reproduce behavioural correlates of body ownership. The synaptic connectivity spontaneously tuned based on Hebbian learning during a period of simulated synapse maturation, induced by multisensory inputs reproducing the statistical regularities between body-related and external information. The network's multisensory neurons that responded to touch also developed overlapping visual and proprioceptive receptive fields. These spontaneously learned characteristics allowed reproducing key neurophysiological and behavioural properties of PPS representation, specifically the emergence of a hand-centred visuotactile representation. Moreover, we found cross-modal influences from visuo-tactile stimulation on proprioceptive encoding that could be mapped to behavioural correlates of body ownership in a RHI like setup. Synchronous visuo-tactile inputs induced a shift of the proprioceptive position, as encoded in multisensory neurons, towards the location where visuo-tactile stimulation occurred. This effect is in line with the proprioceptive drift observed in the RHI (and the IHI), arguably the only correlate of pre-reflexive sense of ownership that can be investigated in an artificial neural network. Furthermore, the plausibility of the proposed network architecture was tested through a dedicated behavioural task, showing that visual stimuli in hand centred coordinates (as signalled by proprioception) modulate tactile reaction times.

2.2. Study 2 - The self and the Bayesian brain: testing probabilistic models of body ownership

Tommaso Bertoni & Giulio Mastria, Henri Perrin, Boris Zbinden, Michela Bassolino, Andrea Serino

Under review, Nature Communications

Personal contribution: designed the study, developed the mathematical model, analysed the data and wrote the paper.

Study 2 is a pre-registered protocol, currently under review. The protocol presents a Bayesian CI model for hand ownership and a set of ad hoc tasks aiming at increasing the evidence base for Bayesian accounts of BSC. Specifically, we will try to overcome the limitations of previous studies by rigorously testing whether the process leading to body ownership is optimal in a Bayesian sense. As mentioned in the introduction, empirical evidence in this sense is still scarce and not satisfactory. The model computes the probability of hand ownership based on the spatial and temporal disparities between visual and proprioceptive information about hand location and movement, and produces estimates of the perceived hand position as a function of the disparity level, and of the associated ownership levels. The model will be tested in a virtual-reality reaching task similar to the one proposed by Fang (2019), but with the addition of temporal delays to visuo-proprioceptive spatial disparities. Importantly, model parameters fitted from reaching errors in this multisensory task will be compared to direct measures in a set of dedicated unisensory tasks. The crucial measures therefore consist of the spatial and temporal uncertainties of proprioceptive and visual estimates, allowing to rigorously address the question of Bayesian optimality. The set of tasks and computational analyses has been optimized through extensive simulations and tested on a smaller pool of healthy participants (N = 10 as a replication of Fang's study, N = 2 to demonstrate the practical feasibility of our task). The preregistered study forecasts a total of 40 healthy participants to be tested to guarantee sufficient statistical power (95 % with $\alpha = 0.05$).

2.3. Study 3 - Sense of agency for intracortical brain machine interfaces

Andrea Serino & Marcie Bockbrader, Tommaso Bertoni, Sam Colachis, Marco Solca, Collin Dunlap, Kaitie Eipel, Patrick Ganzer, Nick Annetta, Gaurav Sharma, Pavo Orepic, David Friedenberg, Per Sederberg, Nathan Faivre, Ali Rezai, Olaf Blanke

Under review, Nature Human Behaviour

Personal contribution: pre-processed the data, conceived and performed analyses, visualization.

Study 3 aims at investigating the behavioural and neurophysiological correlates of sense of agency for bodily actions. The study was carried out in collaboration with the Ohio State University, in a tetraplegic participant using an intracortical brain machine interface that allows restoring functional control of the upper limb. The BMI system reads cortical signals from a chronically implanted array in the right hand area of M1, which are decoded and translated online to functional hand and wrist movements through neuromuscular electrical stimulation (NMES). This setup allows decoupling motor commands and actual body movements, which clearly would not be possible in healthy individuals. Leveraging on this unique opportunity, the participant's sense of agency for BMI generated body movements was manipulated and assessed with either congruent or incongruent somatosensory (as provided by the NMES system) and visual feedback (as provided by a virtual reality animation). As predicted by theory, feedback congruency strongly modulated the participant's sense of agency. Moreover, we found that feedback congruency and sense of agency could be decoded from M1 multiunit activity, and that congruency signals were significantly stronger than visual congruency and pure agency signals. Surprisingly, we found that somatosensory feedback modulated M1 signals with equal or stronger intensity than motor commands themselves: under NMES stimulation, M1 activity with incongruent somatosensory feedback reflected the implemented movement more than the intended one. Finally, we investigated whether sense of agency affected the performance of the BMI decoder, and found evidence suggesting that BMI actions can be implemented more reliably when sense of agency is high.

2.4. Study 4 - The phase of pre-movement mu oscillations predicts sense of agency for an intracortical brain machine interface

Tommaso Bertoni, Marcia Bockbrader, Sam Colachis, Marco Solca, Jean Paul Noel, Nathan Faivre, Ali Rezai, Olaf Blanke, Andrea Serino

In preparation

Personal contribution: conceived and performed the analyses, wrote the paper.

Study 4 focuses on data from the setup presented in Study 3, as well as from another experiment in the same participant, which was analysed specifically to uncover the role of neural oscillations in sense of agency. In the first experiment, we focused on a subset of conditions where sensory feedback weakly correlated with agency ratings, so to maximise the contribution of endogenous oscillations. In the second experiment, we studied the effect of neural oscillations in a BMI version of the Libet paradigm. The participant had to report the perceived timing either of actively performed BMI actions, or of passive movements randomly generated through the NMES. Behavioural results showed that active movements were perceived as occurring earlier in time compared to passive movements, suggesting that agency leads to an anticipation of the perceived movement onset in this setup. Therefore, we used movement timing reports within the active condition as an implicit measure of sense of agency. We found that the phase of mu (8 Hz) oscillations up to 570 ms before movement onset predicted both the explicitly and implicitly assessed sense of agency, with a consistent phase relation across the two measures. When looking at multiunit activity as a function of neural oscillations, we found that the optimal phase for sense of agency is compatible with the phase window in which spikes are maximally facilitated. However, despite it strongly affected subjective (and implicit) ratings of agency, the mu phase at movement onset did not significantly affect subsequent patterns of multiunit activity at the population level. We concluded that a possible explanation is that mu oscillations influence the sense of agency not by directly affecting M1 activity, but by biasing the connectivity between M1 and frontal or parietal regions associated with the sense of agency.

3. Discussion

3.1. From receptive fields to body ownership (via embodiment)

In Study 1, we showed how the presence of tactile information could “align” the artificial visual and proprioceptive receptive fields learned by our model. The network was trained with a combination of visual (about external stimuli), proprioceptive (about hand position) and tactile inputs reproducing the ecological associations between these modalities (i.e., touch was provided when the position of external stimuli coincided with the one of the hand). Neurons that were originally designed to potentially respond to any type of stimulation, and then have learned to respond to touch, concurrently developed overlapping visual and proprioceptive receptive fields. Conversely, neurons that developed weak tactile responses, also developed anti-overlapping visual and proprioceptive receptive fields (Fig. 2c-d, Study 1). In synthesis, this is key to the emergence of a body-part centred visuotactile map, subtending PPS representation (Fig. 3c, Study 1). Despite the individual response properties of a whole population of neurons with complex receptive fields may appear hard to represent intuitively, their emerging collective behaviour can be actually grasped quite easily. The overlapping visual and proprioceptive RFs of touch-responding multisensory neurons implies that they can also be activated when visual stimuli are close to the proprioceptively encoded hand location. Therefore, even in the absence of tactile stimulation in the unisensory population, tactile information will be encoded in that subset of “tactile” neurons in the multisensory layer, whenever a visual input is presented close to the proprioceptively (or visually) encoded hand position. This same property of the network implies that, if touch is provided simultaneously with a visual stimulus far from the hand, the neurons that are preferentially activated are those that code also proprioceptively for that far location in space. Therefore, proprioceptive information in the multisensory layer will shift towards the position of visual stimulation (Fig. 2e, Study 1). We interpreted this as the in-silico analogue of the proprioceptive drift induced in the invisible hand illusion (or the rubber hand illusion, in a further version of the network including visual information about hand position).

There has been considerable debate on whether the proprioceptive drift can be truly used as a proxy of body ownership. Indeed, it has been noted that proprioceptive drift can occur in stimulation conditions that do not elicit subjective body ownership as assessed through questionnaires, such as asynchronous stroking in a RHI setup (Rohde et al., 2011), or simple observation of a rubber hand (Samad et al., 2015). Nevertheless, the amount of drift does correlate with the perceived strength of the illusion (Guterstam et al., 2013; Tsakiris & Haggard, 2005), and it would therefore seem that even if illusory ownership is not the only possible cause of proprioceptive drift, the latter is amongst the robust behavioural correlates of the former. More importantly, proprioceptive drift is arguably the only correlate of illusory body ownership that can be tested in a neural network model, or at least by far the simplest. As a simple example, putting a neural-network in the condition to provide subjective reports would require, at the very least, a mechanistic model of the neural principles of semantics,

which is far beyond our current understanding. As just mentioned, the mechanism that allows replicating proprioceptive drift in our model is instead rather simple, provided some familiarity with the principles of artificial neural networks.

Indeed, one might even argue that a mechanism resting on a multisensory population of neurons with overlapping visual and proprioceptive receptive fields is *too simple* to explain an apparently complex experiential phenomenon such as body ownership. In order to gain a deeper insight into this matter, it may be useful to look for rigorous definitions of the key concepts in philosophical literature, starting from embodiment. This term is often erroneously used as a synonymous of body ownership, but it is more appropriate to use it as an intermediate step towards a rigorous definition of body ownership, through functional properties. According to De Vignemont (2011), “E is embodied if and only if some properties of E are processed in the same way as the properties of one’s body”. The fact that this is a functional definition, based on information processing principles, makes it suitable to be applied to neural network modelling. Under this perspective, the feeling of ownership would refer to the phenomenal experience that arises for embodied objects, under certain circumstances. Therefore, embodiment would be a necessary component for the existence of ownership, but not a sufficient one. As a classical example, a tool can be embodied according to De Vignemont’s definition, but ownership is not typically experienced for tools, with some notable exceptions such as prostheses. Arguably, the key to the fact that embodiment is not sufficient for ownership lies in the fact that the rigorous definition of embodiment (as well as its intuitive notion) allows for the concept of partial embodiment, i.e., when only a subset of an object's properties is processed as if it was a body part. For example, embodiment can concern motor properties (e.g. expecting an object to move upon a motor command as if it was part of the body), or perceptual properties (e.g. expect touch when the object is “stimulated” as indicated by visual cues). Full embodiment would then refer to the case in which *all* of an object’s properties are processed as if it was part of the body (De Vignemont, 2011). Despite not having been explicitly discussed by the author, it is tempting to assume that full embodiment leads inevitably to body ownership.

Applying this framework to our model, we would conclude that it exhibits partial embodiment under two main aspects. First, the prediction of touch when visual stimulation is provided close to a body part. Second, proprioceptive drift towards the location of stimulation if touch is provided simultaneously with visual stimulation at a location shifted from that of the hand. Our work aimed at conceptually demonstrating how a subset of key aspects of BSC can emerge from a neural network learning its synaptic connectivity in a biologically plausible manner, while being simple enough to allow an intuitive understanding of its mechanistic principles. Clearly, this implies that there are many other aspects of embodiment that are not accounted for. For example, we do not model motor aspects, or the role of the visual appearance of body parts in determining their integration into body representations. In order to account for motor properties, the network should at least possess a

proper temporal dynamics (e.g., a recurrent architecture allowing to integrate present and past states), and an input encoding efference copies of motor commands. In order to account for the visual recognition of body parts, visual inputs should be encoded as natural images in retinotopic coordinates. In sum, the simplification of the training inputs and of the network's architecture and dynamics is possibly too extreme to allow an attempt of rigorously validating model predictions, for example by comparing simulated activity in the multisensory layer to response properties of parietal multisensory neurons.

In this paragraph, we will try to argue how these limitations might be overcome by generalizing our approach to a richer architecture and set of training inputs, without modifying the key principle: the Hebbian learning of statistical regularities in multisensory inputs. For example, if the network could successfully be trained to learn the joint probability of tactile, proprioceptive and visual inputs, with the visual encoding consisting of natural images, it would be likely able to reproduce subtler properties of embodiment. For example, proprioceptive drift with visuo-tactile stimulation would occur only if the stimulated visual object was shaped as a hand. Under the statistical perspective, this would be because the network has learned that only in this case the proprioceptive position of the hand typically coincides with the location of visual stimulation. Interestingly, the "visual form" constraint for embodiment was originally proposed by Tsakiris (2010) as a top-down cognitive factor, opposed to "Bayesian" bottom-up factors. In our hypothetical model instead, the same phenomenon would be explained mechanistically as the bottom-up result of statistical computations. Always in principle, the approach could be generalized to all aspects of embodiment for a body part, and even to other aspects of information processing that are relevant for bodily self-consciousness, such as whole body ownership, self-location and possessing a first person perspective. The information processing properties of self-location arguably rest on the learning of associations between self-related cues of orientation and motion (as cued by motor commands and the vestibular system) and external (mainly visual) cues about one's position with respect to the environment. Again, the fundamental principle lies in the associations of information conveyed by self-related and external sensory pathways. Crucially, these associations are arguably learned from the natural statistics of the environment, since they would be too high-dimensional and complex to tune to be genetically hard wired (for a beautiful conceptual argument in this sense, see Hinton, 2014). Indeed, empirical evidence goes in the same direction, showing a gradual development of body representation throughout human life (Cowie et al., 2018; Pagel et al., 2009; Slaughter & Brownell, 2011), and its ability to plastically adapt to novel associations in multisensory inputs likely due to short (Held & Freedman, 1963; M.R. Longo & Serino, 2012; Martel et al., 2016; Redding et al., 2005) and long (T. R. Makin et al., 2015; Ziemann et al., 1998) term plasticity. Autoencoders, such as the restricted Boltzmann machine used in our work, are neural networks designed to learn a compact representation of its inputs in an unsupervised manner. They can be trained based on a simplified version of Hebb's rule, and therefore they constitute promising candidates to model the process of

association learning in a biologically plausible artificial neural network. Moreover, they offer the valuable theoretical advantage of providing a rigorous interpretative framework of the classical notion of “association learning” in statistical terms. Indeed, the learning rule used in our network is not only an approximation of Hebbian plasticity, but it can be shown that, if the training is successful, it should lead the network to be a generative model of the joint probability of its multisensory inputs. This characteristic provided our network with information processing properties that are thought to be the functional counterpart of some key phenomenal components of BSC. Here, we argue that “simply” applying the same principle to a richer set of multisensory inputs, in an appropriate form of autoencoder, might allow to reproduce “all” the information processing properties required for full embodiment, and possibly BSC. At the functional level, the success of this operation would be a matter of computational efficiency more than a conceptual one. In this sense, the continuous advancement of techniques, and the growing interest in the intersection between deep learning and neuroscience might help overcoming technical limitations. A sufficiently complex autoencoder could be then used to formulate subtle predictions about behaviour and the organization of receptive fields, and these predictions could be empirically tested to validate a full mechanistic account of body ownership. Still, there is no doubt that the nature, or even the existence, of a *necessary* link between certain properties of information processing and phenomenal aspects of BSC might remain the argument of eternal debate. Nevertheless, these questions are not subject of scientific investigation, nor can be discussed here, as they pertain the hard problem of consciousness (Chalmers, 2018; Nagel, 1974).

3.2. The role of the motor cortex in encoding sensory feedback, intentionality and sense of agency

As its name may suggest, the motor cortex has been traditionally regarded as a mere executor of movements. Indeed, most of the direct projections to the cortico-spinal tract in the human brain originate from the primary motor area (Porter & Lemon, 2012). Nevertheless, sensory feedback has been long known to be encoded in the motor cortex, and its sensory properties have received increasing attention in recent years (Hatsopoulos & Suminski, 2011). Thanks to the unique setup used in Study 3 (and 4), we could investigate the interplay between motor commands and visual/somatosensory feedback, while being able to manipulate the congruency between motor intentions and their actual bodily consequences. The participant performed BMI actions, which were translated to various combinations of congruent and incongruent visual (through virtual reality) and somatosensory feedback (through muscular stimulation). For each condition, we assessed the participant’s sense of agency. First, we found that the congruency between intended actions and executed movements could be decoded from local field potential (LFP) and multiunit signals in M1, for both visual and somatosensory feedback. Despite the spinal lesion also partially affected afferent pathways, the decoding performance was higher and occurred at earlier temporal delays for somatosensory feedback. In line with this observation, we found that the output of the BMI decoder

was more strongly affected by somatosensory than by visual feedback (Fig. 5, Study 3). When incongruent somatosensory feedback was provided, the decoded strength of the intended movement sharply decreased, whereas it was basically left unchanged by incongruent visual feedback. We more closely studied the neural causes of this effect, by investigating trials with incongruent somatosensory feedback, where one movement was intended but a different movement was actually implemented through NMES stimulation. After movement onset, the population-level activity quickly came to resemble the activity elicited by congruent execution of the actually implemented movement, rather than the intended one (Fig. 6, Study 3). In other words, multiunit population activity seemed to reflect the NMES-realized movement more than the willed one. In contrast, when comparing congruent and incongruent visual feedback, we found little difference in multiunit activity.

These findings have interesting implications for the field of BMI control, as they suggest that decoding systems based on primary motor cortex activity need to be robust to the effects of (potentially unpredictable) somatosensory feedback on population activity. As already noted in other works, neurons in the primary motor cortex exhibit an extremely wide range of responses to sensory feedback, with cells preferentially coding for motor commands, sensory feedback or a combination of the two (Suminski et al., 2010). It has been observed that sensory feedback can even trigger “covert” motor commands, which in a BMI system would need to be appropriately distinguished from actual motor commands to avoid unintentional activation. Indeed, this is in line with the observed increase of the decoder output for the executed movement during somatosensory incongruent stimulation. It is worth noting that the participant was still typically able to activate the correct decoder in case of incongruent somatosensory feedback, despite population activity was similar to the one elicited by the opposite movement. This likely indicates that the BMI decoder is already somehow picking the most robust motor command features, by selecting channels that mainly contain “intention-tuned” cells. Our results may therefore inspire further technical research aiming at optimizing decoder robustness. Interestingly, literature reports a strong modulation of M1 signals by both visual and somatosensory feedback, while in our case the effect seemed to be vastly limited to somatosensory inputs. This may be due to the fact that previous studies mainly used passively implemented movements (Flament & Hore, 1988; Herter et al., 2009; Pruszynski et al., 2011; Suminski et al., 2009), while here we directly assessed the effect of executing (or observing) one movement while the opposite movement was intended. Our results suggest that, under this conditions, only somatosensory feedback has the capability of partially “overwriting” the motor commands encoded in M1.

Besides the implications for the BMI field, this result opens questions that are of interest for the neuroscience of agency and intentionality. Behavioural results clearly showed that incongruent feedback strongly diminished sense of agency, regardless of its visual or somatosensory nature.

Although we did not directly assess the associated phenomenology, we can assume that the participant's subjective feeling of motor intention was left unchanged by incongruent feedback. Otherwise, passive movements would be able to give rise to false intentions, and incongruent feedback would not elicit negative agency, as it would be congruent with these "overwritten" intentions. Nevertheless, activity in the participant's motor cortex in the case of incongruent somatosensory feedback looked as if the he attempted to perform the executed movement, and not the cued (and truly intended) one. Indeed, coding of intentionality is traditionally thought to take place in more frontal areas, such as the SMA, whose electrical stimulation is known to elicit a conscious "urge to move" (Fried, 1993). On the other hand, the same results confirm that M1 does not behave as a mere executor sitting at the bottom of a top-down chain originating in prefrontal areas, but it encodes a complex combination of efferent and afferent information that is also used to compute agency.

Indeed, intriguing questions also arise when turning our attention more directly to the sense of agency. When looking at local LFP (and, to a lesser extent, multiunit) activity, after regressing out the effect of sensory feedback congruency, we found that we could decode the subjective sense of agency. This suggests that the motor cortex may be directly involved in the network of areas that generate sense of agency (Fig. 4B, Study 3). Nevertheless, we were able to discriminate feedback congruency with far greater accuracy compared to "pure" sense of agency. Indeed, it would not be surprising if the actual computations leading to sense of agency took place somewhere else in the brain, and the involvement of the motor cortex was only as a downstream area, encoding a combination of afferent and efferent information. The analysis presented in Study 4 is probably particularly informative in this sense, as it shows that, when removing the variability due to feedback congruency, the most robust correlate of sense of agency in M1 signals is the pre-movement phase of the sensorimotor mu rhythm. In the last fifteen years, it has become increasingly clear that low-frequency neural oscillations are related to the gating of information exchange in long-range connections within the brain (Fries, 2005, 2015). Particularly informative in connection to our result is a study by Hanslmeyr and colleagues (2013), combining EEG for detecting neural oscillations and fMRI for the spatially detailed investigation of functional connectivity. Using an illusory contour detection task, they found that high-level visual processing performance was affected by the phase of 7 Hz oscillations in occipital regions a few hundred milliseconds before stimulus onset. Furthermore, the phase of these oscillations at movement onset determined the subsequent connectivity between low-level visual areas and the right intraparietal sulcus, assessed via fMRI. When the stimulus occurred close to the optimal phase for behavioural performance, the connectivity between these areas was significantly stronger. Transposing this idea to sensorimotor processing, the phase of the mu rhythm might affect the connectivity between M1 and more frontal areas, coding for intention, and/or parietal areas encoding the sensory feedback, allowing for sensorimotor comparisons to take place. In this sense, the primary motor cortex would act as a crucial, and

possibly overlooked hub in the network of regions that integrate afferent and efferent information to generate sense of agency, whose complex interplay is orchestrated by neural oscillations. In our study, only data from one cortical location was available, not allowing to prove that the proposed mechanism is involved in the observed modulation of agency, but this is indeed an intriguing hypothesis. We are currently planning further studies to investigate exactly which brain regions are involved in this process, and whether and how the sensorimotor mu rhythm influences the connectivity between these areas (see paragraph 3.5.).

3.3. Predictive mechanisms for sense of agency

In study 4, we described a reliable endogenous neural marker of sense of agency, occurring before the movement takes place. Both explicitly and implicitly assessed sense of agency was higher when the mu phase at movement onset was near π, the negative trough of oscillations. This is per se a novel and interesting finding, although observed in a single participant, and further investigations should aim at replicating it in a larger cohort. Nevertheless, the key oscillatory signal is most likely subjectively inaccessible, and faster than most cognitive processes. It is therefore reasonable to assume that our finding reflects a general mechanism in human sense of agency, and not some idiosyncratic cognitive bias present only in our participant. At the conceptual level, this result is also relevant for the debate between predictive and postdictive processes in sense of agency. As mentioned in the introduction, the key to such debate is not merely relative to the timing of cues with respect to the movement. Instead, the main focus is about the role of low-level sensorimotor inputs, likely occurring on a fast timescale around movement onset, as opposed to slower cognitive processes. Then, the predictive view would put a strong emphasis on prediction errors, based on detailed computations on afferent and efferent information. The postdictive view would instead favour the overall matching of intentions and external events in lesser sensorimotor detail, which is cognitive in nature and subject to contextual cues and biases such as the perceived value of the action. Here, we found that a low-level, rapidly changing neural quantity such as the phase of mu oscillations was reliably predicting subjective ratings about agency, seemingly providing evidence in favour of the predictive view. Still, the final step in the generation of such ratings is the result of a long timescale, cognitive process, taking place during several seconds after the movement has occurred. Therefore, the neural processes that lead to the modulation of agency judgements as a function of the phase of mu oscillations cannot be known exactly.

The subject becomes even more complex when considering it under the phenomenological perspective. Movements taking place in the optimal phase of mu oscillations may almost instantaneously be accompanied by a stronger pre-reflexive feeling of agency, or simply undergo a different subesquent processing, which biases the production of explicit ratings retrospectively. For example, the mu phase may affect the saliency of sensory feedback, and bias explicit ratings while only indirectly affecting the experience of agency. In any case, the question about *when* exactly the

subjective experience of agency arises, and how mu oscillations influence it, remains inherently philosophical, since only fallible subjective reports allow accessing such information. Our results show that a pre-movement process, which is linked to low-level sensorimotor processing and completely impermeable to conscious awareness, can at the very least bias subjective reports about the feeling of agency, if not directly affect its pre-reflexive qualities. Therefore, the idea that the phase of sensorimotor oscillations at movement onset has an direct effect on the subjective experience of agency is not only fully compatible with our observations, but it arguably constitutes the simplest interpretation. Moreover, the same effect was observed when using an implicit measure of sense of agency, based on the perceived time of the BMI generated movement similarly to the intentional binding paradigm. Therefore, while agency judgements considered in our study may indeed be influenced by retrospective cognitive factors, they likely also inform about genuinely pre-reflexive phenomenological content. Our data suggests that such content is influenced by endogenous, low-level features of sensorimotor processing, occurring before the movement takes place.

3.4. Are Bayesian models of brain function useful for BSC?

In this thesis, we addressed the theme of bodily self-consciousness by touching upon different of its aspects through a variety of techniques. Nevertheless, one main thread arguably emerges as a possible interpretative framework: Bayesian approximations of brain function. In a nutshell, the idea is that the process of congruency detection underlying several key aspects of BSC is performed by the brain as if it was an ideal Bayesian observer, performing statistical inference.
In Study 1, the key idea behind the learning rule of the network was that it should allow the network to behave as a good generative model of its sensory inputs. In mathematical terms, this means that the network should learn to approximately reproduce the joint probability distribution of neural activity in its input layers. In neuroscientific words, this translates to the idea that the network's spontaneous activity in absence of inputs, or its "dreams", in Geoffrey Hinton's words, should resemble the "ecological" sensory stimulation provided during training. Therefore, the idea that the brain might work as an inference machine is already built into the type of network that we chose.

In the studies about sense of agency, the link with Bayesian approaches to brain function is less evident, but not less interesting, and it starts from a peculiar observation that can be made in Study 4. The subset of yielding the key effect contained a mixture of trials with congruent visual and incongruent somatosensory feedback (V+/S-), and incongruent visual and congruent somatosensory feedback (V-/S+). The specific relation between the phase of mu oscillations and sense of agency appeared to be the same in V+/S- and in V-/S+ conditions (Fig. 3f, Study 4). In both cases, sense of agency was highest for movements occurring when the phase was near π, suggesting that the underlying mechanism is independent from the input modality. The mu phase may be therefore best described as a general gating mechanism for the causal binding between motor commands, coded in M1, and events in the external world, coded in sensory cortices. Indeed, this interpretation is

already in line with the rich literature studying the role of neural oscillations in orchestrating brain connectivity. Interestingly, the role of neural oscillations in binding information across brain areas has been recently empirically connected to the Bayesian framework, offering an intriguing interpretation to our result. In a brilliant EEG study, Rohe and colleagues investigated cross-modal interactions between visual and auditory stimuli, by presenting a variable number of visual flashes and auditory beeps in rapid succession, and asking participants to evaluate the total number of visual or auditory stimuli that they perceived (Rohe et al., 2019). As already known from classical studies, the number of flashes can bias the perceived numerosity of beeps, and vice versa (Shams et al., 2000). They found participants' behaviour to be well described by the predictions of a Bayesian CI model, where the perceived number of visual (or auditory) stimuli is estimated as a function of both visual and auditory inputs. The strength of cross-modal influences is modelled as being related to the probability that the stimuli in the two modalities have the same cause. This probability depends on the true numerosity disparity (the larger the disparity, the smaller the probability), and on an individual "prior", governing the overall tendency to attribute events to the same cause. Surprisingly, such prior was not found to be fixed in time, but to depend on the phase of alpha oscillations immediately preceding the first stimulus, and on their power. Similarly, in our study the phase of the mu rhythm may act as an oscillating sensorimotor binding prior, modulating the perceived probability that a given type of sensory feedback has been generated by our motor intentions. This interesting result provides a potential connection between physiological mechanisms and processes of statistical inference in the brain in the most general sense, and our data suggest that the same interpretative framework may be applied also to the emergence of sense of agency. Nevertheless, it also leads to an intriguing and deep epistemological question, tackling the complex relations between description levels that characterize modern neurosciences.

3.4.1. The challenges of scaling up Bayesian models of brain function from multisensory integration to causal inference

In order to introduce the issue, it may be helpful to start with a technical and theoretical introduction on some key aspects of Bayesian models. Bayesian approximations of brain function were originally conceived as approaches aimed a describing the input-output relation of neural systems in terms of formal probabilistic operations. In this sense, the brain itself could be treated as a black box implementing Bayesian computations, since the aim of such theories is to provide a mathematical equivalent of the effective operations carried out by the brain, and not of *how* they are actually implemented by the firing of neurons. Indeed, Ernst and Banks themselves suggested a qualitative model of how neurons may actually implement Bayesian inference, but this was not the main focus of their paper. In the following years, several works followed with the specific aim of proposing biologically plausible models of statistical inference in the brain (Limanowski & Blankenburg, 2013; Ma et al., 2006; Penny, 2012; Rao & Ballard, 1999). These models represent interesting hypotheses, but the path towards a rigorous understanding and empirical demonstration of the actual neural

mechanisms of statistical inference in the brain remains long. Arguably then, the success of Bayesian approximations to brain function is still anchored to behaviour. The key of their success is not only that they provide a compact description of experimental data. This feature would be shared with any alternative model fitting behavioural data sufficiently well, a task not so hard to achieve given the large variability and the relatively small sample size in typical human psychophysical studies. Instead, what arguably makes Bayesian models of behaviour so appealing is that they have a clear, almost self-evident evolutionary motivation: they describe “by definition” the most advantageous policy in a noisy environment processed by a noisy sensory system. In other words, what they may truly inform us about is that evolution must have shaped the brain so that it approximates the behaviour of an ideal Bayesian observer, or rather it provided it the tools to efficiently learn to do so. Indeed, this idea provides a predictive tool allowing to interpret human behaviour with an unprecedented level of generality. However, practically applying this framework to situations more complex than pure multisensory integration presents some non-negligible challenges.

Arguably, the fundamental ingredient to define a Bayesian optimal policy is its underlying generative model (see e.g. Körding et al., 2007 for a practical example). In short, the generative model consists of the statistical properties of sensory inputs (the distribution from which physical stimuli are drawn) and their neural representation (the distribution of neural activity that such stimuli elicit). For example, in Ernst & Banks’ model of visual-haptic integration, the generative model assumes that the visual and haptic representations of the size of the stimulus are drawn from Gaussian distributions with the same mean, and different variances, reflecting the intrinsic precision of the respective sensory modality. Provided that the generative model corresponds to the true “ecological” probability distribution of sensory inputs and their neural representation, the derivation of their expected mutual relations and of optimal estimates through Bayes theorem is unique and (besides mathematical difficulties) straightforward. In order to be able to make predictions about behaviour, the definition of a generative model is therefore a necessary step, and those predictions will only hold if the generative model is a good approximation of the true stochastic process underlying stimulus generation, sampling and encoding. Taking again Ernst & Banks’ study as an example, we can notice that in their case there is very little ambiguity about how the generative model should be structured. The Gaussian assumption is extremely reasonable in almost all complex phenomena, and it is equally reasonable to expect the mean encoded size to be the same in both modalities, as it would not make sense to systematically perceive objects as larger in one modality than the other. The variance of those Gaussians remains as the only free parameter, and it can be simply measured by asking subjects to perform unisensory estimates and measuring their dispersion. All this comes at a price: the forced fusion assumption, which actually moves the problem to expecting that the brain is somehow sure that the sensory information being integrated comes from the same physical source. The forced fusion assumption can be overcome with Bayesian CI models, but the definition of the generative model becomes less straightforward and compelling, as it contains an additional, more

elusive free parameter: the prior probability of common cause. In principle, this should reflect the *overall* probability that the two considered stimuli arise from the same cause. The prior can therefore depend on the nature of the stimuli and on the context, and it is optimal as long as such dependence reflects changes in the *actual* same cause probability. For example, it makes sense to expect a higher audiovisual binding prior when seeing someone hitting a bell, and hearing a bell sound, than when hearing a car honk instead. Similarly, the prior can be expected to be lower if one is explicitly told that a distant loudspeaker may be producing the sound. The main challenge here is that, unlike the probability distributions appearing in Ernst & Banks' model, priors are arguably hard to quantify "from first principles", and are typically left as free parameters to be fit from behavioural data. In sum, the strength of Bayesian models lies in the fact that there is a clear evolutionary motivation for behaviour to be optimal. However, the very definition of optimal behaviour becomes increasingly elusive as the complexity of the modelled context grows, and more "priors" need to be postulated. This poses a significant epistemological challenge to the Bayesian brain hypothesis.

3.4.2. Some practical examples from the literature and our studies

In order to more concretely illustrate this epistemological issue, let us consider again Rohe's 2019 study and their "oscillating prior". As mentioned, in order to be optimal, priors need to reflect actual changes in probabilities, which makes the interpretation of the observed fluctuations in the causal prior rather puzzling. Indeed, it is hard to imagine that the actual probability that two external stimuli originated from the same cause can depend on endogenous neural oscillations. In other words, alpha oscillations rhythmically modulate the prior, but the optimal prior is likely to be constant at fixed experimental conditions, and, in any case, it should not fluctuate with neural oscillations. Moreover, since the alpha power was also found to be correlated with the binding tendency, the authors investigated whether it was influenced by the history of previous trials' congruency, which in an optimal integration framework would be expected to modulate the causal prior. Surprisingly, they did not find the hypothesized correlation. In Study 4, although we did not fit behavioural data with a Bayesian model, we provided further evidence that low frequency oscillations may modulate the overall binding tendency, suggesting that this may be a general mechanism in the brain. Clearly, a possibility is that there is no deeper meaning: the "oscillating prior" may simply represent a small departure from optimality, due to the specific way in which Bayesian inference is implemented in the brain. In this view, the quest for optimality would still be the main driving force that shapes brain function. Even then, though, the ability of the model to fit changes in behaviour through adjustments of the prior carrying no meaningful information should suggest some concern. In the worst case, it may as well be that Bayesian CI just happens to be an empirical approximation that fits well the data (because of the many free parameters), but does not truly inform us about the driving forces that shaped brain function through evolution.

Another interesting example of this interpretative issue emerges in Study 1. When investigating the drift in proprioceptive estimates induced by visual stimuli, coupled with tactile stimulation, we found that cross modal influences decreased at large amounts of visuo-proprioceptive disparitiy. This result closely resembles the predictions of a Bayesian CI model, and behavioural results about visuo-proprioceptive integration such as those presented by Fang and colleagues in a reaching task (Fang et al., 2019), or by Noel and colleagues for a position estimation task (Noel et al., 2018). Nevertheless, assuming a perfect learning of the underlying generative model, the network should have produced the forced fusion model instead. This is because, during training, the true hand and external stimulus positions always coincide when tactile inputs are provided, so the network should have learned that the common cause probability is always one in this case. Therefore, there is no reason to expect the network to be truly performing Bayesian CI. The observed decrease of visuo-proprioceptive binding at large disparities is more likely to be the “accidental” result of the departure from ideal (forced fusion) behaviour in the network, when presented with inputs that are very different from training inputs (i.e.: at large disparities). Indeed, the non-linear activation function of multisensory neurons, coupled with the visuo-proprioceptive overlap of receptive fields, may lead to this result without being in any way related to Bayesian CI.

Our aim here is not to question the overall epistemological value of the Bayesian brain hypothesis, as this interesting topic has been more specifically addressed in more specialized works (Colombo & Wright, 2018; van Es & Hipolito, 2020; Williams, 2020). Karl Friston himself, one of the major players in developing theoretical models within the Bayesian brain approximation, allegedly stated that the whole framework contains an element of tautology, similarly to the natural evolution theory. Rather than a fully developed falsifiable scientific theory, the Bayesian brain hypothesis can be therefore be seen as a useful way to regard brain function, upon which new theories can be built. The fact that Bayesian models can always be adapted to explain new data by modifying their generative model, and that the inference performed by the brain is necessarily approximated, makes it hard to falsify the Bayesian brain hypothesis as a whole. As we mentioned, the claim that the brain performs approximated inference poses remarkable epistemological challenges when the object of investigation is far more complex than bivariate computations typically studied in multisensory integration. In more complex and high-dimensional scenarios (arguably constituting the majority of realistic cases), such as visual recognition, it is very hard to rigorously compute what optimal inference should look like. Therefore, and actual brain function may as well be a very loose approximation of true optimal inference. Then, in case such looser approximations are still accepted as being coherent with the Bayesian brain hypothesis, one can argue that the whole framework just reduces to common sense, as extreme departures from optimality would inevitably be not suitable for survival.

In our opinion, the issues raised here do not deplete the Bayesian framework of its value for neuroscientific investigation. Rather, its flexibility suggests the need for a constant monitoring of its specific applications to given contexts (in our case, BSC). In this sense, a key measure is not only how well a model describes the data, but also how hard it would be to describe the data outside of its framework. Beyond the ultimate epistemic value of the whole Bayesian framework, which may be hard to assess, this approach allows at least to evaluate how credible its applications are in their current formulation, and whether and how they can be further tested and improved.

3.4.3. Testing Bayesian models of body ownership: potential outcomes of the pre-registered study

Classical forced fusion models of multisensory integration extensively described experimental data, and provided surprising new predictions about the optimal reduction of variance in multisensory estimates which hardly any other simple model could convincingly produce. Unfortunately, the same cannot be said for Bayesian models of key components of BSC, namely body ownership. Despite the amount of theoretical works applying a Bayesian framework to the study of BSC (Apps & Tsakiris, 2014; Blanke et al., 2015; Limanowski & Blankenburg, 2013; Moutoussis et al., 2014; Seth, 2013; Seth & Tsakiris, 2018), only two studies (Fang et al., 2019; Samad et al., 2015) attempted to demonstrate this empirically. In both works, optimality was not rigorously proven, as the unisensory precisions were not directly measured or manipulated. This would have allowed, for example, to independently predict the extent of the spatial (or spatio-temporal) window in which the binding of self-related and external information occurs, and compare it to experimental measures. A match would constitute a hallmark that a process very similar to statistical inference truly takes place in the brain, or better said, a process that would be extremely hard to describe and motivate without using the tools of statistical inference. Current results are instead explainable just by combining the forced fusion model (which is Bayesian) and a “smooth” version of the spatial rule of multisensory integration, in which the amount of integration gradually decreases with disparity (which is not necessarily Bayesian). In order for the causal inference part of the model to have an objective added value with respect to a generic “smooth spatial rule”, the extent of the binding window needs to be convincingly predictable by measuring the precision of unisensory estimates. Unfortunately, this has not been tested in previous studies.

Overcoming these limitations is the deep motivation underlying Study 2. Again, our aim here is not to challenge Bayesian models of brain function in general. Instead, we want to test whether current Bayesian CI models can produce meaningful (and testable) predictions when applied to body ownership, and by extension to BSC. As of now, the pre-registered study is under revision and data collection can only start when such protocol is approved. Therefore, we cannot yet discuss the results of the study, but rather the implications of potential outcomes. Essentially, the main hypothesis that we will discuss here is the central requirement for optimality: that the unisensory precisions

measured in unisensory tasks match the ones predicted by fitting the Bayesian CI model on data from the multisensory reaching task. As a first implication, the fact that the properties of an implicit measure of body ownership are truly constrained by optimality requirements would show that Bayesian CI models can be used to produce informative predictions on key components of BSC. Furthermore, since the model only accounts for a part of the mechanisms known to play a role in body ownership, a success in modelling those factors would constitute a starting point and a strong motivation to push the investigation further. For example, aspects such as the temporal history of stimuli could be readily included in a Bayesian "evidence accumulator", to explain why prolonged stimulation is required to elicit a vivid RHI. Other subtler, more "cognitive" factors, such as the visual aspect of the hand (or "hipnotisability", a proposed account of overall sensibility to illusory ownership, see Lush et al., 2020), could be more challenging to model mathematically, but are in principle possible to describe within a Bayesian framework. In this sense, approximate inference through machine learning techniques may help to formulate predictions within the Bayesian brain framework even in the case where an explicit formulation of the underlying mathematical equations is not possible.

On the other hand, a failure to confirm our hypothesis would not necessarily imply that Bayesian models cannot be employed for BSC, as several different reasons may lead to a negative result. First, it is possible that the unisensory tasks do not capture exactly the same unisensory components that intervene in the multisensory task, despite our efforts in this direction. Indeed, measuring the unisensory spatial precision in the visual modality, or even defining its exact nature, is particularly challenging. Moreover, the model we used may be too simplified to accurately describe the data. As we mentioned, Bayesian approaches to brain function are a general and flexible framework which can be implemented in a wide variety of manners, and a negative result should not hinder their use, considering their unprecedented value as a unifying description of brain function. On the contrary, we hope that it will raise awareness about the technical and epistemological challenges of using Bayesian models of brain function not as mere conceptual tools, but as means to formulate new predictions and hypotheses about complex phenomena such as BSC.

3.5. Limitations and further perspectives

Despite the passion for epistemological questions has been one of the main driving forces in this research, it may be useful to conclude this dissertation on a more practical note, and briefly discuss the challenges we encountered, the limitations of this work, and our future plans to overcome them. We already discussed the limitations due to the simplification in the neural network used in Study 1. However, the main limitation of this study is arguably its lack of a rigorous empirical validation. In our work, we proved that visual inputs about external stimuli, coupled with proprioceptive inputs about hand position, could affect tactile processing on the hand, compatibly with the predictions of the

neural-network model. Still, the main purpose (and result) of the behavioural experiment was to justify the architecture of the network, and the choice of the sensory modalities to be modelled. By showing that three sensory modalities are not only conceptually, but also practically sufficient to elicit visuotactile interactions, we were able to support our choice to keep the network as simple as possible. Nevertheless, a rigorous validation of model predictions would require fitting behaviour at a much more granular level, likely needing an amount of behavioural testing which is hardly feasible at the practical level. Furthermore, even in case extensive enough behavioural testing was possible, there is the risk that network properties could be always tuned to reproduce behaviour, due to the large number of free parameters in a neural network model. Possibly, only intracranial recordings, allowing to map the receptive fields and tuning curves of single neurons, can provide data that is high-dimensional and granular enough to provide solid supporting evidence for a neural-network model. An interesting example of a successful attempt to bridge behaviour, neural recordings and machine-learning inspired neural-network models of brain function was proposed by Yamins and colleagues (Yamins et al., 2014). In macaques, they recorded neurons from various levels of the visual hierarchy while the monkeys performed an object recognition task. Then, in a high-throughput simulation, they trained thousands of neural network architectures to recognize the same family of objects used in the behavioural tasks. Simulated and recorded neural activities at various levels of the hierarchy were compared by regressing simulated and true activity on each other, allowing to compare model predictions and neural responses. Furthermore, this allowed identifying the architectures that best described neural activity, and draw general conclusions about the organization of brain networks. In our case, the need to model subjective components of BSC would restrict the investigations to human subjects, making it harder to perform extensive intracranial recordings. However, the growing availability of both chronic and in-surgery neurophysiological recordings makes a similar approach applicable to human populations. Equally importantly, in order for such an approach to be meaningful, the computational modelling side would need to be scaled up from the level of conceptual demonstration to high-throughput simulations, exploring different architectures and sets of training inputs.

Arguably, the research axis offering the most tangible possibilities of further development is the agency axis of studies 3 and 4. Here, the main limitations are the already mentioned facts that our data originates from a single participant, and a single recording site. In Study 4, we found that the pre-movement phase of 8 Hz oscillations predicted agency judgements, and speculated that this may be through an effect of these oscillations on subsequent connectivity between M1 and other brain areas. In order to test such hypothesis, it would be necessary to implement a similar paradigm while acquiring data from multiple recording sites, and possibly multiple participants. To this aim, we are in the process of developing an EEG-BMI paradigm to study oscillatory contributions to the sense of agency in healthy individuals. EEG combines the temporal resolution needed to uncover the role of neural oscillations, and the capacity to record whole brain signals, with a spatial resolution that

should be sufficient to pinpoint the key regions involved in the process, at least at the level of functional areas. Furthermore, EEG based brain-machine interfaces based on kinaesthetic motor imagery are a mature and easy to implement technique (Wolpaw et al., 2002). This would allow overcoming both main limitations of studies 3 and 4, by providing a setup that allows brain-wide recordings and can be generalized to any population non-invasively. Through an EEG-BMI setup, it is possible to translate the motor imagery of a participant to external feedback. In our case, a virtual hand closing will provide the feedback, in order to emulate the embodied setup of our previous studies. We will then tune other parameters (e.g., temporal delays) in order to obtain an appreciable variability in agency ratings, to be correlated with the phase of pre-movement sensorimotor oscillations. Within a similar setup, it will be possible to apply source reconstruction and functional connectivity analyses to test our hypothesis about the relation between pre-movement oscillations, neural connectivity and sense of agency.

4. References

Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

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. https://doi.org/10.1016/j.neubiorev.2013.01.029

Ari, C., & D'Agostino, D. P. (2016). Contingency checking and self-directed behaviors in giant manta rays: Do elasmobranchs have self-awareness? *Journal of Ethology*, *34*(2), 167–174. https://doi.org/10.1007/s10164-016-0462-z

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*(1523), 1499–1506. https://doi.org/10.1098/rspb.2003.2364

Bahrick, L. E. (1995). Intermodal Origins of Self-Perception. In *Advances in Psychology* (Vol. 112, Issue C, pp. 349–373). https://doi.org/10.1016/S0166-4115(05)80019-6

Bahrick, L. E., & Watson, J. S. (1985). Detection of intermodal proprioceptive-visual contingency as a potential basis of self-perception in infancy. *Developmental Psychology*, *21*(6), 963–973. https://doi.org/10.1037/0012-1649.21.6.963

Barlow, H. B. (1961). Possible Principles Underlying the Transformations of Sensory Messages. In *Sensory Communication* (pp. 216–234). The MIT Press. https://doi.org/10.7551/mitpress/9780262518420.003.0013

Bays, P. M., Wolpert, D. M., & Flanagan, J. R. (2005). Perception of the consequences of self-action is temporally tuned and event driven. *Current Biology : CB*, *15*(12), 1125–1128. https://doi.org/10.1016/j.cub.2005.05.023

Bender, M. B. (1952). Disorders in perception; with particular reference to the phenomena of extinction and displacement. In *Disorders in perception; with particular reference to the phenomena of extinction and displacement.* Charles C. Thomas.

Bermúdez, J. L. (1998). *The Paradox of Self-Consciousness* (Vol. 4, Issue 1). The MIT Press. https://doi.org/10.7551/mitpress/5227.001.0001

Berti, A., Bottini, G., Gandola, M., Pia, L., Smania, N., Stracciari, A., Castiglioni, I., Vallar, G., & Paulesu, E. (2005). Neuroscience: Shared cortical anatomy for motor

awareness and motor control. *Science*, *309*(5733), 488–491. https://doi.org/10.1126/science.1110625

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-Temporal Prediction Modulates the Perception of Self-Produced Stimuli. *Journal of Cognitive Neuroscience*, *11*(5), 551–559. https://doi.org/10.1162/089892999563607

Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport*, *11*(11), R11-6. https://doi.org/10.1097/00001756-200008030-00002

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, *1*(7), 635–640. https://doi.org/10.1038/2870

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience*, *13*(8), 556–571. https://doi.org/10.1038/nrn3292

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, *13*(1), 7–13. https://doi.org/10.1016/j.tics.2008.10.003

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness. *Neuron*, *88*(1), 145–166. https://doi.org/10.1016/j.neuron.2015.09.029

Born, J., Galeazzi, J. M., & Stringer, S. M. (2017). Hebbian learning of hand-centred representations in a hierarchical neural network model of the primate visual system. *PLOS ONE*, *12*(5), e0178304. https://doi.org/10.1371/journal.pone.0178304

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756–756. https://doi.org/10.1038/35784

Brain, W. R. (1941). Visual disorientation with special reference to lesions of the right cerebral hemisphere. *Brain*. https://doi.org/10.1093/brain/64.4.244

Buchholz, V. N., David, N., Sengelmann, M., & Engel, A. K. (2019). Belief of agency changes dynamics in sensorimotor networks. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-018-37912-w

Bufacchi, R. J., & Iannetti, G. D. (2018). An Action Field Theory of Peripersonal Space. *Trends in Cognitive Sciences*, *22*(12), 1076–1090. https://doi.org/10.1016/j.tics.2018.09.004

Butler, J. S., Smith, S. T., Campos, J. L., & Bulthoff, H. H. (2010). Bayesian integration of

visual and vestibular signals for heading. *Journal of Vision*, *10*(11), 23–23. https://doi.org/10.1167/10.11.23

Canzoneri, E., Magosso, E., & Serino, A. (2012). Dynamic sounds capture the boundaries of peripersonal space representation in humans. *PloS One*, *7*(9), e44306. https://www.ncbi.nlm.nih.gov/pubmed/23028516

Canzoneri, E., Marzolla, M., Amoresano, A., Verni, G., & Serino, A. (2013). Amputation and prosthesis implantation shape body and peripersonal space representations. *Scientific Reports*, *3*(1), 2844. https://doi.org/10.1038/srep02844

Cavazzana, A., Penolazzi, B., Begliomini, C., & Bisiacchi, P. S. (2015). Neural underpinnings of the “agent brain”: new evidence from transcranial direct current stimulation. *The European Journal of Neuroscience*, *42*(3), 1889–1894. https://doi.org/10.1111/ejn.12937

Chalmers, D. J. (2018). Facing Up to the Problem of Consciousness. In *Consciousness and Emotion in Cognitive Science*. https://doi.org/10.4324/9780203826430-11

Chambon, V., Moore, J. W., & Haggard, P. (2015). TMS stimulation over the inferior parietal cortex disrupts prospective sense of agency. *Brain Structure and Function*, *220*(6), 3627–3639. https://doi.org/10.1007/s00429-014-0878-6

Chang, L., Zhang, S., Poo, M. M., & Gong, N. (2017). Spontaneous expression of mirror self-recognition in monkeys after learning precise visual-proprioceptive association for mirror images. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(12), 3258–3263. https://doi.org/10.1073/pnas.1620764114

Colombo, M., & Wright, C. (2018). First principles in the life sciences: the free-energy principle, organicism, and mechanism. *Synthese*. https://doi.org/10.1007/s11229-018-01932-w

Cooke, D. F., Taylor, C. S. R., Moore, T., & Graziano, M. S. A. (2003). Complex movements evoked by microstimulation of the ventral intraparietal area. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(10), 6163–6168. https://doi.org/10.1073/pnas.1031751100

Cowie, D., McKenna, A., Bremner, A. J., & Aspell, J. E. (2018). The development of bodily self-consciousness: changing responses to the Full Body Illusion in childhood. *Developmental Science*, *21*(3). https://doi.org/10.1111/desc.12557

Crea, S., D’Alonzo, M., Vitiello, N., & Cipriani, C. (2015). The rubber foot illusion. *Journal*

of NeuroEngineering and Rehabilitation, *12*(1). https://doi.org/10.1186/s12984-015-0069-6

Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., & Jeannerod, M. (1997). Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, *65*(1), 71–86. https://doi.org/10.1016/S0010-0277(97)00039-5

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, *7*(5), 889–904. https://doi.org/10.1162/neco.1995.7.5.889

De Vignemont, F. (2011). Embodiment, ownership and disownership. *Consciousness and Cognition*, *20*(1), 82–93. https://doi.org/10.1016/j.concog.2010.09.004

Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science (New York, N.Y.)*, *324*(5928), 811–813. https://doi.org/10.1126/science.1169896

Desmurget, M., & Sirigu, A. (2012). Conscious motor intention emerges in the inferior parietal lobule. *Current Opinion in Neurobiology*, *22*(6), 1004–1011. https://doi.org/10.1016/j.conb.2012.06.006

Dewey, J. A., & Knoblich, G. (2014). Do implicit and explicit measures of the sense of agency measure the same thing? *PLoS ONE*, *9*(10). https://doi.org/10.1371/journal.pone.0110118

di Pellegrino, G., Làdavas, E., & Farné, A. (1997). Seeing where your hands are. *Nature*, *388*(6644), 730–730. https://doi.org/10.1038/41921

Driver, J., & Spence, C. (1998). Cross-modal links in spatial attention. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1373), 1319–1331. https://doi.org/10.1098/rstb.1998.0286

Driver, J., Vuilleumier, P., & Husain, M. (2004). Spatial Neglect and Extinction. In *The cognitive neurosciences, 3rd ed.* (pp. 589–606). Boston Review.

Dummer, T., Picot-Annand, A., Neal, T., & Moore, C. (2009). Movement and the rubber hand illusion. *Perception*, *38*(2), 271–280. https://doi.org/10.1068/p5921

Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. In *Science* (Vol. 317, Issue 5841). https://doi.org/10.1126/science.1142175

Eimer, M., & Schlaghecken, F. (2003). Response facilitation and inhibition in subliminal

priming. *Biological Psychology*, *64*(1–2), 7–26. https://doi.org/10.1016/s0301-0511(03)00100-5

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

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*, *116*(40), 20151–20157. https://doi.org/10.1073/pnas.1902334116

Farrer, C., Bouchereau, M., Jeannerod, M., & Franck, N. (2008). Effect of distorted visual feedback on the sense of agency. *Behavioural Neurology*, *19*(1–2), 53–57. https://doi.org/10.1155/2008/425267

Flament, D., & Hore, J. (1988). Relations of motor cortex neural discharge to kinematics of passive and active elbow movements in the monkey. *Journal of Neurophysiology*, *60*(4). https://doi.org/10.1152/jn.1988.60.4.1268

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48–58. https://doi.org/10.1038/nrn2536

Fornia, L., Puglisi, G., Leonetti, A., Bello, L., Berti, A., Cerri, G., & Garbarini, F. (2020). Direct electrical stimulation of the premotor cortex shuts down awareness of voluntary actions. *Nature Communications*, *11*(1), 1–11. https://doi.org/10.1038/s41467-020-14517-4

Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, *36*(11), 1133–1140. https://doi.org/10.1016/s0028-3932(98)00006-2

Fried, I., Mukamel, R., & Kreiman, G. (2011). Internally Generated Preactivation of Single Neurons in Human Medial Frontal Cortex Predicts Volition. *Neuron*, *69*(3), 548–562. https://doi.org/10.1016/j.neuron.2010.11.045

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*(10), 474–480. https://doi.org/10.1016/j.tics.2005.08.011

Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, *88*(1), 220–235. https://doi.org/10.1016/j.neuron.2015.09.034

Friston, K. (2011). Embodied inference: or “I think therefore I am, if I am what I think .” *The Implications of Embodiment (Cognition and Communication)*, 89–125.

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *355*(1404), 1771–1788. https://doi.org/10.1098/rstb.2000.0734

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Gallup, G. G. (1970). Chimpanzees: Self-recognition. *Science*, *167*(3914). https://doi.org/10.1126/science.167.3914.86

Garrido-Vásquez, P., & Rock, T. (2020). Sense of Agency in Multi-Step Actions. *Advances in Cognitive Psychology*, *16*(2), 85–91. https://doi.org/10.5709/acp-0287-2

Gold, K., & Scassellati, B. (2009). Using probabilistic reasoning over time to self-recognize. *Robotics and Autonomous Systems*, *57*(4), 384–392. https://doi.org/10.1016/j.robot.2008.07.006

Graziano, M. S. A. (1999). Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position. *Proceedings of the National Academy of Sciences*, *96*(August), 10418–10421.

Graziano, M. S. A. (2000). Coding the Location of the Arm by Sight. *Science*, *290*(5497), 1782–1786. https://doi.org/10.1126/science.290.5497.1782

Graziano, M. S. A., Hu, X. T., & Gross, C. G. (1997). Visuospatial Properties of Ventral Premotor Cortex. *Journal of Neurophysiology*, *77*(5), 2268–2292. https://doi.org/10.1152/jn.1997.77.5.2268

Graziano, M. S. A., Reiss, L. A. J., & Gross, C. G. (1999). A neuronal representation of the location of nearby sounds. *Nature*, *397*(6718), 428–430. https://doi.org/10.1038/17115

Grivaz, P., Blanke, O., & Serino, A. (2017). Common and distinct brain regions processing multisensory bodily signals for peripersonal space and body ownership. *NeuroImage*, *147*(December 2016), 602–618. https://doi.org/10.1016/j.neuroimage.2016.12.052

Grünbaum, T., & Christensen, M. S. (2020). Measures of agency. *Neuroscience of Consciousness*, *2020*(1), 1–13. https://doi.org/10.1093/nc/niaa019

Guterstam, A., Gentile, G., & Ehrsson, H. H. (2013). The Invisible Hand Illusion:

Multisensory Integration Leads to the Embodiment of a Discrete Volume of Empty Space. *Journal of Cognitive Neuroscience*, *25*(7), 1078–1099. https://doi.org/10.1162/jocn_a_00393

Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews. Neuroscience*, *18*(4), 196–207. https://doi.org/10.1038/nrn.2017.14

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382–385. https://doi.org/10.1038/nn827

Hanslmayr, S., Volberg, G., Wimber, M., Dalal, S. S., & Greenlee, M. W. (2013). Prestimulus oscillatory phase at 7 Hz gates cortical information flow and visual perception. *Current Biology*, *23*(22), 2273–2278. https://doi.org/10.1016/j.cub.2013.09.020

Hatsopoulos, N. G., & Suminski, A. J. (2011). Sensing with the Motor Cortex. *Neuron*, *72*(3), 477–487. https://doi.org/10.1016/j.neuron.2011.10.020

Held, R., & Freedman, S. J. (1963). Plasticity in human sensorimotor control. *Science Science*, *142*(3591). https://doi.org/10.1126/science.142.3591.455

Herter, T. M., Korbel, T., & Scott, S. H. (2009). Comparison of neural responses in primary motor cortex to transient and continuous loads during posture. *Journal of Neurophysiology*, *101*(1). https://doi.org/10.1152/jn.90230.2008

Hinton, G. (2014). Where do features come from? *Cognitive Science*, *38*(6), 1078–1101. https://doi.org/10.1111/cogs.12049

Holmes, N. P., Martin, D., Mitchell, W., Noorani, Z., & Thorne, A. (2020). Do sounds near the hand facilitate tactile reaction times? Four experiments and a meta-analysis provide mixed support and suggest a small effect size. *Experimental Brain Research*, *238*(4), 995–1009. https://doi.org/10.1007/s00221-020-05771-5

Hur, J.-W., Kwon, J. S., Lee, T. Y., & Park, S. (2014). The crisis of minimal self-awareness in schizophrenia: A meta-analytic review. *Schizophrenia Research*, *152*(1), 58–64. https://doi.org/10.1016/j.schres.2013.08.042

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport*, *7*(14), 2325–2330. http://www.ncbi.nlm.nih.gov/pubmed/8951846

Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behavioural Brain*

Research, *142*(1–2), 1–15. https://doi.org/10.1016/S0166-4328(02)00384-4

Juett, J., & Kuipers, B. (2019). Learning and acting in peripersonal space: Moving, reaching, and grasping. *Frontiers in Neurorobotics*, *13*(February), 1–20. https://doi.org/10.3389/fnbot.2019.00004

Kalckert, A., & Ehrsson, H. H. (2012). Moving a Rubber Hand that Feels Like Your Own: A Dissociation of Ownership and Agency. *Frontiers in Human Neuroscience*, *6*(MARCH 2012), 1–14. https://doi.org/10.3389/fnhum.2012.00040

Kalckert, A., & Ehrsson, H. H. (2014). The moving rubber hand illusion revisited: Comparing movements and visuotactile stimulation to induce illusory ownership. *Consciousness and Cognition*, *26*(1), 117–132. https://doi.org/10.1016/j.concog.2014.02.003

Kang, S. Y., Im, C. H., Shim, M., Nahab, F. B., Park, J., Kim, D. W., Kakareka, J., Miletta, N., & Hallett, M. (2015). Brain networks responsible for sense of agency: An EEG study. *PLoS ONE*, *10*(8), 1–16. https://doi.org/10.1371/journal.pone.0135261

Kannape, O. A., & Blanke, O. (2012). Agency, gait and self-consciousness. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *83*(2), 191–199. https://doi.org/10.1016/j.ijpsycho.2011.12.006

Khalighinejad, N., Di Costa, S., & Haggard, P. (2016). Endogenous Action Selection Processes in Dorsolateral Prefrontal Cortex Contribute to Sense of Agency: A Meta-Analysis of tDCS Studies of “Intentional Binding”. *Brain Stimulation*, *9*(3), 372–379. https://doi.org/10.1016/j.brs.2016.01.005

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE*, *2*(9), e943. https://doi.org/10.1371/journal.pone.0000943

Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialanderungen bei Willkurbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflugers Archiv Fur Die Gesamte Physiologie Des Menschen Und Der Tiere*, *284*(1), 1–17. https://doi.org/10.1007/BF00412364

Kornhuber, H. H., & Deecke, L. (2016). Brain potential changes in voluntary and passive movements in humans: readiness potential and reafferent potentials. *Pflugers Archiv : European Journal of Physiology*, *468*(7), 1115–1124. https://doi.org/10.1007/s00424-016-1852-3

Legaspi, R., & Toyoizumi, T. (2019). A Bayesian psychophysics model of sense of agency. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-12170-0

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: Manipulating bodily self-consciousness. *Science*, *317*(5841), 1096–1099. https://doi.org/10.1126/science.1143439

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain*. https://doi.org/10.1093/brain/106.3.623

Limanowski, J., & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience*, *7*(September), 547. https://doi.org/10.3389/fnhum.2013.00547

Longo, M.R., & Serino, A. (2012). Tool use induces complex and flexible plasticity of human body representations. *The Behavioral and Brain Sciences*, *35*(4).

Longo, Matthew R., & Lourenco, S. F. (2006). On the nature of near space: Effects of tool use and the transition to far space. *Neuropsychologia*, *44*(6), 977–981. https://doi.org/10.1016/j.neuropsychologia.2005.09.003

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*(1), 4853. https://doi.org/10.1038/s41467-020-18591-6

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438. https://doi.org/10.1038/nn1790

Magosso, E., Zavaglia, M., Serino, A., di Pellegrino, G., & Ursino, M. (2010). Visuotactile representation of peripersonal space: a neural network study. *Neural Computation*, *22*(1), 190–243. https://doi.org/10.1162/neco.2009.01-08-694

Makin, J. G., Fellows, M. R., & Sabes, P. N. (2013). Learning Multisensory Integration and Coordinate Transformation via Density Estimation. *PLoS Computational Biology*, *9*(4). https://doi.org/10.1371/journal.pcbi.1003035

Makin, T. R., Scholz, J., Henderson Slater, D., Johansen-Berg, H., & Tracey, I. (2015). Reassessing cortical reorganization in the primary sensorimotor cortex following arm amputation. *Brain : A Journal of Neurology*, *138*(Pt 8), 2140–2146.

https://doi.org/10.1093/brain/awv161

Maravita, A., Husain, M., Clarke, K., & Driver, J. (2001). Reaching with a tool extends visual-tactile interactions into far space: evidence from cross-modal extinction. *Neuropsychologia*, *39*(6), 580–585. https://doi.org/10.1016/s0028-3932(00)00150-0

Maravita, A., Spence, C., & Driver, J. (2003). Multisensory integration and the body schema: close to hand and within reach. *Current Biology : CB*, *13*(13), R531-9. https://doi.org/10.1016/s0960-9822(03)00449-4

Martel, M., Cardinali, L., Roy, A. C., & Farnè, A. (2016). Tool-use: An open window into body representation and its plasticity. In *Cognitive Neuropsychology* (Vol. 33, Issues 1–2). https://doi.org/10.1080/02643294.2016.1167678

Mellor, C. S. (1970). First rank symptoms of schizophrenia. I. The frequnncy in schizophrenics on admission to hospital. II. Differences between individual first rank symptoms. *The British Journal of Psychiatry : The Journal of Mental Science*. https://doi.org/10.1192/s0007125000192116

Metzinger, T. (2003). Being no one: The self-model theory of subjectivity. In *Being no one: The self-model theory of subjectivity.* (pp. xii, 699–xii, 699). MIT Press.

Moore, J. W. (2016). What is the sense of agency and why does it matter? *Frontiers in Psychology*, *7*(AUG), 1–9. https://doi.org/10.3389/fpsyg.2016.01272

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: A review. *Consciousness and Cognition*, *21*(1), 546–561. https://doi.org/10.1016/j.concog.2011.12.002

Moore, J. W., Ruge, D., Wenke, D., Rothwell, J., & Haggard, P. (2010). Disrupting the experience of control in the human brain: Pre-supplementary motor area contributes to the sense of agency. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1693), 2503–2509. https://doi.org/10.1098/rspb.2010.0404

Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and Cognition*, *18*(4), 1056–1064. https://doi.org/10.1016/j.concog.2009.05.004

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*(1), 67–76. https://doi.org/10.1016/j.concog.2014.01.009

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review*, *83*(4). https://doi.org/10.2307/2183914

Nguyen, D. H. P., Hoffmann, M., Roncone, A., Pattacini, U., & Metta, G. (2018). Compact Real-time Avoidance on a Humanoid Robot for Human-robot Interaction. *ACM/IEEE International Conference on Human-Robot Interaction*, 416–424. https://doi.org/10.1145/3171221.3171245

Noel, J.-P., Pfeiffer, C., Blanke, O., & Serino, A. (2015). Peripersonal space as the space of the bodily self. *Cognition*, *144*, 49–57. https://www.ncbi.nlm.nih.gov/pubmed/26231086

Noel, J., Samad, M., Doxon, A., Clark, J., Keller, S., & Di Luca, M. (2018). Peri-personal space as a prior in coupling visual and proprioceptive signals. *Scientific Reports*, *8*(1), 15819. https://doi.org/10.1038/s41598-018-33961-3

Normand, J. M., Giannopoulos, E., Spanlang, B., & Slater, M. (2011). Multisensory stimulation can induce an illusion of larger belly size in immersive virtual reality. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0016128

Pagel, B., Heed, T., & Röder, B. (2009). Change of reference frame for tactile localization during child development. *Developmental Science*, *12*(6). https://doi.org/10.1111/j.1467-7687.2009.00845.x

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Psychological Bulletin Bayesian Approaches to Autism: Towards Volatility, Action, and Behavior Bayesian Approaches to Autism: Towards Volatility, Action, and Behavior. *Psychological Bulletin*, *143*(5), 521–542. https://doi.org/10.1037/bul0000097

Penny, W. (2012). Bayesian Models of Brain and Behaviour. *ISRN Biomathematics*, *2012*. https://doi.org/10.5402/2012/785791

Porter, R., & Lemon, R. (2012). Corticospinal Function and Voluntary Movement. In *Corticospinal Function and Voluntary Movement*. https://doi.org/10.1093/acprof:oso/9780198523758.001.0001

Pouget, A., Deneve, S., & Duhamel, J.-R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews. Neuroscience*, *3*(9), 741–747. https://doi.org/10.1038/nrn914

Pruszynski, J. A., Kurtzer, I., Nashed, J. Y., Omrani, M., Brouwer, B., & Scott, S. H. (2011). Primary motor cortex underlies multi-joint integration for fast feedback control. *Nature*,

478(7369). https://doi.org/10.1038/nature10436

Pugach, G., Pitti, A., Tolochko, O., & Gaussier, P. (2019). Brain-inspired coding of robot body schema through visuo-motor integration of touched events. *Frontiers in Neurorobotics*, *13*(March). https://doi.org/10.3389/fnbot.2019.00005

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Redding, G. M., Rossetti, Y., & Wallace, B. (2005). Applications of prism adaptation: A tutorial in theory and method. In *Neuroscience and Biobehavioral Reviews* (Vol. 29, Issue 3). https://doi.org/10.1016/j.neubiorev.2004.12.004

Rizzolatti, G., Matelli, M., & Pavesi, G. (1983). Deficits in attention and movement following the removal of postarcuate (area 6) and prearcuate (area 8) cortex in macaque monkeys. *Brain*. https://doi.org/10.1093/brain/106.3.655

Rizzolatti, Giacomo, Scandolara, C., Matelli, M., & Gentilucci, M. (1981). Afferent properties of periarcuate neurons in macaque monkeys. I. Somatosensory responses. *Behavioural Brain Research*, *2*(2), 125–146. https://doi.org/10.1016/0166-4328(81)90052-8

Rochat, P., & Hespos, S. J. (1997). Differential rooting response by neonates: Evidence for an early sense of self. *Infant and Child Development*, *6*(3–4), 105–112. https://doi.org/10.1002/(sici)1099-0917(199709/12)6:3/4<105::aid-edp150>3.0.co;2-u

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*(6), e21659. https://doi.org/10.1371/journal.pone.0021659

Rohe, T., Ehlis, A. C., & Noppeney, U. (2019). The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. *Nature Communications*, *10*(1), 1–17. https://doi.org/10.1038/s41467-019-09664-2

Roncone, A., Hoffmann, M., Pattacini, U., Fadiga, L., & Metta, G. (2016). Peripersonal space and margin of safety around the body: Learning visuo-tactile associations in a humanoid robot with artificial skin. *PLoS ONE*, *11*(10), 1–32. https://doi.org/10.1371/journal.pone.0163713

Samad, M., Chung, A. J., & Shams, L. (2015). Perception of body ownership is driven by Bayesian sensory inference. *PLoS ONE*, *10*(2), 1–23.

https://doi.org/10.1371/journal.pone.0117178

Sambo, C. F., Liang, M., Cruccu, G., & Iannetti, G. D. (2012). Defensive peripersonal space: the blink reflex evoked by hand stimulation is increased when the hand is near the face. *Journal of Neurophysiology*, *107*(3), 880–889. https://doi.org/10.1152/jn.00731.2011

Serino, A., Bassolino, M., Farnè, A., & Làdavas, E. (2007). Extended Multisensory Space in Blind Cane Users. *Psychological Science*, *18*(7), 642–648. https://doi.org/10.1111/j.1467-9280.2007.01952.x

Serino, A., Canzoneri, E., Marzolla, M., di Pellegrino, G., & Magosso, E. (2015). Extending peripersonal space representation without tool-use: evidence from a combined behavioral-computational approach. *Frontiers in Behavioral Neuroscience*, *9*, 4. https://www.ncbi.nlm.nih.gov/pubmed/25698947

Serino, A., Sforza, A. L., Kanayama, N., van Elk, M., Kaliuzhna, M., Herbelin, B., & Blanke, O. (2015). Tuning of temporo-occipital activity by frontal oscillations during virtual mirror exposure causes erroneous self-recognition. *European Journal of Neuroscience*, *42*(8), 2515–2526. https://doi.org/10.1111/ejn.13029

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Seth, A. K., & Friston, K. J. (2016). Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1708), 20160007. https://doi.org/10.1098/rstb.2016.0007

Seth, A. K., & Tsakiris, M. (2018). Being a Beast Machine: The Somatic Basis of Selfhood. *Trends in Cognitive Sciences*, 1–13. https://doi.org/10.1016/j.tics.2018.08.008

Sforza, A., Bufalari, I., Haggard, P., & Aglioti, S. M. (2010). My face in yours: Visuo-tactile facial stimulation influences sense of identity. *Social Neuroscience*, *5*(2). https://doi.org/10.1080/17470910903205503

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, *408*(6814). https://doi.org/10.1038/35048669

Slaughter, V., & Brownell, C. A. (2011). Early development of body representations. In *Early Development of Body Representations*. https://doi.org/10.1017/CBO9781139019484

Spaccasassi, C., & Maravita, A. (2020). Peripersonal space is diversely sensitive to a temporary vs permanent state of anxiety. *Cognition*, *195*(June 2019), 104133. https://doi.org/10.1016/j.cognition.2019.104133

Sperduti, M., Delaveau, P., Fossati, P., & Nadel, J. (2011). Different brain structures related to self- and external-agency attribution: A brief review and meta-analysis. *Brain Structure and Function*, *216*(2), 151–157. https://doi.org/10.1007/s00429-010-0298-1

Straka, Z., & Hoffmann, M. (2017). Learning a Peripersonal Space Representation as a Visuo-Tactile Prediction Task. In *Lecture Notes in Computer Science: Vol. 10613 LNCS* (Issue i, pp. 101–109). https://doi.org/10.1007/978-3-319-68600-4_13

Suminski, A. J., Tkach, D. C., Fagg, A. H., & Hatsopoulos, N. G. (2010). Incorporating Feedback from Multiple Sensory Modalities Enhances Brain-Machine Interface Control. *Journal of Neuroscience*, *30*(50), 16777–16787. https://doi.org/10.1523/JNEUROSCI.3967-10.2010

Suminski, Aaron J., Tkach, D. C., & Hatsopoulos, N. G. (2009). Exploiting multiple sensory modalities in brain-machine interfaces. *Neural Networks*, *22*(9). https://doi.org/10.1016/j.neunet.2009.05.006

Synofzik, M., Vosgerau, G., & Voss, M. (2013). The experience of agency: An interplay between prediction and postdiction. *Frontiers in Psychology*, *4*(MAR), 1–8. https://doi.org/10.3389/fpsyg.2013.00127

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 FEB), 1–13. https://doi.org/10.3389/fnhum.2014.00091

Tsakiris, M. (2010). My body in the brain: A neurocognitive model of body-ownership. *Neuropsychologia*, *48*(3), 703–712. https://doi.org/10.1016/j.neuropsychologia.2009.09.034

Tsakiris, M., & Haggard, P. (2005). The rubber hand illusion revisited: Visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.31.1.80

van Es, T., & Hipolito, I. (2020). *Free-Energy Principle, Computationalism and Realism: a Tragedy*. *1124818*, 1–28. http://philsci-archive.pitt.edu/18497/

Walsh, L. D., Moseley, G. L., Taylor, J. L., & Gandevia, S. C. (2011). Proprioceptive signals contribute to the sense of body ownership. *The Journal of Physiology*, *589*(12), 3009–

3021. https://doi.org/10.1113/jphysiol.2011.204941

Wegner, D. M. (2002). *The illusion of conscious will*. MIT Press.

Williams, D. (2020). Epistemic Irrationality in the Bayesian Brain. *The British Journal for the Philosophy of Science*. https://doi.org/10.1093/bjps/axz044

Wittgenstein, L. (1958). *The Blue and Brown Books* (Vol. 34, Issue 131). Harper & Row.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. In *Clinical Neurophysiology* (Vol. 113, Issue 6). https://doi.org/10.1016/S1388-2457(02)00057-3

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111

Yoshie, M., & Haggard, P. (2013). Negative emotional outcomes attenuate sense of agency over voluntary actions. *Current Biology*, *23*(20), 2028–2032. https://doi.org/10.1016/j.cub.2013.08.034

Zaadnoordijk, L., Hunnius, S., Meyer, M., Kwisthout, J., & van Rooij, I. (2015). The developing sense of agency: Implications from cognitive phenomenology. *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 114–115. https://doi.org/10.1109/DEVLRN.2015.7346126

Ziemann, U., Hallett, M., & Cohen, L. G. (1998). Mechanisms of deafferentation-induced plasticity in human motor cortex. *Journal of Neuroscience*, *18*(17). https://doi.org/10.1523/jneurosci.18-17-07000.1998

Zito, G. A., Wiest, R., & Aybek, S. (2020). Neural correlates of sense of agency in motor control: A neuroimaging meta-analysis. *PLOS ONE*, *15*(6), e0234321. https://doi.org/10.1371/journal.pone.0234321

5. Articles

Received: 20 January 2020 | Revised: 3 September 2020 | Accepted: 4 September 2020

DOI: 10.1111/ejn.14981

RESEARCH REPORT

EJN European Journal of Neuroscience FENS WILEY

From statistical regularities in multisensory inputs to peripersonal space representation and body ownership: Insights from a neural network model

Tommaso Bertoni[1] | **Elisa Magosso**[2] | **Andrea Serino**[1]

[1]MySpace Lab, Department of Clinical Neuroscience, Lausanne University Hospital (CHUV), University of Lausanne, Lausanne, Switzerland

[2]Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi", University of Bologna, Cesena, Italy

Correspondence
Andrea Serino, MySpace Lab, Department of Clinical Neuroscience, Lausanne University Hospital (CHUV), Avenue Pierre Decker 5, CH-1011, Lausanne, Switzerland.
Email: andrea.serino@unil.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 163951

Abstract

Peripersonal space (PPS), the interface between the self and the environment, is represented by a network of multisensory neurons with visual (or auditory) receptive fields anchored to specific body parts, and tactile receptive fields covering the same body parts. Neurophysiological and behavioural features of hand PPS representation have been previously modelled through a neural network constituted by one multisensory population integrating tactile inputs with visual/auditory external stimuli. Reference frame transformations were not explicitly modelled, as stimuli were encoded in pre-computed hand-centred coordinates. Here we present a novel model, aiming to overcome this limitation by including a proprioceptive population encoding hand position. We confirmed behaviourally the plausibility of the proposed architecture, showing that visuo-proprioceptive information is integrated to enhance tactile processing on the hand. Moreover, the network's connectivity was spontaneously tuned through a Hebbian-like mechanism, under two minimal assumptions. First, the plasticity rule was designed to learn the statistical regularities of visual, proprioceptive and tactile inputs. Second, such statistical regularities were simply those imposed by the body structure. The network learned to integrate proprioceptive and visual stimuli, and to compute their hand-centred coordinates to predict tactile stimulation. Through the same mechanism, the network reproduced behavioural correlates of manipulations implicated in subjective body ownership: the invisible and the rubber hand illusion. We thus propose that PPS representation and body ownership may emerge through a unified neurocomputational process; the integration of multisensory information consistently with a model of the body in the environment, learned from the natural statistics of sensory inputs.

KEYWORDS

bodily self-consciousness, body representation, Hebbian learning, reference frame transformations, statistical inference

Abbreviations : HMD, head-mounted display; IHI, Invisible Hand Illusion; PPS, peripersonal space; RBM, restricted Boltzmann machine; RF, receptive field; RHI, Rubber Hand Illusion; RT, reaction time; *SD*, standard deviation; VR, virtual reality.

Edited by: Dr. Edmund Lalor

1 | INTRODUCTION

1.1 | Peripersonal space

Peripersonal space (PPS) is typically defined as the region of space immediately surrounding the body, or the space where we can physically interact with external objects, either actively, by reaching to touch them, or passively, when we enter in contact with an incoming object (di Pellegrino & Làdavas, 2015; Serino, 2019). PPS was originally defined in terms of a physical space, with a specific neural representation, following long-known selective impairments of action and perception for stimuli in the near space induced by natural lesions in brain-damaged patients (Brain, 1941) and by experimental lesions in monkeys (Rizzolatti et al., 1983). This concept was then expanded by neurophysiological and behavioural studies focusing on multisensory processing of stimuli within a limited distance from the body. In particular, studies on non-human primates have described a population of multisensory neurons responding to visual and/or auditory stimuli, close to specific body parts, and to tactile stimulation of the same body parts (Duhamel et al., 1998; Graziano et al., 1994; Rizzolatti et al., 1981). That is, they present multisensory receptive fields which are selective for given body parts and anchored to them in space. Such evidence has been interpreted as the demonstration of the existence of a system representing the space around the different parts of the body in the primate brain, whose extent is defined by the extent of the multisensory receptive fields of those neurons. The term PPS then came to define not only a topographical region, but also its neural representation, leading to a variety of different descriptions whose common principles we try to resume here.

Evidence for the existence of an analogous system in humans comes from a body of neuropsychological (Farnè & Làdavas, 2002; di Pellegrino et al., 1997), behavioural (Spence et al., 2000; Zampini et al., 2007), and neuroimaging (Brozzoli et al., 2011; Grivaz et al., 2017; Makin et al., 2007) studies, coherently showing that interactions between tactile processing and visual and/or auditory cues is stronger when these stimuli are presented close to the body, as opposed to far.

Finally, several experimental results suggest to interpret PPS as a shell of interaction between the body and the environment, in which potential contacts between body parts and external objects are processed and predicted, with defensive (prepare reactions to potential threats) or appetitive (e.g., during reaching movements) purposes (Bufacchi & Iannetti, 2018; Cléry et al., 2015; Serino, 2019).

1.2 | Previous models and motivation

Magosso, Ursino, et al. (2010) have developed a neural network model aiming to reproduce the main features of PPS representation in a neurophysiologically plausible computational framework. The model consists of two unisensory neuronal populations (auditory/visual, tactile), connected to a multisensory population: the receptive fields of visual/auditory neurons cover an extended space around the hand (or another target body part), while those in the tactile population code for touch on the same body part (Magosso, Ursino, et al., 2010; Magosso, Zavaglia, et al., 2010). In order to reproduce the space-dependent responses of multisensory neurons in the PPS system, the connectivity of the network was tuned as follows: both tactile and visual/auditory neurons coding for stimuli that are close to the hand project strongly to the multisensory layer, whereas visual/auditory neurons coding for far stimuli project weakly to the multisensory layer. Thus, tactile stimuli on the body and visual/auditory stimuli close to the body induce stronger multisensory interaction than stimuli in the far space. This architecture reproduced neurophysiological (Bernasconi et al., 2018) and behavioural (Serino, Noel, et al., 2015) results of enhanced tactile processing in the presence of stimuli inside versus outside the PPS, and also of plastically induced changes in PPS representation (Magosso, Zavaglia, et al., 2010; Serino, Canzoneri, et al., 2015).

PPS representation is inherently body part centred. While tactile stimuli are directly processed in body-centred reference frames, external auditory and visual stimuli are initially processed in head-centred and eye-centred reference frames. Thus, PPS representation requires a complex set of reference frame transformations on the incoming stimuli in order to estimate their position relative to the different body parts. For the sake of simplicity, the neural network model proposed by Magosso and colleagues assumed static body parts, as if reference frame transformations had been already achieved by means of other mechanisms. Other computational models have proposed to account for reference frame transformations, for instance by Pouget et al. (2002), and Makin et al. (2013). Pouget and colleagues modelled reference frame transformations by simulating three interconnected populations: two of them encode the position of the same stimulus in different reference frames, and the third encodes the offset between the two reference frames. For instance, one population could code for the visual (retinotopic) position of a stimulus, the second population for the auditory (head-centred) position of the same stimulus, while the third could encode the shift between the two reference frames, represented by the gaze angle. By adjusting the weight of feedback and feedforward synapses, the model could either compute the position in a given reference frame based on the activity in the other two populations, or optimally integrate the three of them to increase the reliability of the information in each modality. However, it has not been investigated whether a similar model could also account for the emergence of body-part centred visuo-tactile interactions as the key property of PPS representation.

An additional limitation of the previous model is that the synaptic connections that underlie PPS representation in Magosso, Ursino, et al. (2010) work were hard-wired, and while a second model (Magosso, Zavaglia, et al., 2010) adds Hebbian plasticity, this was only done on top of a pre-defined synaptic connectivity. Therefore, existing models cannot explain how the spatial organization of the multisensory receptive fields underlying PPS representation emerges. Such neural representation has been shown to be highly plastic, e.g., it extends after using a tool to reach far portions of space (Canzoneri et al., 2013; Iriki et al., 1996; Maravita & Iriki, 2004). Interestingly, it was also shown behaviourally that PPS representation can be modified with simple audio-far/tactile-near stimulation, unrelated with tool use (Serino, Canzoneri, et al., 2015). It is therefore reasonable to suppose that that PPS representation might arise from networks of neurons whose large scale architecture, at the level of functional areas, is hard-wired genetically in the brain, but in which the fine structure is based on the spontaneous tuning of synaptic connectivity induced by multisensory inputs through Hebbian learning. Hence, a key question in the field is not only to render how multisensory integration within overlapping visual and tactile receptive fields occurs, but also how such overlap is formed and maintained throughout development and everyday life.

1.3 | Aim of the work

The aim of the present study is therefore to extend the previously established model of PPS representation (Magosso, Ursino, et al., 2010), in order to formalize a neurocomputational framework able to learn visuotactile associations from experience, and maintain them as body parts move in space. More specifically, with the model we aim to show:

1. How the synaptic connectivity that arises from natural stimulation in the environment can account for the emergence of overlapping visual/auditory and tactile receptive fields (RFs) subtending PPS representation.
2. That the same learned associations that build PPS representation implicitly perform reference frame transformations in body-part centred coordinates. Therefore, a key novel point of our study is to demonstrate that PPS representation and reference frame transformations can emerge spontaneously and simultaneously within a unified neurocomputational process, by learning the statistical associations in multisensory inputs that occur naturally when interacting through the body within the environment. As a key example of PPS representation, here we focused on visuotactile integration around the hand, in hand-centred reference frames.

To achieve our aims, we have adapted our previous model (Magosso, Ursino, et al., 2010) via two main modifications. First, proprioceptive inputs, previously neglected, were now taken into account by adding a population of proprioceptive neurons coding the location of the hand in space with respect to the trunk. Second, several psychophysical (Alais & Burr, 2004; Ernst & Banks, 2002), theoretical and computational works (Knill & Pouget, 2004; Ma et al., 2006; Makin et al., 2013) suggested to model multisensory integration in a probabilistic framework. This assumption guided us towards the choice of a plasticity rule designed to learn the statistical properties of visual, proprioceptive and tactile inputs. In the interest of approximating a key feature of biological neural networks that is key to our aims, we imposed the additional constraint that the learning rule should be Hebbian-like, that is, based only on local correlations between neural activities. The network, still keeping the fundamental architecture of unisensory populations reciprocally connected with a multisensory layer, was therefore formalized as a Restricted Boltzmann Machine (RBM), a type of artificial neural network designed to efficiently learn the unknown joint probability distribution of its set of inputs through a local learning rule (Hinton & Salakhutdinov, 2006; Makin et al., 2013). Thus, we did not simulate the response to multisensory (tactile and visual) stimuli close to the hand through pre-programmed synapses between the network's populations. Instead, we simulated a training where multisensory stimuli are randomly presented in space, with the only constraint, based on the physical properties of the body, that tactile inputs are simultaneously associated with visual inputs occurring on or near the hand, and never with far visual stimuli. We then let the model tune its synaptic connectivity to learn the statistical regularities in such a pattern of stimulation. This was compared with another "unconstrained" training model, where tactile and visual inputs were provided randomly and independently. We showed how, after the "body-constrained" training, the model produces multisensory responses to tactile stimuli on the hand and visual stimuli close to the hand, as a function of the position of the hand in space, suggesting the emergence of multisensory, hand-centred, receptive fields. Results from in silico computational simulations were then compared with results from in vivo psychophysical experiments to demonstrate the plausibility of the model. Finally, we also tested the model with analogue patterns of multisensory stimulation as those used to affect the sense of body ownership during the so-called invisible hand illusion (IHI; Guterstam et al., 2013) and rubber hand illusion (Botvinick & Cohen, 1998). By measuring the network's response from the proprioceptive population, we could reproduce a computational analogue of the so-called proprioceptive drift, i.e., a shift in the perceived location of one's own hand that is considered a behavioural proxy of changes in body ownership obtained via the illusions. Furthermore, we showed how the network's principles can be generalized to obtain similar results from more complex architectures. We included a visual

population coding for hand position, changed the encoding schema of proprioceptive inputs to joint angles, and added another reference frame transformation, by encoding visual inputs in eye-centred coordinates and adding a population coding for gaze angle.

2 | MATERIALS AND METHODS

2.1 | Qualitative network description

While built upon Magosso, Ursino, et al. (2010) model architecture, the model presented here substantially differs from the previous one. First, in order to account for the evidence showing that the response of PPS neurons is modulated by proprioceptive inputs, in the present study we included a proprioceptive neural population coding for hand position, in addition to the two unisensory tactile and visual (or auditory) neural populations. Second, in order to overcome the necessity of hard-wired synapses, and model the learning of reference frame transformations and PPS representation from synaptic tuning to external stimuli, we used a Restricted Boltzmann Machine (RBM) with two layers. RBMs are conceptually simple networks, widely used in unsupervised machine learning because of their efficiency in learning complex probability distributions. In their simplest form, they consist of two sets of units arranged in two layers, the lower layer and upper layer (usually called visible and hidden units respectively). A layer is defined as a pool of neurons that have no connections within the layer, but have bidirectional connections with neurons in the other layer. The units in the lower layer code for the components of an observation/event: in our RBM (Figure 1a), the lower layer is composed of the populations of unisensory neurons (proprioceptive, tactile, visual) that code the unisensory components of an event/stimulus. The units in the upper layer, (called multisensory layer in our RBM as it receives convergent inputs from multiple modalities) model the dependencies among these components. Even if there is no strict biological equivalent, the two layers can be seen as two levels in the processing of sensory information, where the lower layer receives the unisensory inputs, and the upper layer integrates them. We chose to restrict the network to two layers in our simpler model, for the sake of the interpretability of the results. Clearly, the sharp

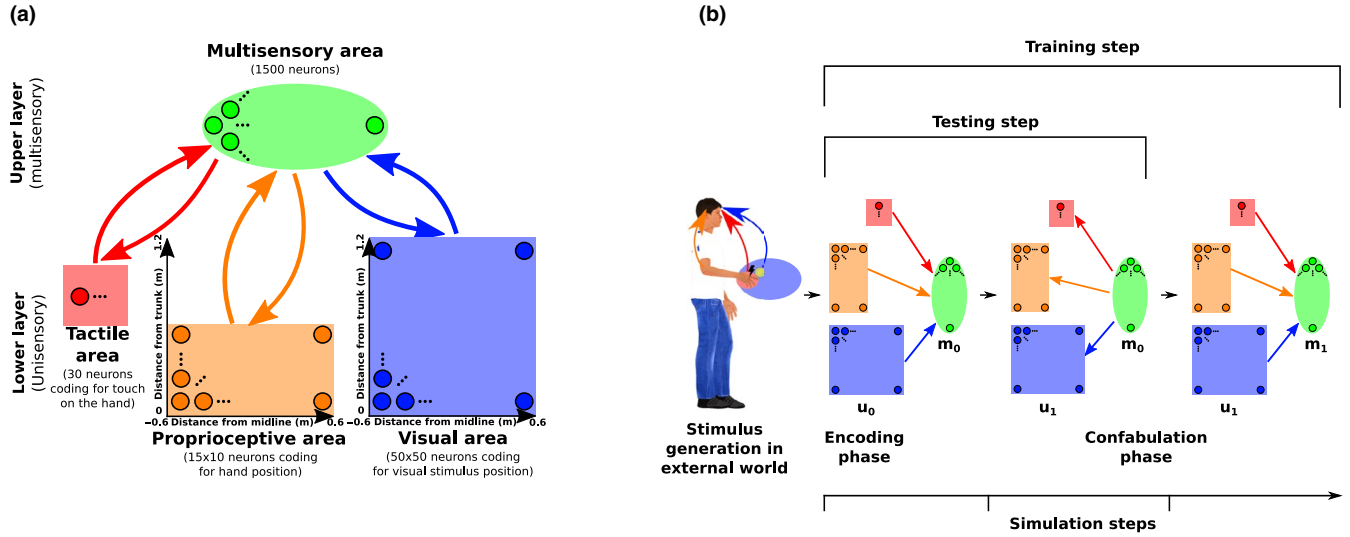


FIGURE 1 Network architecture, training and testing. (a) Architecture of the network. In the lower layer, three unisensory populations encode tactile stimulation on the hand, the proprioceptive position of the hand, the position of a visual stimulus. The upper layer is composed of multisensory neurons, in the sense that they receive inputs from each of the three unisensory populations. Each neuron in the proprioceptive and visual population has a preferred position distributed on a regular grid, with a Gaussian tuning curve of fixed width (~13 cm and ~11 cm respectively). For every stimulus, the number of spikes of neurons in the lower layer is drawn from a Poisson distribution, whose mean is determined by the tuning curve and a randomly selected gain in the range 4–10. The activity of neurons in the tactile population is set to 0 when the distance between the hand and the visual stimulus is greater than 15 cm. If the distance is smaller than 15 cm, the spike count for the tactile neurons is drawn from a Poisson distribution of mean 4–10, with this value randomly selected for each stimulus. Neurons in the lower layer are connected to neurons in the upper layer by bi-directional, symmetric synapses. (b) One training/testing step of the network. During testing, one stimulus is generated and encoded in the lower layer (u_0), and the activity of the upper layer (m_0) is computed based on the unisensory neurons activity. Then, the activity of the unisensory neurons is re-computed based on the multisensory neurons' activity to obtain the read-out of the integrated information encoded in the multisensory population (u_1). During training, an additional encoding step (confabulation phase) is added, where the activity of the multisensory neurons (m_1) is computed based on the reconstructed activity in the unisensory populations (u_1). Then, the synapses are updated with a weight change proportional to the difference in correlations between the lower and upper layer neurons in the two phases

distinction between layers is a purely conceptual construct, and biological multisensory processing takes place in a more complex fashion, involving possibly more "layers" of recurrent processing. Nevertheless, our choice goes in the direction of showing that a simple architecture is general enough to capture the key features of multisensory integration in PPS. In the lower layer, unisensory tactile neurons code for touch on the hand, proprioceptive neurons code for the position of the hand with respect to the trunk, and visual neurons code for the position of an external stimulus in trunk-centred coordinates (Figure 1a). Note that, for the visual population, this implies that inputs are represented as if the head and fixation were kept fixed, omitting for simplicity two additional components of the full reference frame transformation from retinotopic to head-centred to body-centred coordinates. Visual inputs are originally coded in eye-centred reference frames. Thus, to gather proper information about the position of the visual stimuli with respect to the hand's position, visual inputs need to be recoded in more global, trunk centred reference frame. These transformations can be added to our model by including populations coding for head and eye positions, and letting the network learn the joint distribution over all the neural populations. In the last paragraph of the results section, we demonstrate how the main results of this work can be recovered from a network including a fourth population coding for gaze angle, therefore implementing an additional reference frame transformation. The upper layer consists of multisensory neurons, i.e. neurons that receive inputs from the three unisensory populations. The visual and proprioceptive populations represent areas of 1.2×1.2 and 1.2×0.6 meters in front of the trunk, respectively, with the first dimension representing the medial-lateral axis and the second dimension the anterior-posterior axis. The specified sizes refer to the area where stimuli are actually delivered during training, while the area spanned by the neurons' preferred positions is slightly larger due to the margins to prevent edge effects. Similarly to what was done in previous models of multisensory integration, visual and proprioceptive populations use a population coding with Gaussian tuning curves to encode the positions of the visual stimulus and of the hand (Ma et al., 2006). The tactile area simply encodes the presence of tactile stimulation by activating all its neurons with a mean value proportional to the stimulation intensity.

In order to gauge the unisensory inputs' parameters, we required the maximal theoretical visual and proprioceptive precision under an optimal decoder to be consistent with behavioural human studies (Jones et al., 2010; Rincon-Gonzalez et al., 2011; Van Beers et al., 1998). Such value is defined as the standard deviation of the posterior probability of the stimulus location, given the activity of the unisensory population. It depends on the gain (i.e., the strength) of the stimuli and on the density of neurons per unit of space represented, and can be calculated with good approximation on the same bases as in previous works (Ma et al., 2006; Makin et al., 2013) (see Supporting Information for the detailed calculation). With the chosen parameters, the proprioceptive precision at maximal gain is 1.68 cm, and the visual accuracy is 0.45 cm, consistently with what has been reported in human behavioural studies (Jones et al., 2010; Rincon-Gonzalez et al., 2011; Van Beers et al., 1998). The number of multisensory neurons was determined empirically, looking for the optimal tradeoff between minimizing the number of units and maximizing network performance. Specifically, the number of hidden units was set so that a further increase in their number would not lead to significant improvement in the precision of positions encoded in the multisensory layer (see Supporting Information for details). While the receptive fields of the unisensory populations are defined a priori, the receptive fields of the multisensory neurons are learned during training. As widely done in RBMs, we used one-step contrastive divergence as learning algorithm. Contrastive divergence is based on local correlations between neuronal activity and does not require backpropagation, and can be therefore mapped to biologically realistic plasticity mechanisms, namely Hebbian learning. In machine-learning, RBMs are used to learn a generative model of the probability distribution of the inputs presented during the training. This means that, after a successful training, samples taken from the spontaneous activity of the network should come from the same probability distribution as the training examples. Through this mechanism, RBMs have been used to model multisensory integration and reference frame transformations (Makin et al., 2013). Here we test the hypothesis that, in a similar way, the emergence of PPS representation can be simply modelled by letting a neural network learn the regularities of its sensory inputs, represented by correlations across different sensory modalities.

2.2 | Mathematical network description

In a probabilistic population code, such as the one used for the generation of stimuli in our network, the activity of neurons in the unisensory populations can be seen as a probability distribution conditioned on the position of the stimuli in the physical world, from which spike counts are drawn for each population. Let $\mathbf{x_v}$ be the (2D) position of the visual stimulus, and $\mathbf{x_p}$ the position of the hand in the same 2D plane, then the activity of the i-th unisensory neuron u_i is defined by:

$$u_{v_i} = Pois\left(\lambda_{v_i}\right), \lambda_{v_i} = g_v e^{-\frac{\left\|x_v - \hat{x}_{v_i}\right\|^2}{2\sigma_v^2}} \quad (1)$$

$$u_{p_i} = Pois\left(\lambda_{p_i}\right), \lambda_{p_i} = g_p e^{-\frac{\left\|x_v - \hat{x}_{v_i}\right\|^2}{2\sigma_p^2}} \quad (2)$$

$$u_{t_i} = Pois\left(g_t\right) \text{ if} \boldsymbol{x}_v - \boldsymbol{x}_p < 0.15m, 0 \text{otherwise} \quad (3)$$

where u_v, u_p, u_t, respectively, denote neurons belonging to visual, proprioceptive and tactile populations, and $\hat{x}$ denotes the preferred position of a given neuron. The *SD* of the tuning curves (σ_v and σ_p) was set at three neurons for the visual population, and at one neuron for the proprioceptive population (i.e., around 1/15 of the whole population's range, which gives ~13 cm for proprioceptive and ~11 cm for visual neurons). 30 tactile units are used, and the preferred positions of the proprioceptive and visual neurons tile the space on a regular grid of 15 × 10 and 50 × 50 neurons, respectively. This includes the 1.2 × 0.6 and 1.2 × 1.2 meters of space represented by the neural populations, plus a safety margin (approximatively three times the *SD* of the tuning curve, or 30 cm on each side in physical units) to avoid boundary effects. The width of the tuning curve was mainly determined during preliminary testing, on the basis of a set of heuristic criteria. We noticed that in order to allow efficient learning, the average learning signal from units from different populations needs to be approximately the same, hence the width of the tuning curve needs to be a fixed fraction of the total population range. Since the average firing rate of tactile units is fixed by the proportion of training inputs where touch is provided to approximately 5%, the width of visual and proprioceptive tuning curves was chosen to approximately match this value, while not requiring excessively large safety margins. In any case, the network's main predictions were robust with respect to the choice of such parameter, as shown in the Supporting Information. The parameter g represents the stimulus strength (gain), and is varied during training independently for each unisensory population, by drawing a random, uniformly distributed number between 4 and 10 for each stimulus presentation. Note that, alternatively, tactile inputs could have been encoded similarly to visual and proprioceptive inputs, with a population representing the whole hand whose individual neurons respond preferentially to specific locations. In preliminary testing, the two encoding schemas yielded largely overlapping results. However, the current encoding schema was preferred as empirical evidence shows that tactile receptive fields of PPS neurons tend to be large, covering whole body parts, suggesting that their functional role is to roughly predict tactile interaction at the level of entire body parts, more than predicting the specific location of tactile stimulation. Also note that the tactile population needs not be an early tactile area as S1, but possibly a higher level somatosensory area, responding prevalently to tactile stimulation.

The activity of neurons is updated simultaneously in all neurons in a given layer, based on the activity of neurons in the other layer. In other words, the network has no temporal dynamics, and, differently from the previous model, there are no intra-layer connections, as the generation of spread-out population level activation is simulated by the size of unisensory receptive fields. For simplicity, we define as an "up" pass when the activity of the upper layer is computed given the activity of the lower layer, and a "down" pass when the activity of the lower layer is computed given the activity of the upper layer. The up and down passes are defined as follows:

$$Up: \boldsymbol{m} = Bern(\boldsymbol{\mu}),\ \mu = \sigma\left(\boldsymbol{W}\boldsymbol{u} + \boldsymbol{b}_m\right),\ \sigma(x) = 1/(1+e^{-x}) \quad (4)$$

$$Down: \boldsymbol{u} = Pois\ \lambda,\ \lambda = e^{\boldsymbol{W}^T\boldsymbol{m} + \boldsymbol{b}_v} \quad (5)$$

where **u** is the vector of activity of all neurons in the lower layer (unisensory), **m** is the vector of activity of all neurons in the upper layer (multisensory), W_{ij} is the synaptic weight connecting neuron m_i to neuron u_j, and $\mathbf{b_u}$ and $\mathbf{b_m}$ are biases for unisensory and multisensory neurons respectively. Note that the fact that the matrix used in the "down" pass is the transpose of the matrix used in the "up" pass implies that feedforward and feedback synapses are symmetric. In practice, the multisensory neurons' activity, given the unisensory neurons' activity, is a vector of samples of Bernoulli variables, whose mean is a sigmoidal function of the weight matrix acting on the unisensory neurons. Conversely, the unisensory neurons' activity, given the multisensory neurons' activity, is a vector of samples of Poisson variables, whose mean is the exponential function of the weight matrix acting on the multisensory neurons. In RBMs, the choice of sigmoidal and exponential "link" functions is the standard for Bernoulli and Poisson units, respectively (Welling et al., 2004).

2.3 | Training

The network was initialized with random connectivity, with each synaptic weight being drawn from a Gaussian with zero mean and 0.001 *SD*, and all biases were set to zero. Then, it was trained by presenting patterns of stimulations reproducing the natural associations between tactile, proprioceptive, and visual inputs. That is, for each training example, two independent, uniformly distributed positions were randomly generated for the hand and the visual stimulus, and encoded in the visual and proprioceptive populations, respectively. In the "body-constrained" training, tactile stimulation was provided when the distance between the stimulus position and the hand position was smaller than 15 cm, roughly the centre to centre distance at which hand-object tactile interactions are expected to take place. This resulted in tactile stimulation being provided in approximately 5% of the trials. In the control, unconstrained training, we randomly provided tactile stimulation in 5% of the trials, in order to remove the statistical regularity imposed by the body structure while keeping the amount of tactile stimulation constant. The input was encoded in the unisensory populations and then integrated in the upper layer through feedforward synapses, according to the rules defined in the previous paragraph. After this, a "confabulation" phase followed to complete the learning process for a given training example (Figure 1b). In the

confabulation phase, the integrated stimulus was projected back to the lower layer through feedback connections, and again to the upper layer (Hinton, 2000). After a batch of 100 encoding-confabulation sequences, the synaptic weight changes proportionally to the difference in the two phases in correlations between the upper and lower layer:

$$\Delta W = \eta \langle u_0 m_0 - u_1 m_1 \rangle_{batch} \quad (6)$$

$$\Delta b_u = \eta \langle u_0 - u_1 \rangle_{batch} \quad (7)$$

$$\Delta b_m = \eta \langle m_0 - m_1 \rangle_{batch} \quad (8)$$

where the subscript 0 indicates the activity after the first step of encoding the stimulus in the unisensory and multisensory layer, and the subscript 1 indicates the activity in the confabulation phase. It can be shown (Hinton, 2000) that this learning algorithm is approximately minimizing the information loss between the training data's probability distribution, and the lower layer's equilibrium probability distribution (that is, the distribution obtained after a sufficiently large number of up-down iterations). In more neuroscientific terms, when the training is complete, the network's spontaneous activity should closely resemble the activity induced by sensory stimulation. Since this learning rule contains one positive and one negative term proportional to local correlations, this is an Hebbian-anti-Hebbian learning rule. The learning rate η was set to 0.005, and the training was run for 100 epochs in total, with each epoch consisting of 400 batches of 100 samples. The whole process took about two hours on a standard desktop computer.

2.4 | Testing and simulating behaviour

After the training was completed, the network's features were assessed and compared to existing literature. While the receptive fields of the unisensory neurons in the lower layer are set a priori on the basis of prior knowledge from neurophysiological and computational studies, the receptive fields of multisensory neurons are learned during training, and can therefore be tested and compared with data from the literature. Moreover, the network was used to simulate behavioural experiments on multisensory integration, and the results were compared with behavioural data. In order to do so, it is necessary to establish a link between simulated neural activity and visuotactile interactions in behavioural experiments. The general procedure followed in this work was to decode the information contained in the multisensory layer after unisensory inputs are encoded together (i.e. integrated) in its shared representation. Since it would be very difficult to decode such information directly from the multisensory layer, we proceeded as Makin et al. (2013). In order to interpret the activity of multisensory units, their activity was projected down to the unisensory populations via a "down" pass through the feedback synapses (Figure 1b). Here, neural activity could be easily decoded, since the mapping between unisensory activity and the physical stimuli is defined a priori by the Gaussian tuning curves that we chose. It is sufficient to take the barycentre of the neural activity contained in the visual layer to decode the physical location of the visual stimulus encoded in the multisensory layer, the barycentre of the activity in the proprioceptive layer to decode the position of the hand with respect to the trunk, and, finally, the strength of the signal in the tactile layer to decode the intensity of tactile stimulation.

2.5 | Behavioural experiments

2.5.1 | Rationale

The network uses a simplified set of sensory inputs, as visual information about the hand's position and appearance is not present. In literature, hand PPS representation in humans was typically assessed through a simple tactile detection task, in which in which reaction times to tactile stimuli on the hand are measured in the presence of task-irrelevant auditory or visual stimuli, at various distances from the hand (Canzoneri et al., 2012). Using this paradigm, it was found that reaction times speed up (and tactile accuracy increases, as in Salomon et al., 2017) when the tactile stimulation is administered while the auditory or visual stimuli are closer to the body, with a stronger modulation in the case of looming stimuli. To our knowledge, visual information about hand position was always present in such experiments, and therefore the contribution of proprioception alone (simulated by the set of inputs of our model) was never assessed behaviourally. We therefore designed ad-hoc experiments to test whether proprioceptive information alone can generate a hand-centred PPS representation, that can be behaviourally detected through a tactile detection task. This was done by adapting the behavioural task described above to VR, allowing to keep the hand invisible while presenting visual stimuli close or far from its position in space.

2.5.2 | Materials

Tactile stimulation was delivered through rotating mass vibrators (Precision Microdrives), driven by a dedicated microcontroller. A hand-held button was attached to the same microcontroller, in order to collect reaction times to tactile stimulation on the same device and minimize unpredictable delays. Visual stimuli were delivered in a virtual reality scenario. A Head Mounted Display (HMD, Oculus Rift) was used, and rendering of the virtual environment was performed

through a custom made software (ExpyVR; http://lnco.epfl.ch/expyvr) coupled with the Steam VR software (SteamVR; https://www.steamvr.com/en).

2.5.3 | Participants

Forty-three healthy participants (19 females, aged 25 ± 3.7 *SD*, ranging from 23 to 41 years) were recruited for the study, and received monetary compensation for their time. Only right-handed participants with normal or corrected to normal vision were recruited for the study. The study conforms with the World Medical Association Declaration of Helsinki, was approved by the ethical committee of the Vaud canton, Switzerland, and was performed with the understanding and written consent of each subject.

2.5.4 | Procedure

Participants wore the HMD, and had two vibrators taped on the back of their right hand. They saw a virtual scenario reproducing a desk of the same size and location as the physical desk located in front of them, with a fixation cross located 15 cm above the desk and 65 cm in front of their trunk. They were instructed to keep their gaze on the fixation cross, and react as fast as possible when receiving tactile stimulation on the right hand, by pressing a button with the other hand, while trying not to pay attention to visual stimuli moving in their visual field.

2.5.5 | Design

The experiment used a within-subjects design, with two hand positions, run in counterbalanced between-subjects blocks. In "Hand right" blocks, participants placed their right palm on the desk about 30 cm in front of their trunk, and 25 cm right of their midline. In "Hand left" blocks, they placed the hand at the same distance from their trunk and 25 cm left of their midline. Within each block, four types of trials were present: three visuotactile trials and one unisensory. In visuotactile trials, participants saw a tennis ball starting from the fixation cross and moving at constant speed along one of three possible trajectories, directed towards one of three possible targets: "left," corresponding to the hand position in the "Hand left" blocks, "right," corresponding to the hand position in the "Hand right" blocks, and "receding," corresponding to a point located on the midline around 30 cm in front of the fixation cross (see Figure 4a). Participants received a well above threshold 100 ms vibrotactile stimulus (both vibrators were activated at the same time) at one out of three randomized delays from trial onset, to reduce the predictability of tactile stimulation (1.75, 2 or 2.25 s from trial onset). The ball motion started 500 ms after trial onset, and lasted for 2 s at around 22.5 cm/s, so that tactile stimulation was received when the ball was either at 0, 5 or 10 cm from the target. In unisensory trials, the same scenario was displayed, and the tactile stimulus was administered with the same randomized delay, but no tennis ball was displayed. For each hand position block, a total of 21 trials per trajectory (of which 7 per delay) was collected for visuotactile trials, plus 21 unisensory trials. In addition, a total of 36 trials, 30% of the total, were catch trials. In such trials, one of the three usual ball trajectories was displayed (12 trials for each trajectory), but no tactile stimulation was delivered. Each experimental block lasted around 8 min.

2.5.6 | Data preprocessing and analysis

Reaction times (RTs) longer than 700 ms were automatically discarded by the microcontroller. This threshold can be considered safe as the average RT was 264.75 ms, with an average within-subject *SD* of 33.9 ms, making it extremely unlikely to observe a true reaction time longer than 700 ms. Overall, subjects performed the task accurately, with 0.66% of omitted responses to tactile stimulation and 5.6% of false alarms (responses given in catch trials or before the stimulation). Such responses were discarded. We then removed outlier responses by discarding, for each subject and experimental block, RTs falling more than 2 median absolute deviations away from the median RT. This cut-off is a more robust equivalent of the standard cut-off at 2 SDs, as suggested by theoretical and empirical justifications in methodological work (Leys et al., 2013). The three randomized delays between 1.75 and 2.25 s were used purely to reduce the predictability of the task, and are orthogonal to the experimental conditions of interest. Seven trials for each delay were collected for each trajectory and block and the distance covered by the ball in the 500 ms randomization window is small compared to the distance between the three targets. This allowed us to overcome the possible confound introduced when the overt expectations due to the temporal delay of the tactile stimulation correlate with the position of the visual stimulus. Therefore, in our main analyses, we pooled trials from the three delays together, and only focused on the effects of hand position and ball trajectory. Similarly to what was done in previous studies (Serino, Canzoneri, et al., 2015; Serino, Noel, et al., 2015), we defined the multisensory facilitation as the difference between multisensory (visuotactile) and unisensory (tactile only) reaction times, and performed our analyses on this quantity. The multisensory facilitation was computed by averaging unisensory trials for each subject and experimental block and subtracting it from each visuotactile RT.

RTs were measured in a 2×3 design with the factors Hand position (Left, Right), and Trajectory, indicating the direction of the ball, (Left, Right, Receding). Trajectory was recoded as Congruency: Congruent, when the ball was moving towards the tactilely stimulated hand, Incongruent, when it was moving towards the opposite side, and Receding. Reaction times were then analysed by means of linear mixed-effects models. We fit a model on to multisensory facilitation (MF), including Congruency and Position as predictors. Different random structures were tested, assessing all the five possible combinations of Position and Congruency, including their interaction and just a random intercept, and the model giving the best fit was selected. Both in terms of Akaike and Bayesian Information Criterion, the best model was that which considered only the position as a random factor:

$$\mathrm{MF} \sim \mathrm{Position} + \mathrm{Congruency} + (\mathrm{Position}|\mathrm{Subject})\,.$$

Additionally, a model including a Position*Congruency interaction was tested, and statistical testing confirmed the selected random structure. Data preprocessing and further analysis was run in R (R version 3.4.4, for linear mixed-effects model: packages lme4 version 1.1-15 and lmerTest version 2.0-36). Linear mixed-effects models were tested using the Satterthwaite approximation for the degrees of freedom from the lmerTest package.

3 | RESULTS

3.1 | Learned connectivity and receptive fields

The neural network was designed to learn a connectivity scheme that optimizes the reconstruction of the patterns of stimulation observed during the training. This led to a spontaneous diversification of the response of neurons in the multisensory layer to the different sensory modalities. Since at initialization all the multisensory neurons are connected with all the neurons in unisensory populations, after the training all neurons were to some extent multisensory, meaning that they received some input from all the unisensory populations. Nevertheless, a great variability in the modality tuning of different neurons and thus in the pattern of sensory responses emerged. In order to quantify the response of each neuron to a given modality, we computed the sum of the absolute value of the strength of synapses from a given unisensory population, normalized by the mean input for that modality. In particular, we focused on the response of multisensory neurons to tactile inputs. As appears evident from Figure 2a, the response has a bimodal profile, meaning that the population spontaneously diversifies in inhibitory and excitatory neurons as a function of tactile inputs. In general, around 55% of the neurons were found to receive excitatory projections from the tactile area, with the remaining 45% receiving inhibitory projections.

In order to test whether and how the model might build up a PPS representation from capturing regularities in the environment, we tested how the spatial properties of the multisensory neurons depended on their tuning to the tactile modality. It is known from neurophysiological literature that the PPS is represented in the monkey cortex by a set of multisensory neurons that respond both to touch on a given body part, and to visual stimuli close to that body part. In the previous version of the model, this evidence was implemented by a hard-wired connectivity whereby the projections to the multisensory neuron(s) were of fixed strength from the tactile area, whereas from the visual area they decreased as a function of the distance from the hand. We asked whether our model could simply learn a similar pattern of connectivity from the multisensory training, in which neurons that respond more strongly to touch code mostly for the close (trunk-centred) visual space. Since our model uses several multisensory neurons, that spontaneously tune differently to each sensory modality, we tried to define a suitable approach to test this hypothesis. For each multisensory neuron, we defined its preferred visual distance as the preferred distance (along the anterior-posterior axis) of the visual unisensory neuron that projects the strongest excitatory synapse to that same multisensory neuron. Roughly, this corresponds to the peak of the visual RF of the multisensory neuron. This allowed us to explore how the properties of multisensory neurons vary depending on the region of the visual space that is stimulated. We found that, on average, the tactile input computed by multisensory neurons slightly decreases with their preferred visual distance, coded as described above (Figure 2b). Excitatory neurons tend have the peak of their visual RF close to the trunk while inhibitory neurons tend to have it in the far space. This goes in the same direction as the synaptic connectivity in the previous neural network model, but here the distance dependent modulation is much weaker, and does not clearly differentiate the close and the far space. This may seem surprising, but due to the width and complex shape of the RFs learned by most multisensory neurons, the visual preferred distance of a given multisensory neuron is not always informative. A multisensory neuron with the peak of its receptive field in the far space can still have a significant response to close stimuli, and vice versa. More importantly, since in our architecture the visual input was not coded in hand-centred coordinates, the presence of tactile input does not simply depend on the distance in the visual space, but on proprioceptive and visual information combined. Likely, the slight dependence of connectivity on distance is mainly explained by the fact that the proprioceptive hand position cannot be further than 60 cm away from the trunk. However, the presence of the

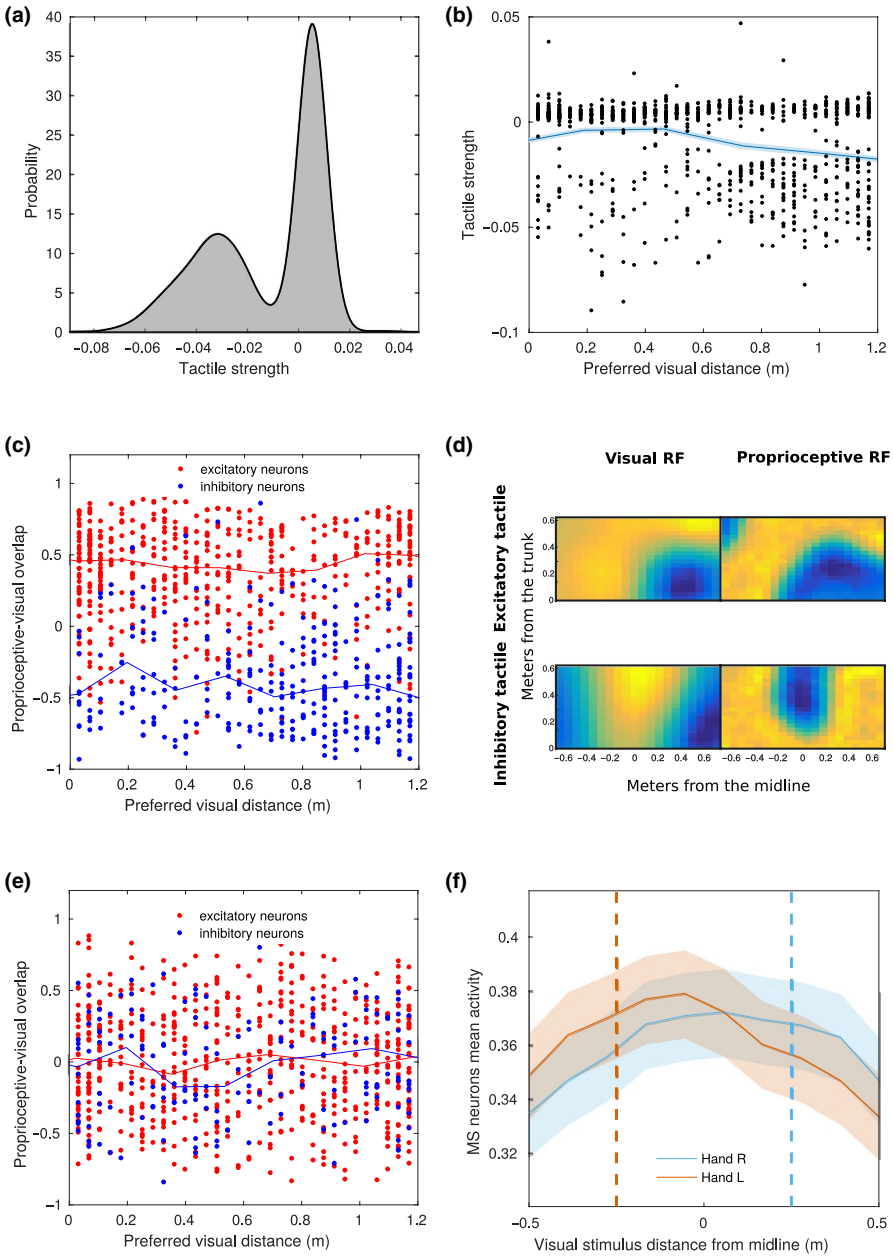


FIGURE 2 Properties of neurons in the upper layer. (a) Distribution of the strength of tactile input across the multisensory neurons. The strength of the input for each multisensory neuron is defined as the average of the synaptic weight of the projections it receives from the 30 tactile neurons. (b) Dependence of the strength of tactile input on the preferred visual distance of the multisensory neurons. The overlaid solid line represents mean values over 10 distance bins and the shade its standard error. (c) Quantification of the overlap of proprioceptive and visual receptive fields as a function of the preferred visual distance. The overlap is defined as the Pearson correlation coefficient of synaptic input to the multisensory neuron over space. Red and blue denote respectively multisensory neurons projecting excitatory and inhibitory synapses towards the tactile area. The overlaid solid lines represent mean values over 10 bins, with the shade representing the standard error. (d) Two exemplary visual (left) and proprioceptive (right) receptive fields of multisensory neurons. in the upper panels, a neuron receiving and sending excitatory projections to the tactile area, with overlapping visual and proprioceptive RFs. In the lower panels, a neuron receiving and sending inhibitory projections to the tactile area, with disjoint visual and proprioceptive RFs. Yellow and blue indicate respectively strong and weak projections from the unisensory areas to the multisensory neurons. (e) Same as panel c, but in a control model where tactile input was provided randomly and uncorrelated with visual and proprioceptive information. (f) Mean activity of the multisensory neurons that positively respond to touch, as a function of the position of the visual stimulus. The orange and light blue curves correspond to two different simulated positions of the hand, respectively, 25 cm left and right of the midline

proprioceptive population introduced an additional level of complexity in the neural network, that can be appropriately addressed only by looking at the associations learned by the network between the visual and the proprioceptive coding. To do this, we computed the overlap between proprioceptive and visual receptive fields of each multisensory neuron,

defined as the spatial correlation of its incoming visual and proprioceptive synaptic weights. This quantity approximatively corresponds to the spatial correlation of its visual and proprioceptive RFs. Since the spacing in the grid of neurons is different for the two populations, the correlation was computed after interpolating the proprioceptive synaptic weights on a grid of points with the same spacing of the visual population. A positive overlap means that the neuron tends to be activated when the hand and the visual stimulus are in the same position, whereas a negative overlap indicates that the neuron responds when the hand and the stimulus are far away. We expected the nature of the learned visuo-proprioceptive associations of a given multisensory neuron to depend on its response to tactile input, therefore we divided the heterogeneous population of multisensory neurons in two groups, based on whether they are inhibited or excited by tactile inputs. Then, we studied the dependence of such overlap on the preferred visual distance. For excitatory neurons, the visuo-proprioceptive overlap was strong regardless of their preferred visual distance. Conversely, for inhibitory neurons, the overlap was strongly negative at all preferred visual distances (Figure 2c). An exemplary pair of inhibitory and excitatory neurons with preferred distance in the close space are shown in Figure 2d. These results suggest that, more than differentiating between neurons coding for the close and the far space overall, the network spontaneously organized them in two populations of overlapping and anti-overlapping visual and proprioceptive RFs. Again, due to the width and complex shape of RFs, the presence of neurons with strong visuo-proprioceptive overlap, and preferred visual distance in the far space should not surprise. Crucially, in the present model, the alignment (or anti-alignment) of receptive fields emerges from the statistical regularity of touch with respect to an external visual stimulus and proprioceptive information. In order to demonstrate this, we replicated the simulation represented in Figure 2c after a control training with the same visual and proprioceptive stimuli, but in which touch was provided randomly and independently from the hand-centred coordinates of the visual stimulus. The visuo-proprioceptive overlap was always close to zero for both excitatory and inhibitory neurons (Figure 2e). In order to establish a comparison with the neurophysiological literature, we then studied the subset of neurons in the multisensory layer that positively respond to touch, to compare our artificial neural population to the one typically studied in primates (Fogassi et al., 1996; Graziano et al., 1994, 1997). In Figure 2f we show the average response of such neurons as a function of the position of the visual stimulus, in two conditions: hand to the left and to the right of the body midline. The average receptive field of the population shifts according to the hand position, in a similar way to what was reported by Graziano for individual neurons (Graziano et al., 1997).

3.2 | The network encodes tactile predictions in hand-centred coordinates

In an RBM, information from the different unisensory populations of the lower layer is encoded in the upper layer in a unified and compressed representation, embedding the statistical relations between the unisensory inputs. This allows the network to build a more compact and accurate representation of the input than each of its unisensory components (Makin et al., 2013), and can be seen as a predictive form of multisensory integration, in which inputs from different modalities influence and complement each other to better fit a global model of sensory inputs. We hypothesize that PPS representation spontaneously emerges when a neural network learns to integrate in such a way external and body-related information, being trained on sensory inputs that reflect the natural statistics of body-environment interactions. In practice, to efficiently encode incoming sensory information, multisensory neurons in our network must learn an encoding schema that embeds the statistical relations observed between tactile stimulation on the hand and visual stimuli close to the hand (the hand position being specified via proprioceptive information). As a consequence, when a visual stimulus is present close to the proprioceptively encoded hand position, we expect the multisensory neurons to start coding for the presence of tactile stimulation at a sub-threshold level, before or even in absence of contact. This prediction constitutes a possible explanation of the well-reported effect of a facilitation of reaction time to tactile stimulation in the presence of an external stimulus approaching the stimulated body part (Canzoneri et al., 2012; Serino, 2019). Indeed, the "pre-encoding" of tactile information in multisensory neurons might be not sufficient to elicit conscious tactile perception, but might boost responsiveness to tactile stimulation, thus speeding up reaction times when tactile stimulation is delivered. Following this line of reasoning, testing our hypothesis becomes equivalent to performing in-silico simulations of tactile detection tasks such as in (Canzoneri et al., 2012). Practically, this can be done by providing proprioceptive and visual information to the neural network, while suppressing the input from the tactile area, so as to measure only the contribution of vision and proprioception on the tactile information encoded in the multisensory layer. The activity we read out from the tactile population (even in absence of tactile stimulation) is used as a proxy of multisensory facilitation, i.e., faster reaction times in a tactile detection task. We call this read-out tactile information evoked tactile activity, and treat it as a in silico behavioural correlate of PPS representation. Note that this does not necessarily mean that behavioural effects in reaction times reduction are linked to actual activity in tactile unisensory areas, as behaviour may be based on the amount of tactile information contained in the multisensory layer, that we only decode through feedback synapses. Consistently with our previous theoretical reasoning

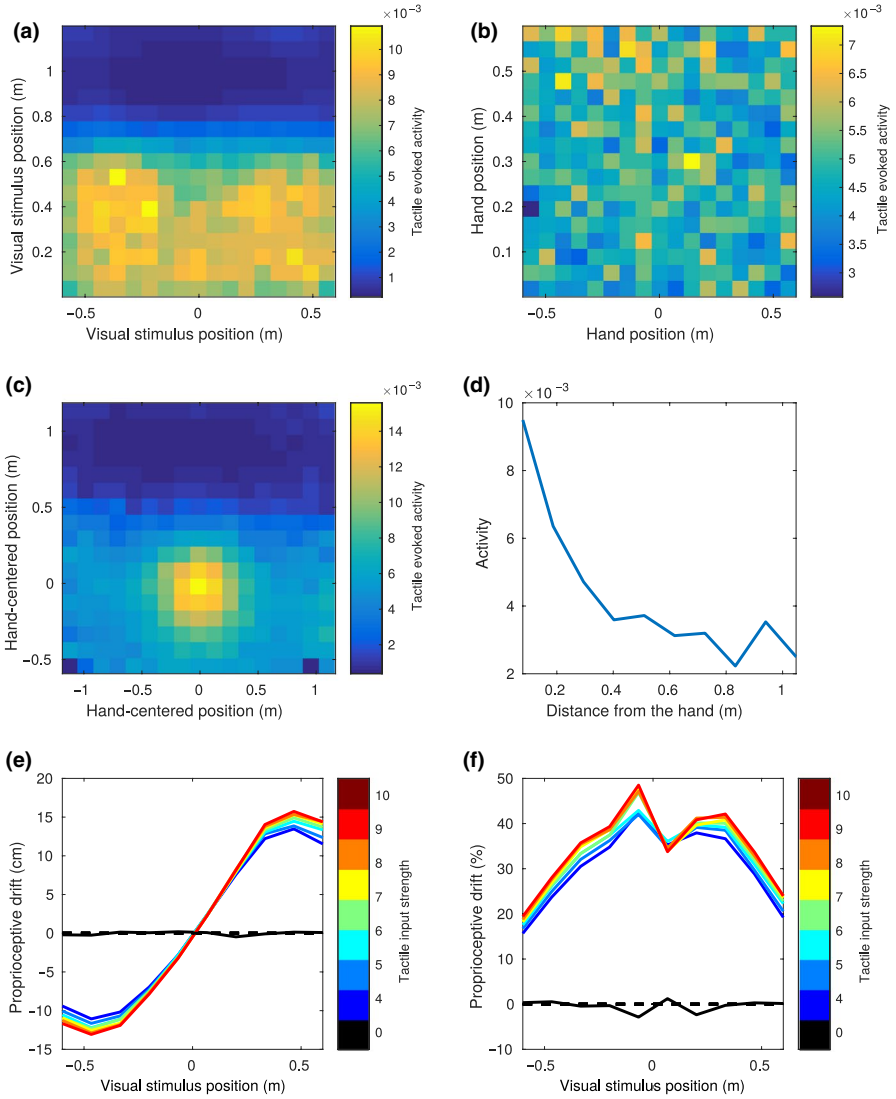


FIGURE 3 Simulated behavioural experiments. (a and b) Tactile evoked activity - multisensory facilitation as a function of visual stimulus position (in trunk-centred coordinates) and hand position. The evoked tactile activity is obtained by setting the tactile input to zero, encoding a visual and a proprioceptive input, and reading out the tactile information encoded in the multisensory area from the tactile area (i.e.: its mean activity after a "down" pass). In trunk-centred coordinates (a) stronger activity for close positions of the visual stimulus can be observed, but no modulation as a function of the position along the anterior-posterior axis. Virtually no modulation is observed as a function of hand position (b). (c) The same tactile evoked activity, plotted as a function of the visual stimulus position in hand-centred coordinates. (d) Tactile evoked activity as a function of the distance from the centre of the hand of the visual stimulus. (e) Simulated proprioceptive drift in the invisible hand illusion. The proprioceptive input is fixed at the midline, and the position of the visual stimulus is shifted across the midline. The plot shows the proprioceptive position reconstructed by the network after integrating the three sensory inputs. The *x* axis represents the distance from the midline of the visual stimulus. Different colours represent different levels of intensity for the tactile input, starting from black (no touch/asynchronous stimulation), to red (maximal intensity of tactile stimulation). (f) Same as panel d, but the proprioceptive drift is expressed as the percentage of the distance between the visual and proprioceptive stimuli

and with neurophysiological and behavioural findings, we expect the evoked tactile activity to depend strongly on the location of the external stimulus with respect to the hand position, i.e., on its hand-centred coordinates.

To test this hypothesis, we ran a simulation in which the hand and the visual stimulus were placed at random positions within the respective areas and measured the evoked tactile activity as a function of the position of the hand and of the visual stimulus. As expected, the region of space in which visual stimuli elicit an evoked activity in tactile neurons was spatially anchored to the hand. In particular, if the evoked tactile activity was displayed as a function of hand position or visual position of the stimulus in trunk-centred coordinates (Figure 3a,b), there was a weak, non-coherent modulation of activity. Instead, if the activity was displayed as a function of position of the visual stimulus with respect to the hand position as coded by the proprioceptive population, i.e., hand-centred, (Figure 3c), the modulation became stronger and coherent,

with a maximal level of activity when the stimulus was in the origin (i.e. the centre of the hand), sharply decreasing with distance. Figure 3d shows the trend of tactile evoked activity as a function of the distance of the visual stimulus from the centre of the hand. This curve shows a similar trend to what reported for some neurons mapping the PPS representation around the monkey face (Graziano et al., 1997).

3.3 | In-silico results match in-vivo hand-centred coding of multisensory facilitation

In order to confirm that the proposed architecture can model actual behaviour in a meaningful way, we ran an ad-hoc behavioural experiment on healthy participants. The aim of the experiment was to show that proprioceptive information is integrated with information about an incoming visual stimulus, affecting tactile processing on the hand. As a behavioural proxy of multisensory integration, we measured reaction times to tactile stimulation on the right hand, while the subjects were seeing task-irrelevant visual stimuli (tennis balls) in virtual reality. RTs were compared for Hand position (Left, Right) and Congruency of the ball trajectory (Congruent, Incongruent and Receding). A linear mixed-effects model on the multisensory facilitation (MF), including Congruency and Position as predictors (see Methods for details), showed a significant main effect of Congruency ($F(2, 4{,}831.8) = 6.389$, $p = 0.0017$) and a marginally significant effect of Position ($F(1, 42.1) = 3.59$, $p = 0.065$). When looking at individual coefficients, using the Receding trajectory as a reference, we found Congruent trials to be significantly faster (−4.115 ms, $SE = 1.184$ ms, $T = -3.477$, $p < 0.001$), and Incongruent trials to be not significantly different (−1.216 ms, $SE = 1.188$ ms, $T = -1.024$, $p = 0.30$) from receding trials. In order to directly compare Congruent and Incongruent trials, and assess the role of proprioception in visuotactile integration, we fit the same model on the subset of Congruent and Incongruent trials. Again, the main effect of Congruency was significant ($F(2, 3{,}202.4) = 5.912$, $p = 0.015$), with Congruent trials faster than Incongruent trials (−2.889 ms, $SE = 1.188$ ms, $T = -2.432$, $p = 0.015$). This is in line with the model's qualitative predictions, shown in Figure 4c.

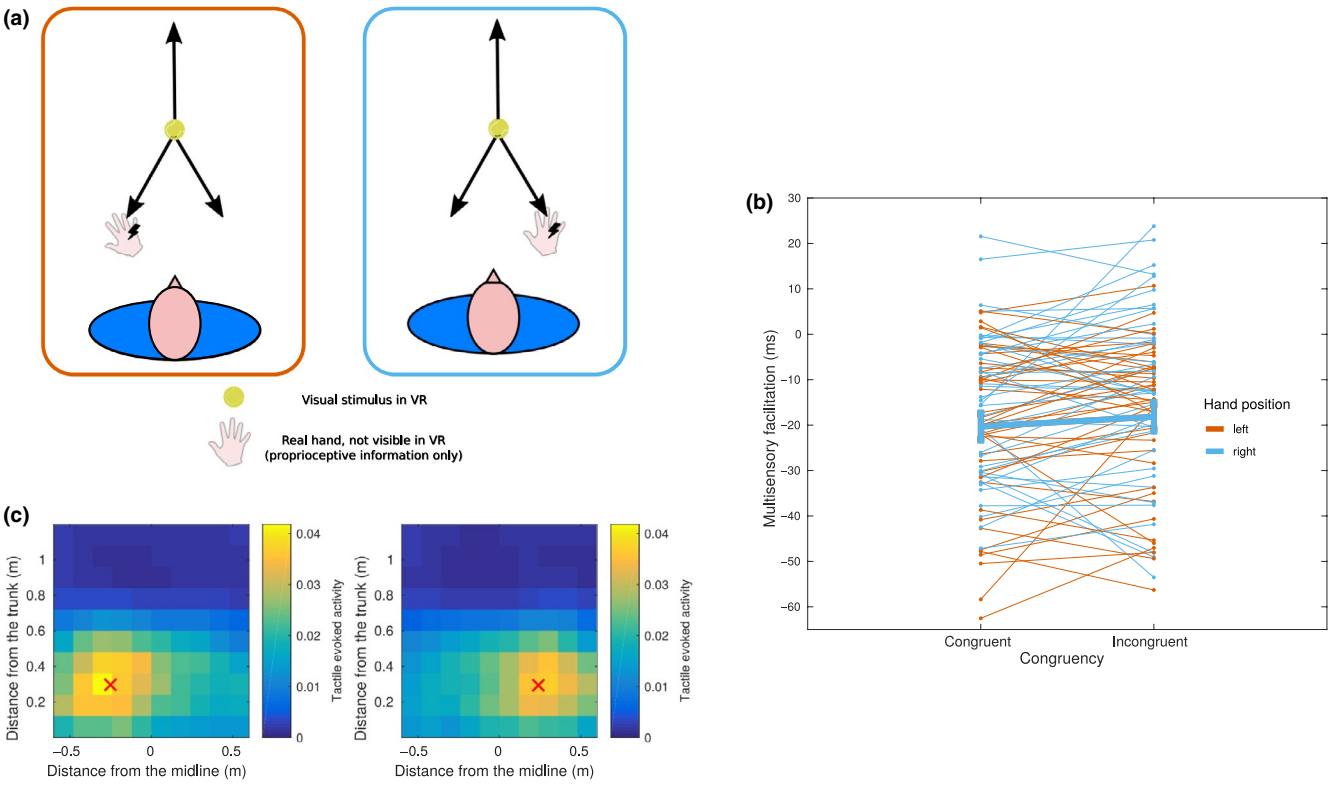


FIGURE 4 Results of the behavioural experiment. (a) Schematic experimental setup. The subjects placed their right hand approximatively 30 cm in front of their trunk, either 25 cm left or right of their midline. The origin of the arrows represents the starting point of the different trajectories, coinciding with the fixation cross. The total length of the trajectories was approximatively 50 cm. (b) Modulation of average reaction times for the 43 participants as a function of hand position and ball trajectory congruency with hand position. For simplicity, we show only the two conditions that are relevant for confirming our hypothesis, and leave out the receding condition. Thick lines indicate global means by condition. (c) Expected results from model simulations for the same experimental setup. Red crosses represent the position of the real hand's centre, the colour coding represents the predicted multisensory facilitation. Yellow areas represent zones of higher facilitation/faster reaction times

Additionally, to rule out the possibility that the congruency effect may be present only on one side of the midline, we run the same model including a Position*Congruency interaction. This did not change the main effect of Congruency ($F(2, 4{,}831.8) = 6.389$, $p = 0.0017$), nor the comparison Congruent versus Incongruent($F(1, 3{,}201.3) = 5.90$, $p = 0.015$), and the interaction was not significant ($p = 0.76$). Note that the facilitation when the visual stimulus is on the opposite side of the midline (Incongruent trials) is close to zero, similarly to when the visual stimulus is in the region outside the PPS (Receding trials), in line with the non-significant difference found in our experiments. It is worth noting that the multisensory facilitation compared to unisensory trials was significantly below zero in all conditions, including Receding trials (−14.6 ms, $SE = 2.45$ ms, $T = -5.96$, $p < 0.001$). This seems to contradict model predictions, as no significant tactile evoked activity is expected in the far space. We hypothesize that this may be due to overall stronger expectation effects in multisensory trials, compared to unisensory trials, due to the presence of the virtual ball providing a more precise cue about the likely time of stimulation. Our experimental design, unlike most previous studies, makes the delay of stimulation orthogonal to the three conditions of interest, which allows comparing them while controlling for expectation. We also investigated more in detail the possible interactions between PPS representation and expectation effects, by analysing trials separately by stimulation delay. We found an overall effect of stimulation delay reducing reaction times, compatibly with the presence of expectation effects. However, the decrease in reaction times with increasing delay (and decreasing distance from the hand) was significantly stronger in the Congruent condition, further confirming the presence of proximity effects in modulating reaction times (see Supporting Information and Figure S4 for details).

3.4 | The network encodes proprioceptive inputs as a function of visuotactile integration – The IHI

Our previous computational (and behavioural) results show how visual and proprioceptive information combined can affect the encoding of tactile information to reproduce the associations learned during the training (or real-life experience). Since the learned associations have no preferential direction, we expect the transfer of information between sensory modalities to take place also in the opposite direction: from the tactile to the visual and proprioceptive modalities. In particular, we focused on how visuotactile inputs affect the encoded proprioceptive information, as this link has been previously investigated in several behavioural works exploring the multisensory bases of body representation (Guterstam et al., 2013; Salomon et al., 2017). We fixed the input hand position at the midline and provided visual stimulation at different positions along the anterior-posterior axis. This was done in association with no tactile inputs (touch OFF), or at various levels of intensity of tactile stimulation (touch ON). We can consider the "touch ON" conditions as synchronous stimulation, in which touch and visual stimulation occurred at the same time, and the "touch OFF" as asynchronous stimulation, meaning that visual stimulation and touch were sufficiently separated in time to have no residual activity in the tactile area when visual stimulation occurred. Then, as we previously did with the tactile population, we projected the multisensory activity to the proprioceptive population, and computed the integrated proprioceptive position as the barycentre of neural activity. In the "touch ON-synchronous" condition, we found that the proprioceptively encoded position of the hand gets attracted towards the position of the visual stimulation. This result held with little changes at different intensities of tactile stimulation, as if the presence of tactile stimulation was treated as an all or none variable to generate the attractive pull (Figure 3e,f). Only at zero tactile intensity, in the "touch OFF-asynchronous" condition, was the reconstructed proprioceptive position roughly unbiased and did it correspond to the actual proprioceptively encoded hand position. These results resemble behavioural findings reported by Guterstam et al. (2013) when introducing the so-called "IHI." In the IHI, the hand of a participant is hidden, and tactile stimulation is provided while synchronously stroking the empty space next to the location of the real hand. Thus, as in our model, the subjects receive visual information about an external stimulus, touch on the hand, while processing proprioceptive cues, while they do not get any visual information about the hand position. Participants report feeling to have an "invisible hand" and when asked to point at the location of their real hand, they aim to a location shifted towards the point in space where the visual stroking occurred, a phenomenon known as proprioceptive drift. The output of the proprioceptive population in our model simulation in the "touch-ON" condition replicates proprioceptive drift in the IHI.

3.5 | Development of the key features of the network during training

After outlining the main features of the network, we explored how, during the training, these develop from the initial random connectivity. The results are summarized in Figure 5. To simply quantify the overall progress in the training of the network we computed the reconstruction error. This quantity is defined by encoding a sensory input in the multisensory layer and then projecting it back to the unisensory areas. The mean squared difference between the original input and the reconstructed activity is called reconstruction error, and it is

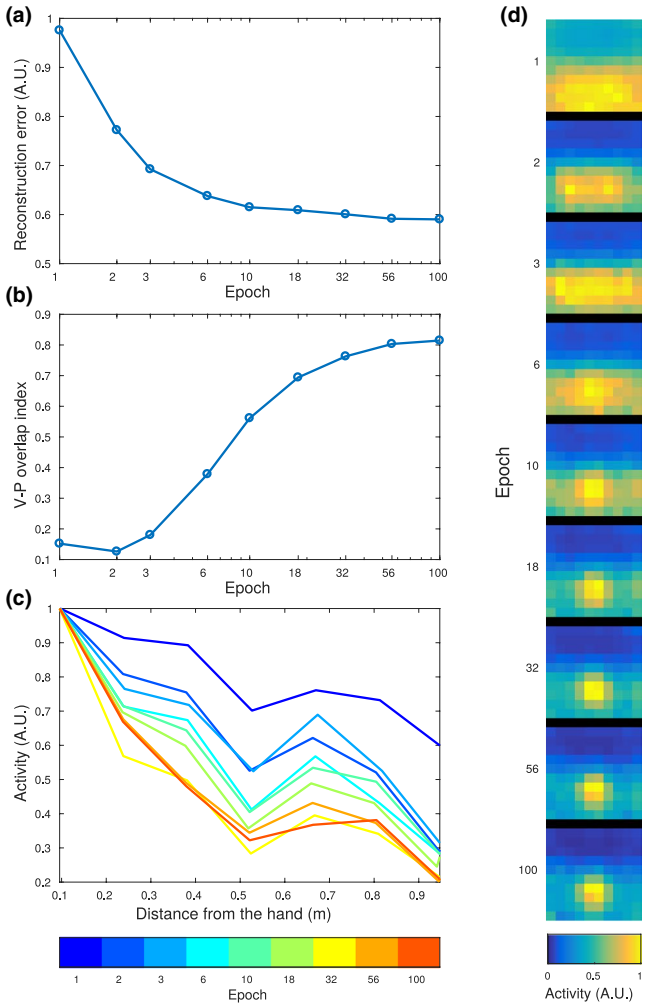


FIGURE 5 Evolution of the network during training. (a) Reconstruction error of the network plotted as a function of the training epoch. The reconstruction error is defined as the mean squared difference between the training sensory input and its reconstruction in the confabulation phase. (b) Visuo-proprioceptive overlap index across the 9 training epochs. The visuo-proprioceptive overlap index is defined as the difference between the average visuo-proprioceptive overlap of tactile excitatory and tactile inhibitory neurons. The stronger the overlap for tactile excitatory neurons, and the stronger the anti-overlap for tactile inhibitory neurons, the higher the index is. (c) Evoked tactile activity as a function of the distance from the hand of the visual stimulus, across the same nine epochs of training. (d) Evoked tactile activity as a function of the position of the stimulus expressed in hand centred coordinates. The activity is plotted for the same nine stages of training

expected to decrease during the training as the network learns to more efficiently encode its sensory inputs. As the training progressed, the reconstruction error decreased (Figure 5a), meaning that the network learned to reproduce more reliably the information contained in the unisensory inputs, after encoding it in the multisensory layer. After the initial strong decrease of the reconstruction error (from epoch 1 to epoch 6), the learning slowed down, and continued at a reduced pace throughout the whole training, probably towards the saturation value due to the stochasticity of the network's update rule. In order to synthesize the information about the overlap of visual and proprioceptive receptive fields, and display its evolution across epochs, we define a visuo-proprioceptive overlap index. The visuo-proprioceptive overlap index is defined as the difference between the average visuo-proprioceptive overlap of tactile excitatory and tactile inhibitory neurons. At the beginning of the training, the overlap index was low and close to zero, meaning that inhibitory and excitatory tactile neurons are not differentiated in terms of visual and proprioceptive RFs. During training, the value progressively increased, reaching almost the final value after epoch 18 (Figure 5b). This seems to coincide with the emergence of a strong tuning of the tactile evoked response to the distance from the hand (Figure 5c). As seen in Figure 5d, in the first stages of training, the reconstructed tactile activity was coarsely determined by the distance from the body of the visual stimuli. At this stage, the network has only learned that touch is more likely to occur if a visual stimulus is in the closed space, and still does not take proprioceptive information into account. Starting from epoch 10, and more clearly from epoch 18 and onwards, the network's response became tuned to hand-centred coordinates, as determined by proprioceptive signals.

3.6 | Visually encoded hand position

In the present work, we limited the inputs about hand position to proprioceptive information. This was done mainly to minimize the network's complexity and the number of input populations, facilitating the task of reverse engineering the network's functioning. Nevertheless, it is known from neurophysiological literature that visual input about arm (or even artificial reproductions of the arm) position affects the response of some PPS neurons (Graziano, 2000). However, since proprioceptive and visual information are redundant, at least in normal conditions, we predicted that adding visual cues about the hand position would not affect significantly the main properties of the network. To show this, we trained another network identical to the one shown in the previous paragraphs, with the addition of another visual population, coding for the location of the hand in space, through the same population coding and tuning curves used for the external visual stimulus (Figure 6a). In 75% of the training examples, the additional visual population coded for the same position in space as the proprioceptive population. In addition, to model occlusion of the hand by other objects or its exclusion from the visual field, we suppressed visual information about hand position in 25% of the training examples. To model the vision of other people's hands, the visually and proprioceptively encoded positions of the hand were independent in

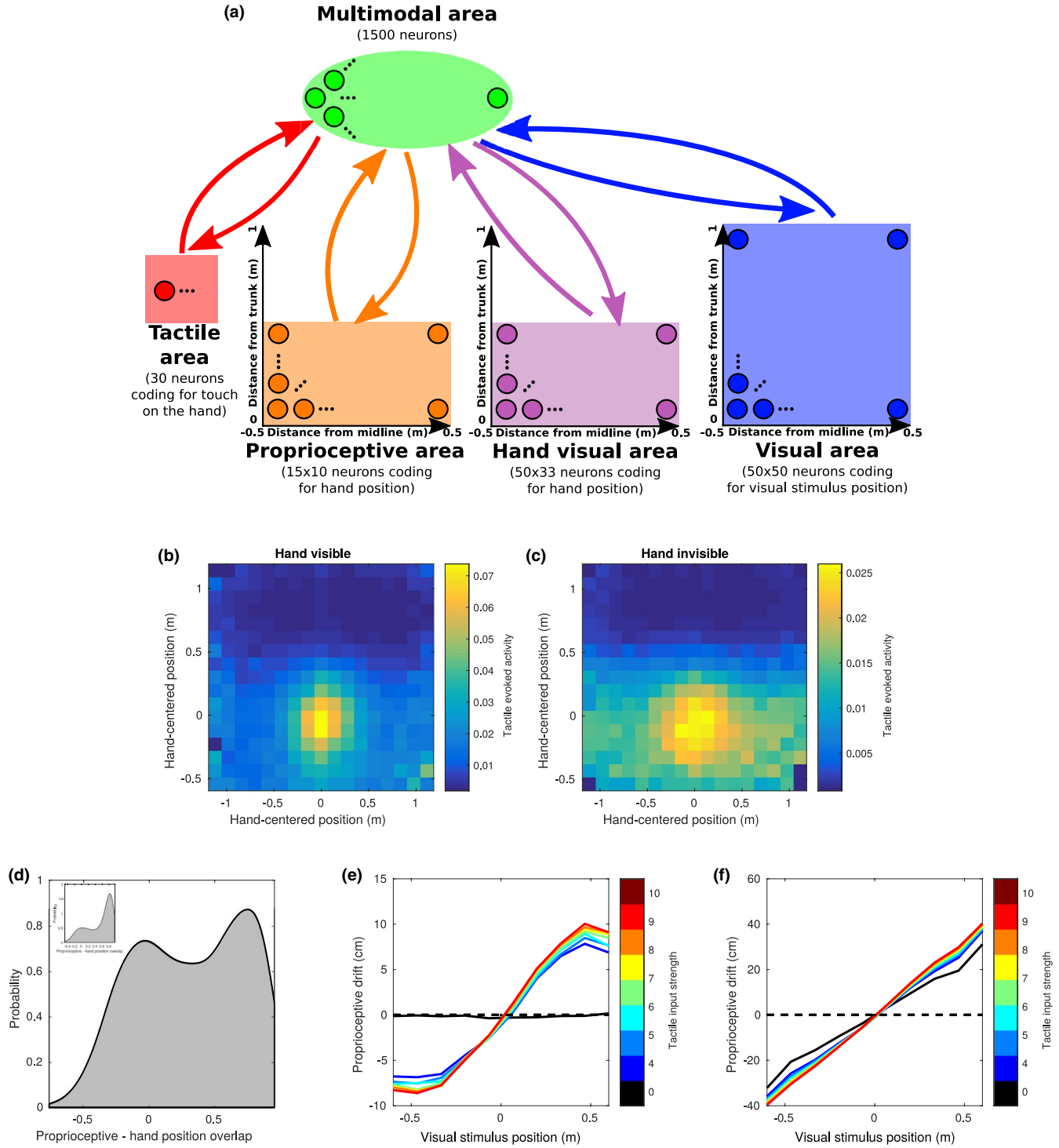


FIGURE 6 Network including visual information about hand position (a) Architecture of the network. In addition to the previous model, this network has one visual population (purple one) coding for the position of the hand. The tuning curves of neurons in this population have the same width as in the visual population coding for the position of the external stimulus. Other populations' tuning curves and training parameters were the same as in the previous model. (b) Tactile evoked response as a function of the position of the stimulus expressed in hand centred coordinates. (c) Same as panel (b), but the activity in the visual population coding for the hand was set to 0, simulating the occlusion of the hand and reproducing the sensory input of the previous model. (d) Distribution of the overlap between proprioceptive receptive fields and the receptive fields of the visual population coding for hand position. The inset shows the same result, in a network in which the proprioceptive and visual hand positions were never dissociated. (e) Proprioceptive drift in the simulated invisible hand illusion. We followed the same procedure as for Figure 3e, and set the activity in the visual population coding for hand position to 0 to simulate the occlusion of the hand in this network. The x axis represents the distance from the midline of the visual stimulus. (f) Proprioceptive drift in the simulated rubber hand illusion. The procedure was the same as for the invisible hand illusion, with the exception that the visual hand area was now coding for the same location as the external visual stimulus

25% of the trials. Therefore, we modelled the hand visual area as a neural population coding for the spatial location of hand-like objects in space, without recognizing the specific visual features of one's own hand. Then, we run the same set of analyses as in the previous paragraphs. In Figure 6b,c, we show how the network encodes information in hand-centred coordinates, similarly to what shown in Figure 3b. We tested the network both in the case of visible (Figure 6b) and invisible (occluded) hand (Figure 6c) and found comparable results, the only difference being a weaker evoked activation of the tactile area when the hand was not visible. Even when the network was trained with both visual and proprioceptive information, proprioception alone was sufficient to build a visuotactile PPS representation. In Figure 6d we provide a simple explanation for this: in the majority of multisensory neurons, the learned proprioceptive and hand-visual receptive fields were strongly overlapping, as the two populations typically code for the same spatial location. Interestingly, the visuo-proprioceptive overlap distribution in Figure 6d presents a secondary peak at zero overlap, besides the main peak around 0.75, showing that the receptive fields were completely dissociated in a minor yet significant fraction of the multisensory neurons. Further testing showed that this was the case only when the network had been exposed to the dissociated visual and proprioceptive hand positions (others' hands) during training. When training an identical network, in which visual and proprioceptive hand positions were always overlapping, the zero overlap peak was greatly reduced, as seen in the inset of Figure 6d. This may reflect the network learning to differentiate between integration and segregation of visuo-proprioceptive information (see Section 4). We then tested the IHI, by providing the same inputs as previously done for Figure 3e,f in the visual, tactile and proprioceptive populations, and no input in the hand visual population. The results closely matched the ones of the previous model (Figure 6e). Moreover, this extended network architecture reproduced the stimulation pattern of the rubber hand illusion. We fixed the proprioceptive hand position, while encoding an incongruent hand position in the visual hand area, representing the rubber hand. At the same time, we provided visual stimuli at the same location as the visual hand, representing the stimulating brush, and either no tactile stimulation or touch at various intensities. Then, we read out the proprioceptive hand position, by projecting multisensory activity down to the proprioceptive population. We observed a significant proprioceptive drift towards the rubber hand in the touch ON-synchronous that was weakly modulated by tactile intensity (Figure 6f). A significant proprioceptive drift was observed also in the touch OFF-asynchronous condition, although clearly smaller than in the synchronous condition. This result, seemly surprising, is actually in line with behavioural reports of a significant proprioceptive drift towards the rubber hand even in the case of no or asynchronous visual stimulation (Rohde et al., 2011; Samad et al., 2015).

3.7 | Shifting receptive fields at the level of single neurons

In the previous paragraphs, we showed how the network can encode information in hand-centred coordinates at the population level. This allowed to reproduce some important behavioural and neurophysiological aspects of PPS representation. However, while neurophysiological studies reported individual neurons with visual receptive fields spatially anchored to body parts in space (Graziano, 1999, 2000), the receptive fields of individual multisensory neurons in our network cannot be spatially "shifted" by proprioceptive inputs. Mathematically, this is a direct consequence of the fact that the network has only two layers, and that the response of one multisensory neuron is a sigmoidal function of the sum of its inputs, with the visual and proprioceptive inputs being independent. Since the sigmoid is a monotonically increasing function, when changing the proprioceptively encoded hand position, the neuron's response as a function of the visual stimulus' position would either increase or decrease everywhere, but do not change its global spatial properties. More specifically, the peak of the receptive field would not change. However, since the two-layers network learned to encode information in hand-centred coordinates at the population level, we expect that the addition of a third multisensory layer could lead to individual neurons with visual receptive fields anchored to body parts in space. We therefore trained a further model to provide an example of how fully hand-centred receptive fields at the single neuron level can be achieved by simply expanding our two-layers architecture. The new network had the same architecture as in our previous model, but with reduced overall number of neurons, to keep its computational complexity manageable during the learning task. We then added a second, "higher level," multisensory layer, receiving inputs from the first multisensory layer and from the tactile area (Figure 7a). The training was performed in two steps. In the first step, connections between unisensory areas and the first multisensory layer were trained as shown before, with contrastive divergence and coupled unisensory inputs (Figure 1b). After completion of the first step of training, connections from the tactile area and the first multisensory layer to the second multisensory layer (denoted by black arrows in Figure 7a) were again trained with contrastive divergence. The stimuli were generated by encoding unisensory inputs from the usual training set in the first multisensory layer and using the so obtained activity, coupled with activity in the tactile area, as training input for the second multisensory layer. The hypothesis underlying the emergence of shifting RFs from this architecture

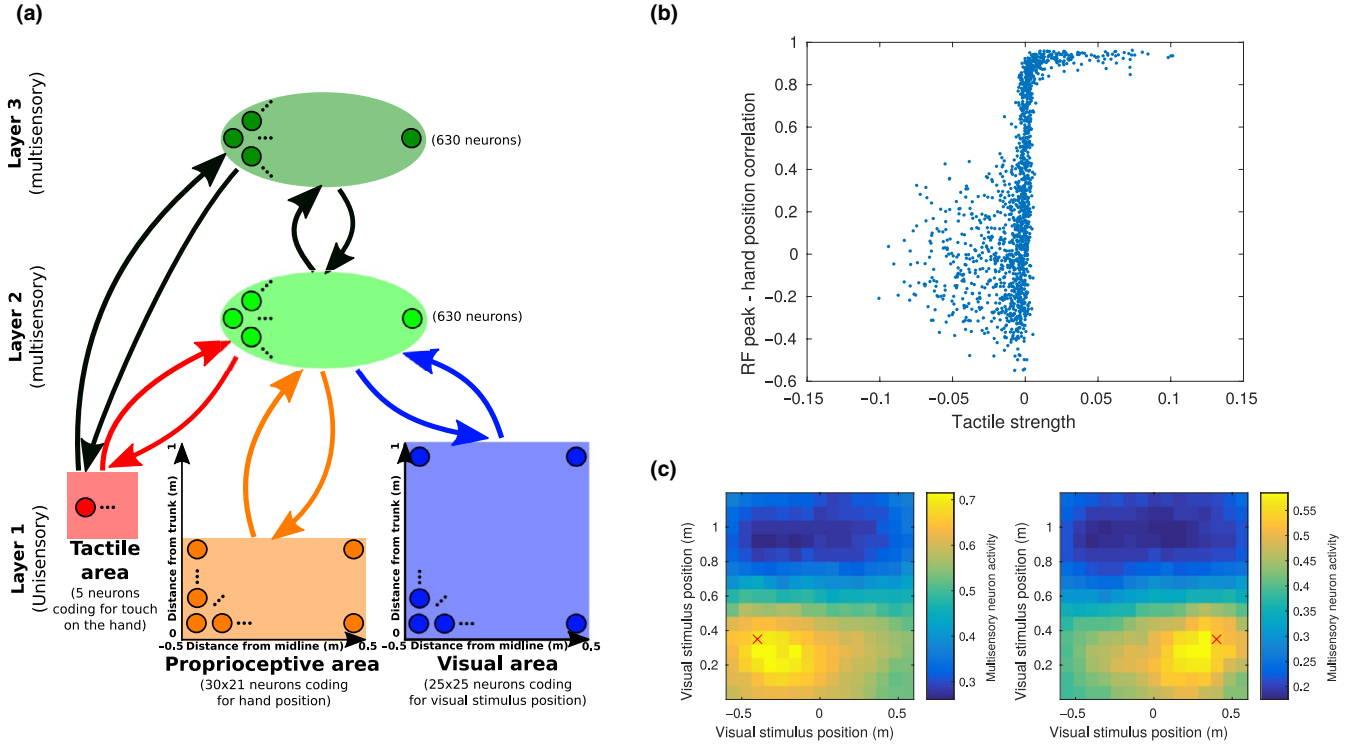


FIGURE 7 Individually shifting receptive fields. (a) Architecture of the network. The first two layers have the same architecture as in the main model, but fewer neurons to facilitate the training. The third layer is connected to the second multisensory layer and to the tactile population in the unisensory layer. The training was performed in two steps. The first step was identical to the original model. In the second step, training inputs for the second multisensory layer were constituted by the joint activity of first multisensory layer neurons and unisensory tactile neurons. (b) Correlation between hand position and RF peak of second multisensory layer neurons, as a function of the strength of the input they receive from unisensory tactile neurons. The correlation is defined as the average between correlations along the *x* and *y* directions. (c) Visual receptive field of one exemplary multisensory neuron in the third layer, receiving strong excitatory projections from the tactile area, for two different hand positions. Each subplot corresponds to different position of the hand, indicated by the red cross overlaid to the receptive field

stems from our previous observation that multisensory neurons that respond to touch have overlapping visual and proprioceptive receptive fields. We expect individual neurons in the second multisensory layer to learn the associations between (unisensory) tactile activity and the activity of neurons in the first multisensory layer that code for touch and whose visual and proprioceptive receptive fields overlap. If this is the case, third-layer neurons would learn to be active when *any* of the neurons coding for touch in the first multisensory layer is active. That is, they would respond whenever vision and proprioception are aligned, by shifting the peak of their visual receptive field. We therefore expected that the neurons receiving the strongest projections from tactile units would exhibit a stronger tuning to hand position. We explored this hypothesis by setting tactile inputs to zero, and mapping the peak of the RF for 100 different hand positions. We then computed the average correlation (along the *x* and *y* axis) between hand position and RF peak, as an index of hand position tuning. As shown in Figure 7b, neurons receiving the strongest projections from unisensory tactile neurons show the highest degree of hand position tuning, with 27.5% of them having an average correlation coefficient above 0.6. An example of a hand position tuned neuron can be seen in Figure 7c.

3.8 | Encoding proprioceptive input in joint angles

In the previously presented models, we chose to encode proprioceptive input as a population coding in Cartesian space. This was done to simplify the interpretation and visualization of the results. However, the encoding of raw proprioceptive input likely resembles joint angles more than Cartesian space. Here, we demonstrate how the main results of our work can be recovered when training a network with proprioceptive inputs encoded under the form of more biologically realistic joint angles. In this version of the network, the proprioceptive population was still 15 × 10 neurons, representing, respectively, the angle of the shoulder in the horizontal plane and the angle of the elbow. Shoulder angles ranged from $-\pi/4$ to $\pi/2$, where 0 represents the arm straight ahead and negative and positive angles represent a deviation towards the body midline or away from it, respectively. Elbow angles ranged from $-\pi/2$ to 0, where 0 represents

the arm fully extended. Visual and tactile inputs were encoded through a population coding with Gaussian tuning curves, with the same width as in the main network. For training, pairs of proprioceptive and visual positions were drawn from uniform distributions in the respective range (therefore, joint angles were randomly drawn for the arm instead of positions in the Cartesian space). Then, feedforward kinematics were computed to determine hand position in Cartesian coordinates, and tactile units were activated if the distance between the visual stimulus and the hand was smaller than 15 cm.

$$x_{hand} = 0.3\sin(\theta_1) + 0.35\sin(\theta_1 + \theta_2)$$
$$y_{hand} = 0.3\cos(\theta_1) + 0.35\cos(\theta_1 + \theta_2) \quad (9)$$

where 0.3 and 0.35 represent the length of the arm and forearm (up to the centre of the hand), respectively, and θ_1 and θ_2 represent respectively shoulder and elbow angles.

Figure 8b,c shows the same results as Figure 3c,e, for this version of the network. The network is able to compute hand-centred coordinates of visual stimuli in a similar way to what already shown in the main results. Also when testing the proprioceptive drift in the IHI similar results were obtained, except for a slight bias in the reconstructed hand position. This is in line with our expectation that, as long as the network is able to learn a good generative model of its sensory inputs, the specific encoding schema should not matter, provided that information about the physical stimuli can be recovered from the unisensory populations.

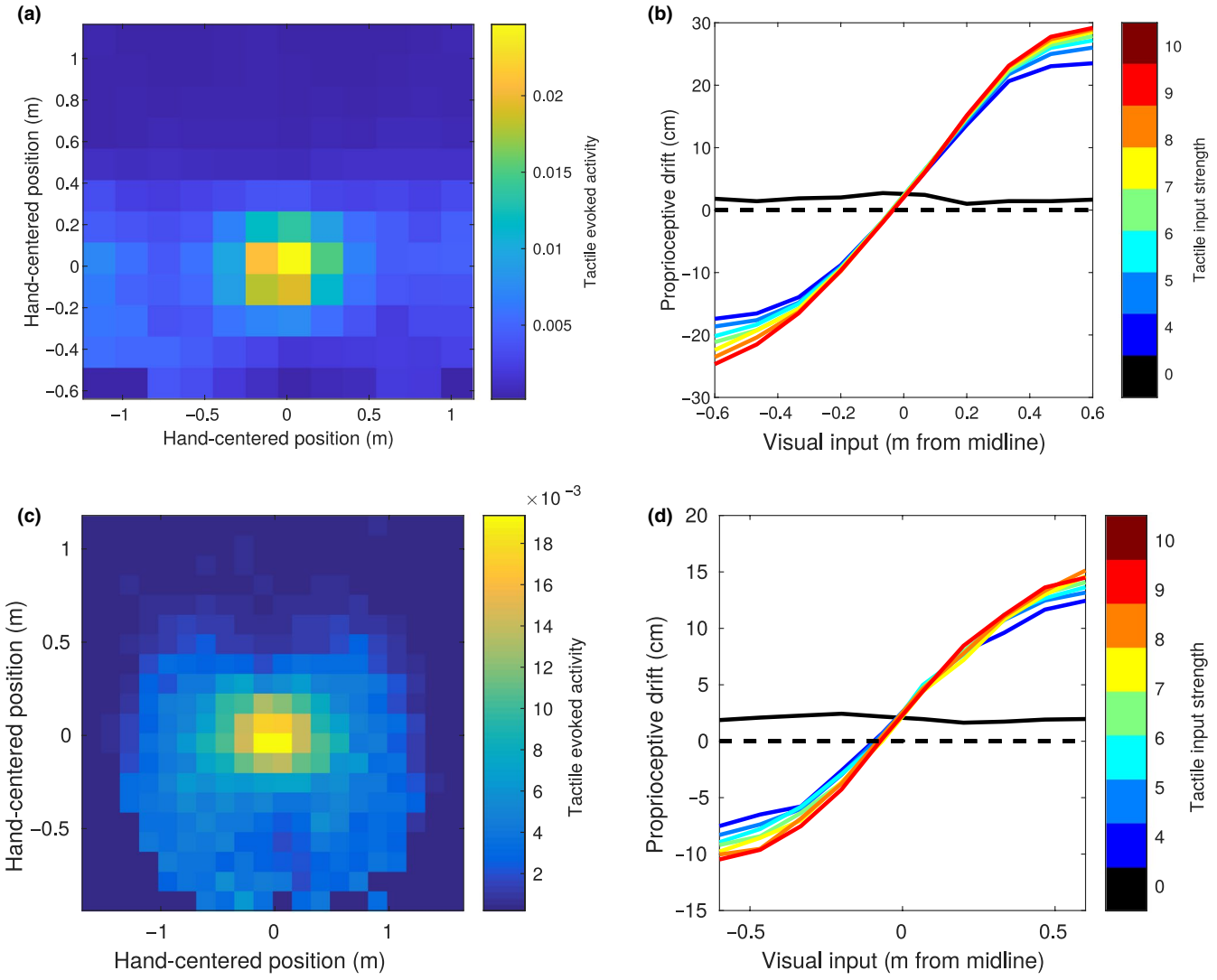


FIGURE 8 Further network generalizations. Panel (a) shows of the same analyses shown in Figure 3c, for a network in which hand position was encoded under the form of shoulder and elbow joint angles. Panel (b) reproduces Figure 3e for the same network. Panels (c) and (d) demonstrate the same results as panels (a) and (b) in a network where, in addition to encoding proprioceptive inputs under the form of joint angles, a fourth population coding for gaze position was added. This requires the network to compute a further reference frame transformation

3.9 | Adding gaze angle and further generalizations

Similarly, the network presented previously had to learn a simplified version of the actual reference frame transformations that are necessary to link retinotopic visual input and tactile input through proprioception, as we chose to ignore gaze angle. We therefore explored whether the network, in addition to encoding proprioceptive input in joint angles, could handle the additional degree of freedom of gaze direction in the horizontal plane. We trained another network that was identical to the one presented in the previous paragraph, except for a fourth population coding for gaze direction. This population consisted of 120 neurons, representing gaze angles from $-\pi/4$ to $\pi/4$, with 0 indicating looking straight ahead, and a tuning curve width of 1 neuron. The small width of the tuning curve, as explained in the methods, is motivated by the necessity to keep the average firing rate approximately constant across the different populations to allow efficient learning. Visual and proprioceptive inputs were encoded as in the previous section, with the difference that the coordinates of the visual input were now eye-centred. Visual inputs and gaze angles were again uniformly distributed in the respective range, and the body-centred coordinates of the visual stimulus were determined by rotating its eye-centred position by the negative gaze angle around the body axis.

$$\boldsymbol{x}_{body-centred} = R_z(-\theta)\,\boldsymbol{x}_{eye-centred} \tag{10}$$

where θ represents the gaze angle, and R_z the rotation matrix along the vertical axis. As usual, tactile stimulation was present if the body-centred positions of the hand and the visual stimulus differed by less than 15 cm. As summarized in Figure 8c, the network still learned to predict touch in hand-centred coordinates. We also tested the proprioceptive drift induced in the invisible hand setup (Figure 8d). With proprioception fixed, we measured the proprioceptive drift at different locations of visual stimulation in body-centred coordinates, but with a random gaze angle (and therefore different retinotopic coordinates) at each trial. The results were again in line with our main findings, with an attractive pull towards the location of visual stimulation but only in the case of tactile stimulation. However, there was a substantial, constant bias also in the case of no-touch, possibly demonstrating the limits of the network in handling the additional complexity.

4 | DISCUSSION

4.1 | Motivation and approach

The multisensory bases of PPS representation have been studied first in animal neurophysiological studies (see Cléry et al., 2015; Graziano & Cooke, 2006) and later in human neuropsychological, behavioural and neuroimaging studies (see Serino, 2019 for a review). Only more recently, efforts have been made to build neural-network models accounting for the properties of PPS representation in a computational framework (Magosso, Ursino, et al., 2010; Roncone et al., 2016; Straka & Hoffmann, 2017). Shortly after, computational models inspired by visuotactile PPS properties were proposed for impact avoidance (Nguyen et al., 2018), reaching (Juett & Kuipers, 2019) or development of a body schema (Pugach et al., 2019) in robotics. Here we focused on neuroscientific implications of neural network models of PPS representation, by tackling two main questions. First, we asked how the reference frame transformations that are needed to represent visual, proprioceptive and tactile inputs in a common, body-centred reference frame, could be implemented in a conceptually simple and biologically plausible neural network. We proposed that spatially aligned visual and proprioceptive multisensory receptive fields collectively account for the reference frame transformations that allow the maintenance of the overlap between visual and tactile receptive fields, which is at the core of PPS representation. Second, such alignment of reference frames was obtained through the spontaneous tuning of the synaptic connectivity within the neural network as a function of statistical regularities in the environment. Empirical evidence on the high plasticity of PPS representation (Cléry et al., 2015; Maravita & Iriki, 2004; Serino, 2019) suggests that the synaptic changes due to multisensory stimulation during interactions with the environment play a major role in shaping PPS representation. Here, we argue that the same mechanism can be used to explain how PPS representation is formed at a first stage. Therefore, the learning component is fundamental in a neural network model aimed at describing the key proprieties and the emergence of PPS representation. To achieve these goals, we combined findings and methods from two different approaches applied to model multisensory integration and reference frames transformations. We started from the neural network model developed by Magosso and colleagues (Magosso, Ursino, et al., 2010). The model represents PPS representation as the interaction between unisensory areas processing tactile and visual/auditory information and a multisensory layer, integrating the two unisensory inputs in pre-computed spatially overlapping receptive fields. We integrated this approach with further computational models of reference frame transformations, proposed by Ma et al. (2006) and Makin et al. (2013). Ma and colleagues were able to generate coordinate transformations in a neural network model using three interconnected populations of neurons with Gaussian receptive fields, whose synaptic weights were hard-wired. Instead, to model reference frame transformations as learned from sensory inputs, Makin et al. (2013) adapted a neural-network (RBM) that has been widely used to model complex probability distributions

in machine learning. They showed that, indeed, coordinate transformations can be learned from a sensory stimulation based on population coding. Here, we applied the same principles to the key set of sensory inputs that we assumed to be sufficient to build a PPS representation, by implementing an RBM in the architecture proposed by Magosso, Ursino, et al. (2010). In addition to unisensory tactile and visual populations, a proprioceptive population was added allowing the model to processes information related to the position of body parts in space. Importantly, the synaptic connectivity between the unisensory and the multisensory populations was learned through a biologically plausible learning rule, using a set of ecological stimuli as training inputs.

4.2 | Visuo-tactile facilitation in hand-centred reference frames emerges from statistical regularities in the environment

Following classical behavioural and neurophysiological assessments, we focused on visuotactile interactions, and how they are modulated by proprioception, to test PPS representation as emerging from the network. To this aim, visual and proprioceptive inputs in the multisensory layer were encoded in the network, while tactile input were fixed at zero, and the activity induced in the tactile population (through feedback projections) was measured. Such tactile induced activity can be interpreted as the network's prediction of tactile stimulation, based on the integration of visual and proprioceptive information. We found the network's tactile predictions to be based on the hand-centred coordinates of the visual stimulus, with a maximal strength when visual stimuli are close to the hand and an activation profile depending on the distance from the hand, closely resembling what reported from single cell responses by neurophysiological studies in monkeys (as in Graziano et al., 1997). This pattern of response can be linked to the well-known behavioural finding that visual (or auditory) stimuli close to a body part induce a facilitation of tactile processing for the same body part (Canzoneri et al., 2012; Spence et al., 2004). Here, we directly replicated this effect in a behavioural experiment on healthy participants. By suppressing visual information about hand position, which is rarely done in similar behavioural studies, we confirmed the relevance of the proprioceptive-visual associations (as learned by our model) for multisensory integration in the PPS. Our new behavioural data show that tactile responses were facilitated selectively when the side of visual stimulation matched that of the hand position as specified by proprioception. The fact that congruent visual and proprioceptive spatial cues affect multisensory processing is well-known in experimental psychology, typically shown by the crossmodal congruency effect (Pavani et al., 2000; Spence et al., 2000). However, this had never been demonstrated in a tactile detection task, where the presence of visual cues about hand position is typically thought to be the main driving force. Nevertheless, the comparison between model predictions and behavioural data remains qualitative at the present stage, as the main goal of the experiments presented in this paper was to demonstrate the plausibility of the model's architecture. Further efforts should focus on finding better methods to link model predictions to behavioural data, and increasing the granularity of behavioural measures.

Importantly, the fact that the receptive fields are learned and not hard-wired allows us to treat their properties as predictions generated by the model, and not assumptions that are set a priori. Specifically, the model predicts the existence of neurons responding to touch, with overlapping visual and proprioceptive RFs, and neurons not responding to touch with dissociated visual an proprioceptive RFs. The collective behaviour of such neurons leads to the encoding of tactile information being influenced by the hand-centred coordinates of visual stimuli. Their receptive fields are broad and complex in shape, and neurons do not individually encode information in body-part centred coordinates. This is consistent with what was found in literature in multisensory neurons, displaying broad RFs and only partially shifting reference frames (Avillac et al., 2005). Nevertheless, seminal neurophysiological studies, such as by Graziano and colleagues (Graziano et al., 1999) showed how proprioceptive inputs can shift the visual receptive fields of individual neurons. While in our two layers network fully-shifting reference frames can emerge only at the population level (Figure 2f), in further simulations we showed how individual neurons with receptive fields anchored to the hand in space can be spontaneously obtained by letting a third layer learn the associations between tactile inputs and the multisensory representation of sensory inputs. With a three-layers architecture, we therefore showed how neurons with fully and partially shifting RFs may simply be successive levels of information processing. Interestingly, this also implies that canonical PPS neurons may not be needed for generating hand-centred visuotactile interactions. Importantly, we showed that the presence of tactile stimulation that is coherent with visual and proprioceptive inputs can lead to the alignment of visual and proprioceptive receptive fields in multisensory neurons, constituting a possible explanation for both PPS representation and reference frame transformations. Moreover, we have shown how changing the encoding schema of proprioceptive inputs, the unisensory tuning curves, or even adding an additional reference frame transformation does not change such a finding, thus strengthening the idea that learning of statistical regularities is indeed the key mechanism of the network. A notable exception to such generalizations was, however, the challenge encountered when we attempted to extend the network to a 3D spatial representation. This may be due to computational limitations, but further investigations would

be needed to rule out the possibility that this limit may be intrinsic to the network.

4.3 | Visuo-tactile integration explains proprioceptive drift

Similarly to what we did with touch, we then tested the effect of visuotactile stimuli on proprioceptive encoding, by providing visual and proprioceptive inputs, and studying the effect of tactile input on the read-out proprioceptive information. We found that, in the presence of touch, the encoded proprioceptive position got attracted towards the position of the visual stimulus, replicating the proprioceptive drift induced in the IHI. The maximal magnitude of the forecasted shift is around 40% of the visuo-proprioceptive disparity, in line with behavioural data (Guterstam et al., 2013). By adding to the model another unisensory population, encoding the location of the hand in space as specified by visual information only, we also reproduced a proprioceptive drift as during the RHI. The IHI and RHI have been used to experimentally study body ownership, as a key component of bodily self consciousness (Blanke, 2012). It has been suggested that the multisensory stimulation underlying those illusions rely on the same multisensory principles at the bases of PPS representation (Blanke et al., 2015; Grivaz et al., 2017; Makin et al., 2007). Interestingly in this sense, visuotactile stimulation can induce a subset of PPS neurons to anchor their RFs to dummy hands (Graziano, 2000). Here, we show how the same computational mechanisms that generate the reference frame transformations needed to represent the PPS also can explain the proprioceptive drift in the IHI (or RHI). Clearly, we cannot infer subjective states from neural network simulations. However, it is known that multisensory bodily illusions induce a proprioceptive shift consistent with the model's predictions, and, on the subjective side, alter the sense of body ownership. While it has been argued that proprioceptive drift can occur in the absence of (explicitly assessed) body ownership (Rohde et al., 2011), the amount of drift is known to correlate with the perceived strength of the illusion (Guterstam et al., 2013; Tsakiris & Haggard, 2005). In other words, while it is a distinct neural phenomenon, it seems to participate to the phenomenology of ownership, and it is arguably its only known correlate that can be assessed in a neural network model. Here, we have demonstrated how such correlate of body ownership can emerge on the basis of simple multisensory integration in PPS. Previous mathematical studies proposed Bayesian inference on the incoming sensory information as a mechanism to explain illusory ownership in the rubber hand illusion (Samad et al., 2015). The crucial difference and novelty of the present work is that our results were instead obtained in an artificial neural network with a biologically plausible learning rule. Unlike mathematical models, the network is not designed for (and probably does not achieve) optimal Bayesian inference, but it shares the same underlying probabilistic approach to brain function. The network reproduces behavioural findings by learning a generative model of sensory inputs, capturing subtle and highly non-linear relations between patterns of neural activity. For example, the effect of touch on the proprioceptive drift was of the "all or none" kind (Figure 3d–f). Such effect, whose finely tuned non-linearity would be hard to obtain by chance, reflects the fact that, in the training probability distribution, the spatial coherence of visual and proprioceptive inputs only depends on the presence of tactile stimulation, and not on its intensity. Interestingly, the proprioceptive drift decreased when the distance between the hand (defined via proprioception) and the visual stimuli was larger than around 30 cm. This is coherent with the idea that visuotactile interactions occur only within spatially and temporally compatible regions (Holmes & Spence, 2005; Stein et al., 1989), and possibly explains why the RHI and IHI can only take place if the distance between the real and the fake (invisible) hand is limited (Lloyd, 2007). A recent work (Noel, Samad, et al., 2018) found a pattern of spatially decreasing integration of visual and proprioceptive inputs that closely resembles the one found in our simulations. They suggested that the observed behaviour would be in line with a Bayesian causal inference (BCI) model of the world, whose predictions are the weighted average of two alternative sub-models. In one sub-model, the two stimuli are assumed to have the same cause, and their positions are integrated in space, whereas in the alternative sub-model they are treated as separate events. In this perspective, the mathematical counterpart of body ownership would be the weight attributed to the "one-cause" sub-model, as already suggested in (Samad et al., 2015). Recent work by Fang et al. (2019) provided neurophysiological support to this proposal. They trained macaques to perform a reaching task, while recording from their premotor cortex in the presence of different levels of disparity between proprioceptive and visual feedback about hand position. As the level of disparity increased, visuo-proprioceptive integration progressively decreased. In the same study, in a complementary behavioural assessment in humans, the amount of visuo-proprioceptive integration was demonstrated to correlate with subjective ownership ratings, and was therefore taken as an implicit measure of ownership. They showed that the amount of integration, discriminating between "same cause" versus "different cause" responses, that is arm ownership versus no-ownership, could be explained by using a BCI model similar to the one used proposed by Noel, Samad, et al. (2018). Single neurons response also followed two patterns: some neurons tended to integrate visuo-proprioceptive information, suggesting tuning to the "same cause" model, while others tended to segregate them by responding to proprioceptive input only, suggesting tuning to the "separate causes" model. Interestingly, when we included visual information about arm position in the model,

we also found two different patterns of responses from neurons in the multisensory layer: one population of neurons with overlapping and another with dissociated visual (coding hand position) and proprioceptive RFs (Figure 6c). Here, we demonstrated how qualitatively similar results can be obtained in a neural network model that shares with Bayesian models the use of a probabilistic framework to describe brain function, but is not tuned for optimality. Similarly, Ursino and colleagues (Ursino et al., 2017) recently showed that a multisensory effect (i.e., the ventriloquism effect), that has been traditionally explained in the framework of Bayesian inference (Alais & Burr, 2004) can emerge from the organization of multisensory receptive fields.

4.4 | Ownership and embodiment are grounded in a probabilistic model of the physical structure of the body

As we introduced in the previous paragraph, it may be useful to approach the problem beyond the focus of Bayesian optimality, and under a more general perspective. A key function of the brain is to learn the regularities in the probability distribution of its sensory inputs. Those regularities are then exploited to compress inputs in a simpler, more compact representation, retaining the relevant information about their causes in the external world (Attneave, 1954; Barlow, 1961; Simoncelli & Olshausen, 2001). Here, we applied this general principle to a set of sensory inputs – mimicking real-life natural stimulation – that we assumed to be sufficient for building a PPS representation. We fed simple representations of visual, proprioceptive and tactile inputs to a network designed to fit them to a statistical model of their interdependences. The key to the emergence of such statistical model is the network's biologically plausible plasticity rule: by adjusting synaptic weights until its spontaneous activity resembles the training inputs, the network learns the joint probability distribution of multisensory signals. The statistical relations between such training inputs were not arbitrarily chosen, as they are constrained by the physical structure of the body and its interactions with the environment: touch is always on the body, thus environmental stimuli associated to touch must occur close to the physical body, and their proximity is encoded based on visual and proprioceptive cues. We then showed how, under such limited hypotheses, both PPS representation and the IHI (or the RHI) spontaneously emerge as consequence of a single and unified inference process, where sensory inputs are treated differently depending on their relation to the body. This means that our network complies to some extent with F. de Vignemont's minimal definition of embodiment, arguably the only one that can be applied to a neural network simulation: "E is embodied if and only if some properties of E are processed in the same way as the properties of one's body" (de Vignemont, 2011).

There are other important features that were not directly modelled here, but could be implemented in a model with an architecture similar to ours, designed to learn the probability distribution of its sensory inputs, in order to extend its level of compliance with such definition. For instance, the present model allows to accurately model only simultaneous stimuli. However, the combination between temporal and spatial processing is key for a dynamical model of the body in space, which is deeply linked to PPS representation, as well as for bodily self-consciousness. Moreover, a perfect generative model, which does not include temporal features, should in principle not fully account for a PPS representation extending beyond the skin, as it would only learn associations between touch and stimuli currently causing it. Again, this could be instead achieved in the general framework of the learning of a complex probability distribution, extending not only in space, but also in time. Interestingly, the idea of PPS representation as a spatio-temporal prediction system finds empirical support in the observation that it expands when faster stimuli approach (Noel, Blanke, et al., 2018). Straka and Hoffmann (2017) investigated the dynamical properties of visuotactile integration in PPS by coupling an RBM with a feedforward layer undergoing supervised learning. Alternatively, implementing a recurrent dynamics in our RBM would allow to handle the temporal dynamics with more biological realism (see for example Makin et al., 2015). Similarly, the complexity of the visual input could be increased to replace the population coding of pre-computed positions of objects in space, with more realistic inputs, starting from retinotopic representations to egocentric representations. This way, the visual appearance of the body would be embedded in the training inputs' distribution, possibly allowing explaining why multisensory bodily illusions work less well (or do not work at all) with objects that do not resemble body parts (Tsakiris et al., 2010). Again, the network's conceptual functioning would still hold on the learning of a joint probability distribution, whose variables would be the neural activities of a retinotopic intensity coding. In principle, this framework could be extended to build a neural network that learns a model of all the possible interactions between the body and the environment. We argue that such a process, of which we successfully modelled few key aspects here, might constitute the neurocomputational basis of body representation, and a substrate for the subjective experience of possessing a body, that is felt as one's own, in interaction with the external world.

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.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

TB, EM and AS developed the conceptual framework. TB designed and implemented the neural network, and performed the simulations, behavioural data collection and data analysis. EM and AS provided supervision in data analysis and interpretation of the results. TB drafted the manuscript, all the authors edited and contributed to the critical revisions of the manuscript and read and approved the final version for submission.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.14981

DATA AVAILABILITY STATEMENT

MATLAB code for training the main network presented in the paper and to reproduce Figures 2 and 3 are available in the following OSF repository: https://osf.io/w6edv/?view_only=135df79adbd84ce79d01111a1663b7de

Further MATLAB code, behavioural data and detailed instructions for the replication of other results are freely available from the first author (TB) upon request.

ORCID

Tommaso Bertoni https://orcid.org/0000-0002-4798-7733
Elisa Magosso https://orcid.org/0000-0002-4673-2974
Andrea Serino https://orcid.org/0000-0001-7475-6095

REFERENCES

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183–193. http://dx.doi.org/10.1037/h0054663

Avillac, M., Denève, S., Olivier, E., Pouget, A., & Duhamel, J.-R. (2005). Reference frames for representing visual and tactile locations in parietal cortex. *Nature Neuroscience*, *8*, 941–949.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication*, 217–234.

Bernasconi, F., Noel, J., Park, H. D., Faivre, N., Seeck, M., Spinelli, L., Schaller, K., Blanke, O., & Serino, A. (2018). Audio-tactile and peripersonal space processing around the trunk in human parietal and temporal cortex: An intracranial EEG study. *Cerebral Cortex*, *28*, 3385–3397. https://doi.org/10.1093/cercor/bhy156

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience*, *13*, 556–571.

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron*, *88*, 145–166. https://doi.org/10.1016/j.neuron.2015.09.029

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*, 756. https://doi.org/10.1038/35784

Brain, W. R. (1941). Visual disorientation with special reference to lesions of the right cerebral hemisphere. *Brain*, *64*(4), 244–272. https://doi.org/10.1093/brain/64.4.244

Brozzoli, C., Gentile, G., Petkova, V. I., & Ehrsson, H. H. (2011). FMRI adaptation reveals a cortical mechanism for the coding of space near the hand. *Journal of Neuroscience*, *31*, 9023–9031.

Bufacchi, R. J., & Iannetti, G. D. (2018). An action field theory of peripersonal space. *Trends in Cognitive Sciences*, *22*, 1076–1090.

Canzoneri, E., Magosso, E., & Serino, A. (2012). Dynamic sounds capture the boundaries of peripersonal space representation in humans. *PLoS One*, *7*, 3–10. https://doi.org/10.1371/journal.pone.0044306

Canzoneri, E., Ubaldi, S., Rastelli, V., Finisguerra, A., Bassolino, M., & Serino, A. (2013). Tool-use reshapes the boundaries of body and peripersonal space representations. *Experimental Brain Research*, *228*, 25–42.

Cléry, J., Guipponi, O., Wardak, C., & Ben Hamed, S. (2015). Neuronal bases of peripersonal and extrapersonal spaces, their plasticity and their dynamics: Knowns and unknowns. *Neuropsychologia*, *70*, 313–326. https://doi.org/10.1016/j.neuropsychologia.2014.10.022

de Vignemont, F. (2011). Embodiment, ownership and disownership. *Consciousness and Cognition*, *20*(1), 82–93. http://dx.doi.org/10.1016/j.concog.2010.09.004

di Pellegrino, G., & Làdavas, E. (2015). Peripersonal space in the brain. *Neuropsychologia*, *66*, 126–133. https://doi.org/10.1016/j.neuropsychologia.2014.11.011

di Pellegrino, G., Làdavas, E., & Farné, A. (1997). Seeing where your hands are. *Nature*, *388*, 730. https://doi.org/10.1038/41921

Duhamel, J.-R., Colby, C. L., & Goldberg, M. E. (1998). Ventral intraparietal area of the macaque: Congruent visual and somatic response properties. *Journal of Neurophysiology*, *79*, 126–136. https://doi.org/10.1152/jn.1998.79.1.126

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433. https://doi.org/10.1038/415429a

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.

Farnè, A., & Làdavas, E. (2002). Auditory peripersonal space in humans. *Journal of Cognitive Neuroscience*, *14*, 1030–1043.

Fogassi, L., Gallese, V., Fadiga, L., Luppino, G., Matelli, M., & Rizzolatti, G. (1996). Coding of peripersonal space in inferior premotor cortex (area F4). *Journal of Neurophysiology*, *76*, 141–157.

Graziano, M. S. A. (1999). Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 10418–10421.

Graziano, M. S. A. (1999). Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position. *Proceedings of the National Academy of Sciences*, *96*(18), 10418–10421. http://dx.doi.org/10.1073/pnas.96.18.10418

Graziano, M. S. A. (2000). Coding the location of the arm by sight. *Science*, *290*, 1782–1786. https://doi.org/10.1126/science.290.5497.1782

Graziano, M. S. A., & Cooke, D. F. (2006). Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia*, *44*(6), 845–859. http://dx.doi.org/10.1016/j.neuropsychologia.2005.09.009

Graziano, M. S. A., Hu, X. T., & Gross, C. G. (1997). Visuospatial properties of ventral premotor cortex. *Journal of Neurophysiology*, *77*, 2268–2292.

Graziano, M. S. A., Yap, G. S., & Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science*, *266*, 1054–1057. https://doi.org/10.1126/science.7973661

Grivaz, P., Blanke, O., & Serino, A. (2017). Common and distinct brain regions processing multisensory bodily signals for peripersonal space and body ownership. *NeuroImage*, *147*, 602–618. https://doi.org/10.1016/j.neuroimage.2016.12.052

Guterstam, A., Gentile, G., & Ehrsson, H. H. (2013). The invisible hand illusion: Multisensory integration leads to the embodiment of a discrete volume of empty space. *Journal of Cognitive Neuroscience*, *25*, 1078–1099. https://doi.org/10.1162/jocn_a_00393

Hinton, G. E. (2000). Training products of experts by minimizing contrastive divergence. *Neural Computation*, *14*(8), 1771–1800. http://dx.doi.org/10.1162/089976602760128018

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*, 504–507. https://doi.org/10.1126/science.1127647

Holmes, N. P., & Spence, C. (2005). Multisensory integration: Space, time and superadditivity. *Current Biology*, *15*, R762–R764. https://doi.org/10.1016/j.cub.2005.08.058

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *NeuroReport*, *7*, 2325–2330. https://doi.org/10.1097/00001756-199610020-00010

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.

Juett, J., & Kuipers, B. (2019). Learning and acting in peripersonal space: Moving, reaching, and grasping. *Frontiers in Neurorobotics*, *13*, 1–20.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*, 764–766.

Lloyd, D. M. (2007). Spatial limits on referred touch to an alien limb may reflect boundaries of visuo-tactile peripersonal space surrounding the hand. *Brain and Cognition*. *64*, 104–109. https://doi.org/10.1016/j.bandc.2006.09.013

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432–1438.

Magosso, E., Ursino, M., di Pellegrino, G., Làdavas, E., & Serino, A. (2010). Neural bases of peri-hand space plasticity through tool-use: Insights from a combined computational–experimental approach. *Neuropsychologia*, *48*, 812–830. https://doi.org/10.1016/j.neuropsychologia.2009.09.037

Magosso, E., Zavaglia, M., Serino, A., di Pellegrino, G., & Ursino, M. (2010). Visuotactile representation of peripersonal space: A neural network study. *Neural Computation*, *22*, 190–243.

Makin, J. G., Dichter, B. K., & Sabes, P. N. (2015). Learning to estimate dynamical state with probabilistic population codes. *PLOS Computational Biology*, *11*(11), e1004554. http://dx.doi.org/10.1371/journal.pcbi.1004554

Makin, J. G., Fellows, M. R., & Sabes, P. N. (2013). Learning multisensory integration and coordinate transformation via density estimation. *PLoS Computational Biology*, *9*, e1003035. https://doi.org/10.1371/journal.pcbi.1003035

Makin, T. R., Holmes, N. P., & Zohary, E. (2007). Is that near my hand? Multisensory representation of peripersonal space in human intraparietal sulcus. *Journal of Neuroscience*, *27*, 731–740.

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, *8*, 79–86.

Nguyen, D. H. P., Hoffmann, M., Roncone, A., Pattacini, U., & Metta, G. (2018). Compact real-time avoidance on a humanoid robot for human-robot interaction. *ACM/IEEE Int. Conf. Human-robot Interact.*, 416–424.

Noel, J.-P., Blanke, O., Magosso, E., & Serino, A. (2018). Neural adaptation accounts for the dynamic resizing of peripersonal space: Evidence from a psychophysical-computational approach. *Journal of Neurophysiology*, *119*, 2307–2333. https://doi.org/10.1152/jn.00652.2017

Noel, J., Samad, M., Doxon, A., Clark, J., Keller, S., & Di Luca, M. (2018). Peri-personal space as a prior in coupling visual and proprioceptive signals. *Scientific Reports*, *8*, 15819. https://doi.org/10.1038/s41598-018-33961-3

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

Pouget, A., Deneve, S., & Duhamel, J.-R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, *3*, 741–747.

Pugach, G., Pitti, A., Tolochko, O., & Gaussier, P. (2019). Brain-inspired coding of robot body schema through visuo-motor integration of touched events. *Frontiers in Neurorobotics*, *13*, 5.

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*, e25214. https://doi.org/10.1371/journal.pone.0025214

Rizzolatti, G., Matelli, M., & Pavesi, G. (1983). Deficits in attention and movement following the removal of postarcuate (area 6) and prearcuate (area 8) cortex in macaque monkeys. *Brain*, *106*(3), 655–673. https://doi.org/10.1093/brain/106.3.655

Rizzolatti, G., Scandolara, C., Matelli, M., & Gentilucci, M. (1981). Afferent properties of periarcuate neurons in macaque monkeys. II. Visual responses. *Behavioural Brain Research*, *2*, 147–163.

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*, e21659. https://doi.org/10.1371/journal.pone.0021659

Roncone, A., Hoffmann, M., Pattacini, U., Fadiga, L., & Metta, G. (2016). Peripersonal space and margin of safety around the body: Learning visuo-tactile associations in a humanoid robot with artificial skin. *PLoS One*, *11*, 1–32. https://doi.org/10.1371/journal.pone.0163713

Salomon, R., Noel, J. P., Łukowska, M., Faivre, N., Metzinger, T., Serino, A., & Blanke, O. (2017). Unconscious integration of multisensory bodily inputs in the peripersonal space shapes bodily self-consciousness. *Cognition*, *166*, 174–183. https://doi.org/10.1016/j.cognition.2017.05.028

Samad, M., Chung, A. J., & Shams, L. (2015). Perception of body ownership is driven by Bayesian sensory inference. *PLoS One*, *10*, 1–23. https://doi.org/10.1371/journal.pone.0117178

Serino, A. (2019). Peripersonal space (PPS) as a multisensory interface between the individual and the environment, defining the space of the self. *Neuroscience and Biobehavioral Reviews*, *99*, 138–159.

Serino, A., Canzoneri, E., Marzolla, M., di Pellegrino, G., & Magosso, E. (2015). Extending peripersonal space representation without tool-use: Evidence from a combined behavioral-computational approach. *Frontiers in Behavioural Neurosciences*, *9*, 4.

Serino, A., Noel, J.-P., Galli, G., Canzoneri, E., Marmaroli, P., Lissek, H., & Blanke, O. (2015). Body part-centered and full body-centred peripersonal space representations. *Scientific Reports*, *5*, 18603.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*(1), 1193–1216. http://dx.doi.org/10.1146/annurev.neuro.24.1.1193

Spence, C., Pavani, F., & Driver, J. (2000). Crossmodal links between vision and touch in covert endogeneous spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1298–1319.

Spence, C., Pavani, F., Maravita, A., & Holmes, N. (2004). Multisensory contributions to the 3-D representation of visuotactile peripersonal space in humans: Evidence from the crossmodal congruency task. *Journal of Physiology - Paris*, *98*, 171–189. https://doi.org/10.1016/j.jphysparis.2004.03.008

Stein, B. E., Meredith, M. A., Huneycutt, W. S., & McDade, L. (1989). Behavioral indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. *Journal of Cognitive Neuroscience*, *1*, 12–24. https://doi.org/10.1162/jocn.1989.1.1.12

Straka, Z., & Hoffmann, M. (2017). Learning a peripersonal space representation as a visuo-tactile prediction task. ICANN 2017: 26th International Conference on Artificial Neural Networks, Alghero, Italy, September 11-14, 2017, Proceedings, Part I, str. 101–109, Cham. Springer International Publishing.

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*(3), 343–352. http://dx.doi.org/10.1007/s00221-009-2039-3

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. https://doi.org/10.1037/0096-1523.31.1.80

Ursino, M., Cuppini, C., & Magosso, E. (2017). Multisensory bayesian inference depends on synapse maturation during training: Theoretical analysis and neural modeling implementation. *Neural Computation*, *29*(3), 735–782. http://dx.doi.org/10.1162/neco_a_00935

Van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1998). The precision of proprioceptive position sense. *Experimental Brain Research*, *122*, 367–377.

Welling, M., Rosen-Zvi, M., & Hinton, G. (2004). Exponential family harmoniums with an application to information retrieval. *NIPS*, *17*, 1481–1488.

Zampini, M., Torresan, D., Spence, C., & Murray, M. M. (2007). Auditory-somatosensory multisensory interactions in front and rear space. *Neuropsychologia*, *45*, 1869–1877. https://doi.org/10.1016/j.neuropsychologia.2006.12.004

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Bertoni T, Magosso E, Serino A. From statistical regularities in multisensory inputs to peripersonal space representation and body ownership: Insights from a neural network model. *Eur. J. Neurosci.* 2021;53:611–636. https://doi.org/10.1111/ejn.14981

Supporting information

From statistical regularities in multisensory inputs to peripersonal space representation and body ownership: insights from a neural network model

Detailed calculation of unisensory precisions

Here we provide the detailed calculation of the localization precision of unisensory inputs, σ_x, assuming a flat prior on their position. When a stimulus located in **x** is encoded in a unisensory population, it generates a conditional distribution of neural activity in the respective unisensory population **u**. The activity of each unisensory neuron u_i is drawn from an independent Poisson distribution whose mean is determined by the stimulus location and the neuron's tuning curve. In our case, we have:

$$P(u|x) = \prod_i P(u_i|x)$$

$$P(u_i|x) = Pois(g \cdot exp\left\{\frac{-||\boldsymbol{x} - \widehat{\boldsymbol{x}}_i|^2}{2\sigma_{TC}^2}\right\})$$

Where $\widehat{\boldsymbol{x}}_i$ denotes the i-th neuron's preferred position, g is the gain of the stimulus, and σ_{TC} is the standard deviation of the tuning curve. Note that here we assume all the tuning curves of neurons within each unisensory population to be identical, except for the preferred position. By combining the two expressions we get:

$$P(u|x) \propto \prod_i Pois(g \cdot exp\left\{\frac{-||x - \widehat{\boldsymbol{x}}_i|^2}{2\sigma_{TC}^2}\right\}) =$$

$$= \prod_i \frac{g}{u_i!} exp\left\{-g e^{\frac{-||x-\widehat{\boldsymbol{x}}_i|^2}{2\sigma_{TC}^2}}\right\} exp\left\{-u_i \frac{||x - \widehat{\boldsymbol{x}}_i|^2}{2\sigma_{TC}^2}\right\}$$

$$= (\prod_i \frac{g}{u_i!}) exp\left\{-g \sum_i e^{\frac{-||x-\widehat{\boldsymbol{x}}_i|^2}{2\sigma_{TC}^2}}\right\} exp\left\{-\sum_i u_i \frac{||x - \widehat{\boldsymbol{x}}_i|^2}{2\sigma_{TC}^2}\right\}$$

$$\approx (\prod_i \frac{g}{u_i!}) exp\{g\sqrt{2\pi\sigma_{TC}}\} exp\left\{-\sum_i u_i \frac{||x - \widehat{\boldsymbol{x}}_i|^2}{2\sigma_{TC}^2}\right\}$$

Where the approximation consists in assuming that the sum in the first exponential consists of enough terms to depend weakly on **x**. This is true if the neurons are tiled densely enough that a

large number of them contribute to the sum, and the approximated value of the sum can be computed by an integral. As long as this value is constant, it is not needed to compute the posterior variance, which can be obtained by re-writing the exponent of the second term of the expression as follows:

$$-\sum_i u_i \frac{||x-\hat{\boldsymbol{x}}_i|^2}{2\sigma_{TC}^2}$$

$$= \frac{\sum_i u_i}{2\sigma_{TC}^2}\left[x^2 - 2x\frac{\sum_i \hat{\boldsymbol{x}}_i u_i}{\sum_i u_i} + \frac{\sum_i \hat{\boldsymbol{x}}_i 2u_i}{\sum_i u_i}\right]$$

$$= \frac{\sum_i u_i}{2\sigma_{TC}^2}\left[x - \frac{\sum_i \hat{\boldsymbol{x}}_i u_i}{\sum_i u_i}\right]^2 + C$$

$$= \frac{\sum_i u_i}{2\sigma_{TC}^2}(x - x_b)^2 + C$$

Where C does not depend on the stimulus location **x**, and $x_b = \frac{\sum \hat{x}_i u_i}{\sum u_i}$ is the barycenter of neural activity. The posterior is therefore Gaussian, with mean x_b and standard deviation $\sigma_x = \sqrt{\frac{\sigma_{TC}^2}{\sum u_i}}$. The relevant quantity for estimating σ_x becomes then the total spike count of each sensory input. If the number of active neurons is large enough, the expected value for this quantity can be approximated by an integral

$$E[\sum_i u_i] = g\sum_i e^{-\frac{\hat{x}_i^2}{2\sigma_{TC}^2}} \approx g\int e^{-\frac{\hat{x}^2}{2\sigma_{TC}^2}} d\hat{\boldsymbol{x}} = g(2\pi\sigma_{TC}^2)^{n/2}$$

Where n is the dimensionality of the physical space of stimulus position (2 in our case), and for simplicity we have considered a stimulus centred in 0, and performed the calculation in units equal to the neuron grid spacing. Therefore, in such units, we have

$$\sigma_x = \sqrt{\frac{\sigma_{TC}^2}{g2\pi\sigma_{TC}^2}} = \sqrt{\frac{1}{g2\pi}}$$

Note that σ_x, in general, depends on σ_{TC}^2, but not in a 2D grid of neurons. Therefore, in our case, the only way to adjust the stimulus precision is by changing the gain or the density of neurons.

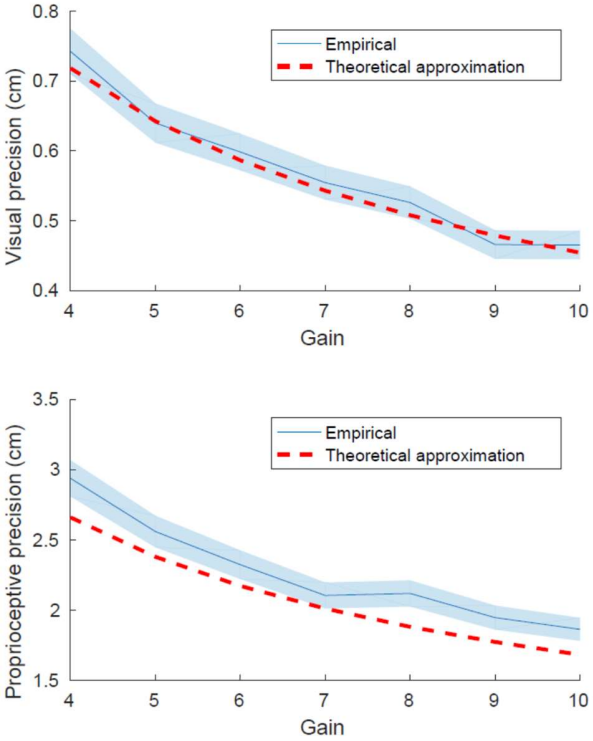


Figure S1: To evaluate the goodness of the approximation, we performed simulations by generating 1000 sets of unisensory stimuli in a fixed position, for different values of the gain. Then, we decoded the maximum-likelihood position of the stimulus as the barycentre of the neural population, and estimated its standard deviation along the x axis. The values were compared to the results obtained from eq. (3), after conversion from neuron grid units to physical space units (see Fig. S1). Overall, the approximation was good, even in the more extreme case of the proprioceptive population, where the relatively small number of neurons could have challenged the assumptions of the approximation.

Effect of the width of unisensory tuning curves on the results

While the multisensory receptive fields in our network were entirely learned from sensory stimulation, unisensory receptive fields were set a priori, and despite being based on neurophysiological knowledge they present a certain degree of arbitrariness. Namely, the width of the Gaussian tuning curves has been determined mainly on technical grounds, to allow efficient training of an RBM. One may therefore wonder to which extent our results depend on the choice of the unisensory tuning curves. Namely, the spatial extent of the hand-centred region in which visual stimuli elicit tactile predictions (the size of the in-silico PPS) may depend on the width of unisensory receptive fields. We therefore trained a series of replicas of our main network, in which we only changed the width of the tuning curves of the unisensory visual and proprioceptive population, and plotted the evoked tactile activity as a function of the distance from the hand. The range of explored widths has a lower limit in that it cannot get much smaller

than one neuron in the proprioceptive population, because the stimulus encoding would become extremely irregular, and it cannot get too big as this would require huge safety margins to avoid edge effects. Within this reasonable range, there was virtually no sensitivity to the width of the tuning curve (see Figure S2). Again, this is in line with the idea that the encoding schema should not matter too much, as long as the network is able to learn a good generative model of its inputs.

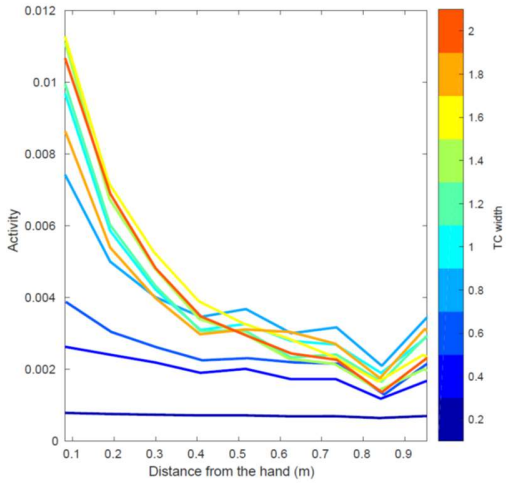


Figure S2: PPS spatial properties as a function of the width of unisensory tuning curves. We trained 10 replicas of our main network, and multiplied the width of visual and proprioceptive tuning curves by a fixed factor ranging from 0.2 to 2. Here we show the dependence of tactile evoked activity on the distance from the hand of the visual stimulus, as a function of the tuning curve width. The multiplicative factor is colour coded as depicted in the colorbar, with the lowest value (0.2) corresponding to dark blue and the highest value (2) corresponding to red.

Optimal number of hidden units and precision

Here we illustrate how the number of hidden units influences the precision with which unisensory positions are encoded in the multisensory layer. This analysis was used to determine the number of hidden units to use in our network, aiming at reaching a sufficiently low information loss when passing from the lower to the upper layer, while respecting a biological principle of efficient encoding and keeping computational demands not too high. In order to do so systematically, we trained 20 other replicas of our network, and systematically changed only the number of hidden units from 10 to 3000. We then used the precision with which the position of unisensory stimuli can be recovered, after encoding them in the multisensory layer as a main proxy of information loss. Such precision has a lower bound in the theoretical precision illustrated in section 1, due to noise in unisensory inputs, so when such bound is reached no information loss takes place in the encoding. Practically, this was assessed by generating random positions for visual and proprioceptive stimuli, encoding them (with noise) in the unisensory layer. Then, unisensory activities were projected in the multisensory layer and read out again from the unisensory populations through the usual procedure. However, since

we are interested in the information loss in an “up” pass, the read out is done noiselessly, by taking mean values instead of Poisson samples. For our main analysis, we considered results obtained by performing noisy “up” passes, as the efficient encoding principle needs to take noise into account. After a sharp decrease in the encoding error until 800 hidden units (see Figure S3), the performance starts saturating, especially for visual inputs. We therefore determined that 1500 hidden units would be a good trade-off between complexity and performance. Additionally, we performed the same analysis in the case of noiseless “up” passes, to see how quickly the network approaches the theoretical limit (that can only be achieved in the case of noiseless “up” passes) when it is not limited by noise.

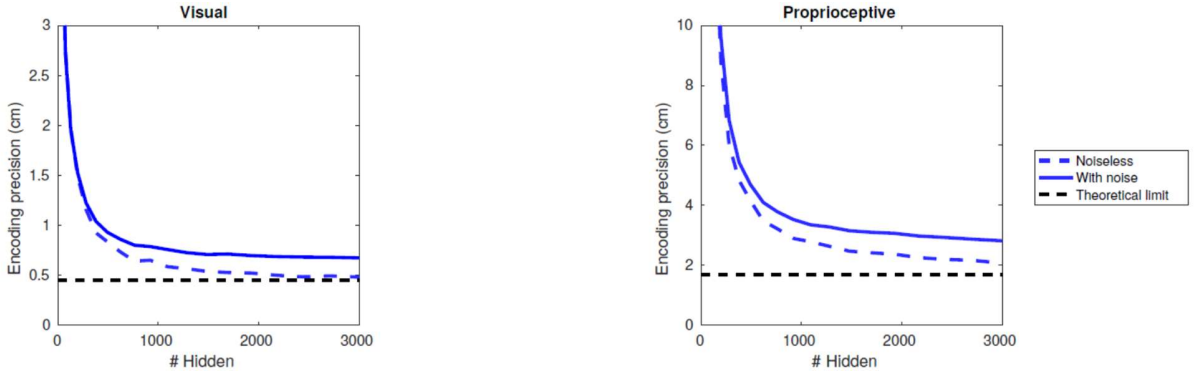


Figure S3: Precision in the encoding of visual (left) and proprioceptive (right) inputs, as a function of the number of hidden units in the network. The encoding precision is defined as the standard deviation (average between x and y directions) of the positions obtained after generating unisensory inputs, encoding them in the multisensory layer and then decoding them again by projecting multisensory activity down to the unisensory layer and taking the barycentre of the generated activity. The ``down'' pass is always noiseless as it only acts as a decoding step, while we show results for both a noisy (solid blue line, used for determining the number of hidden units) and noiseless (dashed blue line) ``up'' pass. The maximal theoretical precision as obtained in Section 1 is shown as the black dashed line.

Additional behavioural analyses

As mentioned in the main text, here we analyse the effect of the temporal delay of stimulation in more detail. First of all, we performed a Delay*Position*Congruency 3x2x3 ANOVA. Since there was no significant three-way interaction ($p = .72$), we pooled the two hand positions together, and performed a two-way Delay*Congruency ANOVA. We observed significant main effects of Congruency, as already confirmed by linear mixed models in the main text ($F(2, 84) = 4.04$, $p = .0209$). Moreover, we observed a significant main effect of Delay ($F(2, 84) = 17.36$, $p < .001$), possibly indicating overall expectation effects. Interestingly, the Delay*Congruency interaction was also significant ($F(4, 168) = 3.77$, $p = .0057$), with an overall stronger effect of temporal delay (or distance) in the congruent condition (see Figure S4).

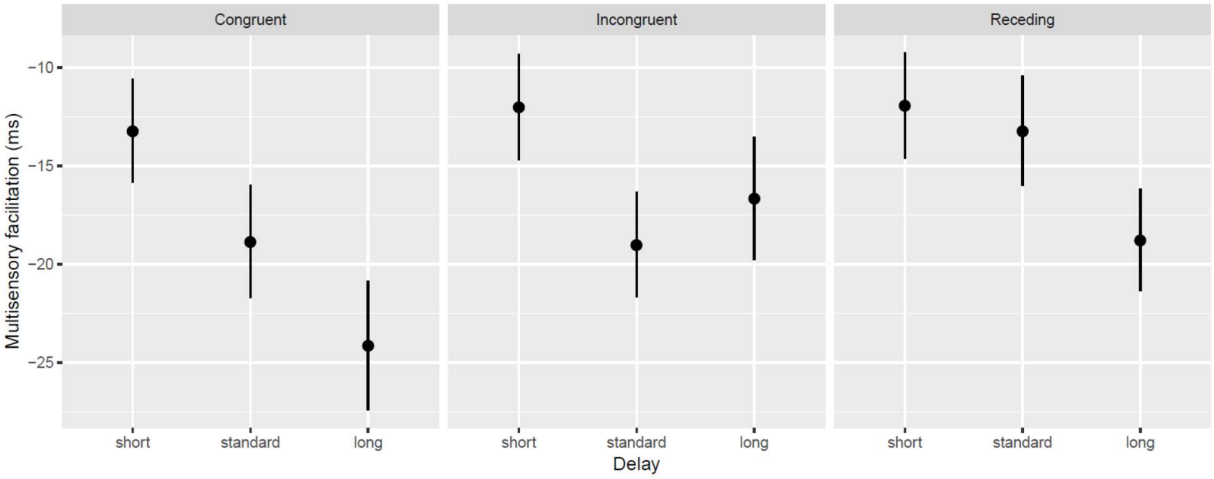


Figure S4: Multisensory facilitation plotted by congruency and by delay. In the congruent condition, the 'short' delay corresponds to approximately 10 cm from the hand, while the 'long' delay corresponds to 0 cm from the hand. Errorbars represent standard errors.

The self and the Bayesian brain: testing probabilistic models of body ownership

Tommaso Bertoni[1*] & Giulio Mastria[1*], Henri Perrin[2], Boris Zbinden[1], Michela Bassolino[3], Andrea Serino[1]

*Authors equally contributed

[1] MySpace Lab, Department of Clinical Neurosciences, University Hospital of Lausanne, University of Lausanne, Lausanne, Switzerland

[2] School of Medicine, Faculty of Biology and Medicine, Unversity of Lausanne, Lausanne, Switzerland

[3] School of Health Sciences, HES-SO Valais-Wallis, Sion, Switzerland

Abstract

Simple multisensory manipulations can induce the illusory misattribution of external objects to one's own body. Thus, the process leading to body ownership has been compared to optimal Bayesian inference: the online estimation of the probability that one object belongs to the body from the congruence of multisensory inputs. This idea was highly influential, providing the basis for bottom-up accounts of self-consciousness, but empirical evidence in its support is scarce. Here we will test a Bayesian model of hand ownership based on visuo-proprioceptive congruency. Model predictions will be compared to data from a virtual-reality reaching task, whereby reaching errors induced by a spatio-temporally mismatching virtual hand will be used as implicit proxy of hand ownership. We will independently assess unisensory components to rigorously test optimality. This is crucial to provide conclusive evidence on whether key components of self-consciousness can be truly described as the bottom-up, behaviorally optimal processing of multisensory inputs.

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 this experiences has been termed bodily self-consciousness (BSC) and is considered the minimal building block of consciousness and self-awareness [1].

BSC and the identification with our own body are so rooted in our everyday experience that they are easily given for granted; 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 for an external object [2] or dis-embodiment for an actual body part [3,4]. 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 [5]. Probably on the wake of the success of Bayesian accounts of low-level multisensory integration [6], this theoretical principle has been soon qualitatively rephrased in terms of Bayesian inference on sensory information. Ramachandran [5] was the first to propose 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 is part of one's own body as the most probable. What was striking in this account of the RHI is that the bottom-up Bayesian logic seemed to overcome top-down cognitive constraints, such as those originating from years worth of experience

about the visual appearance of one's own body. In the following decade, this observation sparked a lively debate opposing "bottom-up" Bayesian approaches focusing on visuo-tactile congruencies [5,7,8], and "top-down" cognitive constraints focusing on posture and visual appearance [9–11]. Later on, "top-down" constraints 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 [12–14].

Bayesian accounts have been extended to explain BSC as a whole as the result of the comparison of sensory inputs and internal states to a "self-model" [12]. Over the years, such inference-based accounts have also included the integration of mental states [15] and interoceptive signals [16–19], linking inference on internal bodily and neural states to self-consciousness itself. Arguably, the key to the success of Bayesian approximations to brain function is that they constitute a normative framework, as they find a clear evolutionary motivation in the need to behave optimally in a noisy sensory environment. In its initial field of application, i.e. low-level multisensory integration, the validity of such approach has been rigorously proven experimentally [6,20–22]. Instead, although Bayesian descriptions of BSC have been popular for almost two decades, most accounts are purely conceptual [12,15–17,23,24] or mathematical [13]. Experimental studies in their support are still rather scarce [25,26] and do not provide conclusive proofs of the optimality of behaviour, weakening the motivation for the use of a normative model.

Here, we aim at extending the evidence base for Bayesian theories of BSC by focusing on body ownership. Body ownership is a key, quantifiable component of BSC that can be experimentally manipulated, and its qualitative underlying principles are relatively well understood. Therefore, quantitative theories of ownership constitute the ideal connection point to generalize models of multisensory integration to higher levels of conscious experience. To test the validity of such

generalization, we propose a Bayesian model of hand ownership and rigorously test it through a set of virtual reality-based tasks.

In classical models of multisensory integration [6] 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 [27]. It has been suggested that this problem may also be solved in a probabilistic framework, i.e. Bayesian Causal Inference (Bayesian CI), in which the brain infers the likelihood that two unisensory stimuli originate from the same cause, based on their spatial and temporal congruencies [28]. 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.) in different experimental setups. For example, Samad and colleagues 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. 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. This qualitative observation is the only empirical evidence supporting the validity of the model.

More recently, Fang and colleagues [25] modelled ownership of a virtual hand as a function of visuo-proprioceptive disparity during a reaching task. In their experiment, by adapting classic paradigms of visuo-motor rotation [29], 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 increased as a function of the virtual hand's displacement within a given range of visuo-proprioceptive disparity, but decreased for large levels of disparity. Such behaviour was well modelled by the predictions of a Bayesian CI model where the probability of visual information from a virtual hand and somatosensory information from the real hand originating from the same physical cause (P_{com}) decreases proportionally to the visuo-proprioceptive disparity, because of the lower weight attributed to vision at large levels of disparity. Explicit ownership ratings in humans covaried with P_{com}, suggesting that this parameter might provide an implicit measure of ownership probability at a trial-by-trial level.

As already anticipated, Bayesian models assume optimality as a normative constraint to brain functions. Tipically, optimality is defined as the behaviour minimizing squared errors on estimates (in this case: position estimates) depending on a set of (unknown) free parameters (in this case: unisensory precisions). In order to rigorously assess the optimality of the observed behaviour, it is necessary either to manipulate the free parameters in a controlled manner or directly measure them [30]. However, previous studies on body ownership directly fitted unisensory precisions from multisensory tasks [25], or assumed fixed and arbitrary parameters [26], making the presence of optimality hard to determine with certainty. Interestingly, Costantini and colleagues [31] found that the strength of the RHI is constrained by the precision in perceiving the synchrony of visuo-tactile

stimuli at the single subject level. This is consistent with the predictions of integration based on Bayesian CI, even if the authors have not discussed such hypothesis. Conversely, a recent study found that susceptibility to the RHI did not depend on proprioceptive precision [32]. Throughout the literature, Bayesian approximations have been applied to a variety of other brain functions [15], spanning from pain to social and narrative aspects of the self, although optimality is rarely assessed experimentally. In most cases these models are complex and flexible enough to fit well the experimental data, however directly testing the key optimality assumption is crucial to provide a normative justification to their proposed mechanisms.

In the present work, we aim at providing empirical evidence for the hypothesis that body ownership emerges from a Bayesian inference process, as a key instance in the family of models that extend Bayesian approaches from simple multisensory perception to self-consciousness. To do this, we will extend and revise the existing models and behavioural validations, focusing on spatial and temporal features of visual and proprioceptive inputs. We will introduce a virtual reality adaptation of the reaching task used by Fang and colleagues – i.e., visuo-proprioceptive disparity task (VPD, Figure 1 a) - as a base for the assessment of the Bayesian CI model predictions. In addition to the spatial manipulation of visuo-proprioceptive disparity, we will also modulate 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 will also strengthen the interpretation of the behaviour observed in the VPD task and the underlying model. Ownership (or disownership) is 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. Threfore, the first main hypothesis of this study is that a model taking into account both the spatial and temporal manipulation will outperform both the forced fusion model (replicating Fang 2019) and a model including only spatial disparity, demonstrating the combined effect of spatial and temporal congruencies in eliciting body ownership.

With respect to direct assessments of ownership through multisensory illusions such as the RHI, our task has two main advantages: first, the short duration of each trial and the quantitative nature of the variables at play, which allow collecting data with sufficient granularity for modelling; second, the measure of ownership provided by the reaching errors is implicit and not based on subjective ratings, whose validity has been recently put into question. Lush and colleagues [33] found that explicit reports of ownership during the RHI could be biased by subjects' suggestibility to illusion. In this view, the RHI would be the consequence of a form of phenomenological control more than of a genuine subjective experience of ownership. Here, subjective ratings will only be collected in a second moment and put in relation to reaching bias in order to ascertain its link with subjective ownership. Therefore, our study's second hypothesis is that explicit evaluation of ownership feeling during the VPD task will match the probability of multisensory integration (P_{com}) estimated by the Bayesian CI model both in the spatial and temporal domain.

Our model will provide predictions of the reaching bias as a function of both spatial and temporal disparity, depending on four free parameters that will be fitted from the data: σ_v, the unisensory visual precision, σ_p, the unisensory proprioceptive-motor precision, the temporal precision σ_t and a global prior about ownership of the virtual hand P_π. The parameter σ_p incorporates the accuracy of both the kinaesthetic and movement execution that determine the final precision in reaching a target, in order to disambiguate it from static hand position sense (σ_{ps}) that will be introduced later

on. The free parameters will be fitted from experimental data at the individual level by finding the parameter set that maximizes the match between model predictions and reaching bias.

As mentioned previously, to strengthen model validation by directly assessing optimality, it is necessary either to manipulate or to independently measure unisensory precisions. Here we chose the latter approach due to the impossibility of manipulating with the same level of accuracy unisensory noise for both vision and proprioception in healthy humans during our multisensory task (VPD). Therefore, we designed a set of unisensory tasks to independently extract model parameters and compare them with those fitted from the multisensory task. Such tasks were designed to match the VPD setup, to capture the relevant unisensory components as accurately as possible. However, their measure through independent tasks 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 fitter parameters is to be expected. Therefore, we will test for such correlation as an indication that the measured visuo-proprioceptive integration in the multisensory task relies on the proposed Bayesian CI process.

Two unisensory tasks are dedicated to the assessment of the proprioceptive-motor and visual precision. An open-loop reaching task (OL) will be used to isolate the motor-proprioceptive component (Figure 1 c). Participants will perform the same reaching movements towards virtual visual targets as in the multisensory task (VPD), but in the absence of visual feedback about hand position, allowing us to measure the proprioceptive-motor precision.

Concerning visual precision, it is worth noting that, despite visual acuity is extremely high compared to proprioceptive precision in humans [34,35], 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 [36]. In this view, we believe that participants' precision in visually determining their body midline, as assessed by a midline judgement (MJ) task (Figure 1 d), 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 will be measured through a simultaneity judgement task (SJ), whereby participants will evaluate the synchrony between the onset of a voluntary reaching movement and the displacement of the hand displayed in virtual reality (Figure 1 e).

As previously discussed, also the recognition of the higher-order visual features of the body may be seen as part of a (possibly optimal) inference process. 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 [10] and as suggested by Fang's study [25], 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 visual morphing (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 will measure participants' accuracy in recognizing their own hand. 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 will correlate with prior probability of cue combination fitted from the multisensory task.

Finally, several studies based on explicit localization tasks have found that the perceived hand position is biased towards the body midline [3,26,37–39]. In Bayesian approximations of brain function, priors are expected to reflect the marginal distribution in the natural statistics of sensory inputs. Therefore, we could conceive such systematic bias in hand proprioception as an informative prior

deriving from the overall distribution of the hand in space over a lifetime. Medial positions are privileged during grasping and object manipulation, which occupy 60% of human daily activities [40]. It is therefore reasonable to assume the overall distribution of hand positions to peak around the body midline. A Bayesian model including a midline prior would be able to explain why the RHI is attenuated when the rubber hand is placed laterally [10,41] and disappears when the rubber hand is at more than 30cm from the trunk [42] as well as why pathological forms of embodiment [43] emerge only when the alien limb is placed medially with respect to the patient's real hand. Yet, we are not aware of systematic investigations to assess the existence and the characteristics of such a prior, nor of its use to model behavioural and neuropsychological evidence. On the contrary, previous studies made this parameter statistically irrelevant, by using intentionally uninformative priors of large and fixed width [25,26]. This hypothesis may be justifiable in some contexts but arguably incorrect in others. Indeed, proprioceptive biases have been mainly observed in static localization tasks, while they do not seem to directly reflect in motor tasks [44,45], such as in our VPD. Therefore, in the present work, we propose to assess the presence of systematic errors in hand position sense through a proprioceptive judgement (PJ) task (Figure 1 f), and relate them to the Bayesian CI model in terms of a midline prior. For parsimony, we model such prior as a Gaussian of finite width, centred on the body midline. The signature allowing distinguishing such prior from what is typically designated as a generic bias would be for such bias to be systematically medial, and to increase linearly with the distance from the midline. We will further assess whether, counter our expectations, such prior generalizes to a motor task, by analysing data from the OL task. In that case, we would expect to observe an equivalent bias in the opposite direction during reaching movements. The presence of a prior might not cause noticeable effects in healthy participants, where the proprioceptive precision is far greater than the expected width of the prior, but it may become relevant in the case of a strongly impaired proprioception, as in brain-damaged patients. If the prior hypothesis is confirmed,

further versions of the Bayesian CI model used in patients will be modified to incorporate the value of the proprioceptive prior predicted by the PJ task.

In summary, the central hypothesis of our study is that body ownership arises from a quantifiable Bayesian CI process. This will be assessed by fitting our model to the results from a multisensory task (VPD), and comparing model predictions to an independent measure of the (unisensory) model parameters (Figure 1). Moreover, prior beliefs about hand ownership and position will be assessed and compared with their mathematical counterpart as fitted from the model.

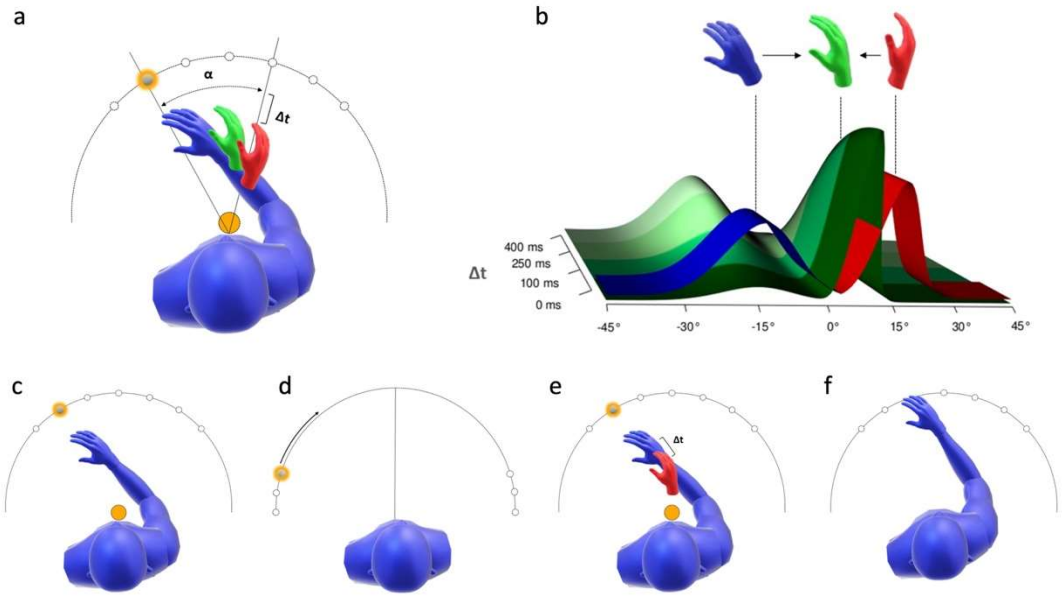


Figure 1. Experimental tasks and model rationale. In the visual-proprioceptive disparity task (multisensory task, a), a variable angle disparity α and temporal delay Δt are introduced between the real and the virtual hand during reaching movements towards a set of visual targets. The red hand represents 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 (c): in the open-loop (OL) reaching participants reach targets without visual feedback of the hand. Visual precision (d): 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 (e): 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) will be used to measure the accuracy in the encoding of the visual features of ones' own hand, as part of the prior. Finally, the presence of a proprioceptive prior centred on the body midline is assessed by a proprioceptive judgement task (PJ) (f). In this task, a virtual hand is displayed in virtual reality at the left or the right of participants' real hand, the perceived position of the hand will be determined using a two-alternative forced-choice converging algorithm.

2. Methods

2.1. Ethics information

Participants will be 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 will be run in accordance with the ethical guidelines of the ethical committee and the Declaration of Helsinki. Participants will be recruited using the online platform SonaSystem of the University of Lausanne (https://epflunil.sona-systems.com), and compensated 20 CHF per hour for their time.

2.2. 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 visuo-proprioceptive disparity). Besides the experimental design in terms of spatial (angular) disparities, the task was the same as described in section 2.3.3. 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°, ±13.3°, ±26.6° or ±40°). Each disparity was presented 7 times, each time with a different target out of 7 equally spaced targets between -45 and 45 degrees. 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 Figure 2 a, are 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 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 [25], we then compared the Bayesian CI model to the forced fusion 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 [46] (original code available from: https://github.com/sjgershm/mfit). We found this analysis to favour the Bayesian CI with an exceedance probability = 0.894, in line with the value reported by Fang [25] in his second experiment with 8 human participants (0.954), 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 (Figure 2 b 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°/40°. 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 = 0.940 and 0.905 respectively), yielding values of the parameters in line with our expectations (S01: σ_v = 7.1, σ_p = 4.63, σ_t = 0.118, P_π = 0.938, S02: σ_v = 9.64, σ_p = 5.98, σ_t = 0.101, P_π = 0.826).

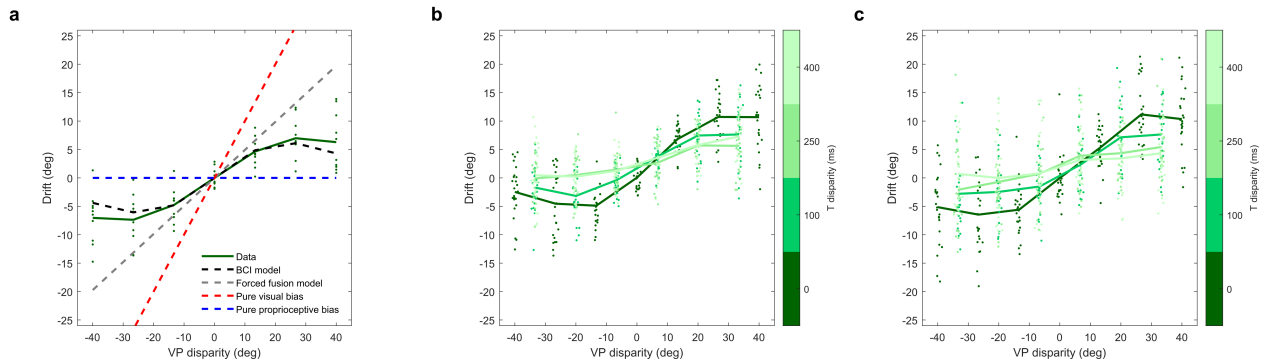


Figure 2. Pilot data for a purely spatial (a) and spatio-temporal (b,c) disparity setup. Panel (a) shows the results from a larger pilot study on 10 participants with the VPD task, with no temporal delay (as in Fang et al., 2019). The x axis indicates visuo-proprioceptive disparities, defined as the 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 forced fusion model of visual-proprioceptive integration, while the black dashed line shows the averaged predictions of our Bayesian CI 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 Figure 1 b). As expected, drift values increased at increasing temporal delays.

2.3. Design

2.3.1. Bayesian CI model

Following several successful approaches to model multisensory integration in probabilistic terms [25,26,28,47], we modelled the process of visuo-proprioceptive integration in a Bayesian Causal Inference (Bayesian CI) 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 will be applied to the integration of visual and proprioceptive inputs about the hand, focusing 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 visuo-proprioceptive spatial congruencies and visuo-motor temporal congruencies. Therefore, to derive the equations of the causal inference model, we will start 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 (t_v and t_p). 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 degrees range, approximating the set of reachable angles. Previous works [25,26] used a Gaussian centred in 0 and with very large standard deviation ($\sigma = 10000$) 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 information 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 to 30 seconds, 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 $x_v = s_v + N(0, \sigma_v)$, $x_p = s_p + N(0, \sigma_p)$, $t_v = \tau_v + N(0, \sigma_{tp})$, $t_p = \tau_p + N(0, \sigma_{tv})$. 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 degrees and 0/30 seconds range, where the likelihood functions quickly fall off, the likelihood functions defined by our generative model are:

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

$$\alpha = \int_{-90}^{90}\int_{0}^{3} P(x_p, x_v, t_p, t_v|C = 1) \approx 2\pi\sigma_s\sigma_t \cdot 30 \cdot 180 \quad (4)$$

Where α is the normalization constant, δ_s and δ_t denote spatial and temporal disparities respectively, and σ^2_s and σ^2_t are short forms for the sum of spatial and temporal variances. See supplementary information for details about the approximation in equation 4. When there are two separate causes, we simply have:

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

Therefore, the probability of common cause is given by:

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

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:

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

Then, we explored our model's predictions about sense of ownership (P_{com}) in a spatio-temporal disparity setup through numerical simulations. To illustrate model predictions, shown in Figure 3, 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 [28], 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 was extracted following equation 6. We show the average P_{com} values as a function of spatial and temporal disparity in Figure 3a. 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 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 the Bayesian CI 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. Figure 3 c and d show the ownership probability, as extracted from our simulated behavioural experiment and the reaching bias associated with the spatio-temporal disparities explored in our setup.

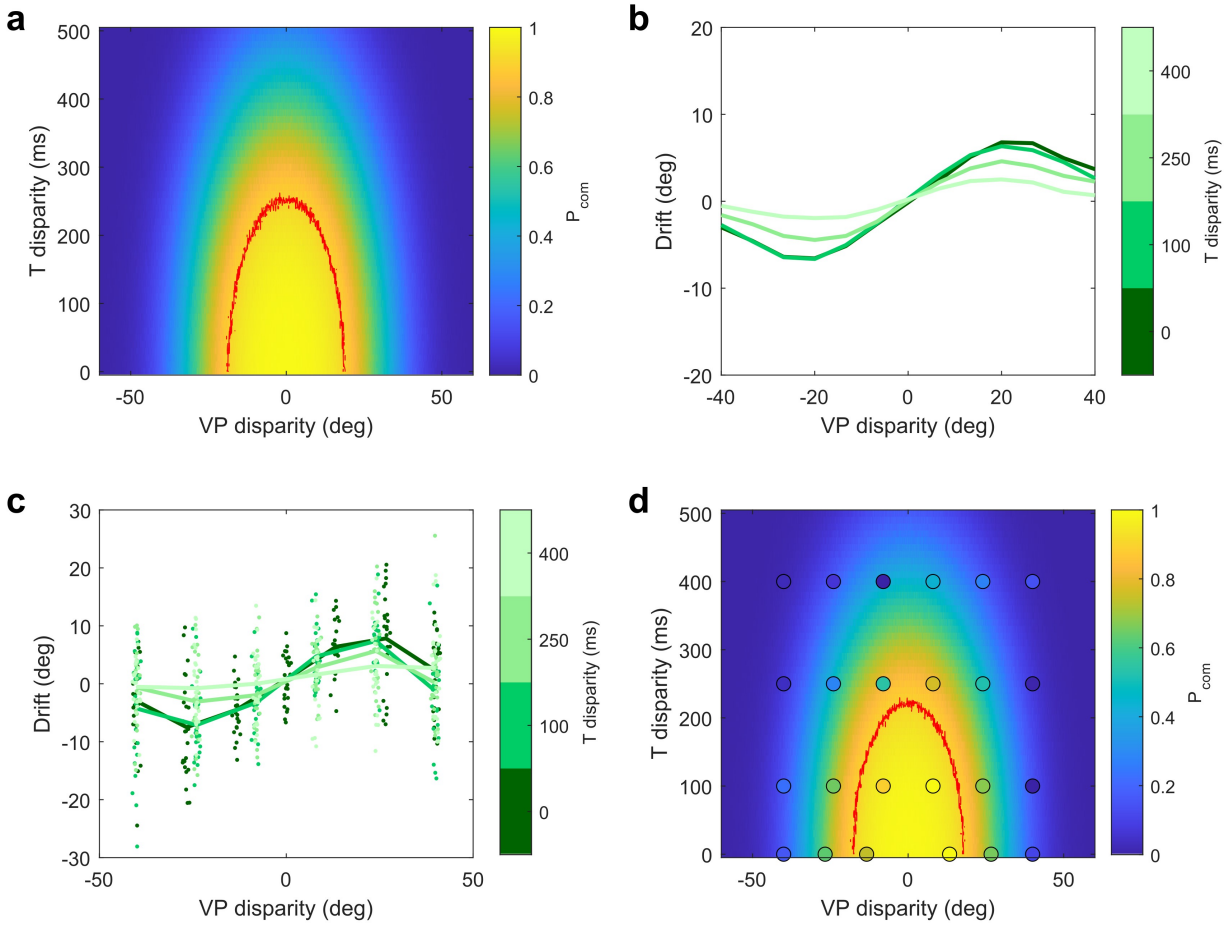


Figure 3. Model predictions and simulated experimental results for typical values of σ_v, σ_p, σ_t, P_π. Panel (a) shows the average probability of common cause as a function of spatial and temporal disparity, obtained after simulating model predictions for 1000 trials (in order to obtain a virtually 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 Bayesian CI 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 in order to extract model parameters, as they are recovered more robustly by directly fitting reaching errors from the VPD task.

2.3.2. Materials and General procedure

The battery of tasks is administered via the Oculus Rift S Virtual Reality system, comprising: Oculus Rift S head-mounted display (HMD) and two Oculus Touch, or motion controllers. 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 high tracking accuracy achieved by modern HMD allows to record kinematics with a precision that should be largely sufficient for our tasks (< 1 cm)[48,49]. Second, the usage of a commercial, readily available apparatus, allows a quick and stantardized 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 will consist of a main task (VPD) and 5 complementary tasks (OL, MJ, SJ, VM and PJ). VPD yields an estimate of ownership as the individual tendency to integrate proprioceptive and visual information of a virtual hand. 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 other tasks will assess independently the relevant parameters included in the Bayesian CI model: hand position sense in OL, visual precision in MJ, the temporal precision in SJ, and the encoding of the visual appearance of the hand in VM. The presence of a proprioceptive prior will be assessed in the PJ task. (see Figure 1).

2.3.3. 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 (Figure 1 a).

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 degrees 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 ms is added between the onset of the movement and the displacement of the virtual hand. For the 0 ms delay condition, 7 spatial disparities are used: 0°, ±13.3°, ±26.6° or ±40° (+: clockwise, CW; −: counter clockwise, CCW). For 100, 250 and 400 ms delay conditions, 6 spatial disparities, uniformly distributed on the same range, are used: ±8°, ±24° or ±40. 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*(7+6*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 minutes, and 600 trials will be collected over

approximately 75 minutes (including 5 minutes 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 second, and the target appears. This mismatch between the real (proprioceptive) and the virtual (visual) hand is maintained for 1.5 seconds 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 hand and return to the resting position within 1.5 seconds, ending the trial. Participants receive a positive feedback if they successfully reach in the target area and a negative feedback otherwise. The reaching target area is defined as a range between the target's angular position and the current angular disparity, ±5° of tolerance in both in clockwise and counter-clockwise directions. 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 verbally report their subjective feeling of ownership for the virtual hand, evaluating their agreement with the statement (adapted to VR from [25]) "I felt as if the virtual hand was my hand" on a -3 to 3 Likert scale. Values are manually recorded by the experimenter.

2.3.4. Open-Loop reaching task (OL)

The set-up of this task is similar to the VPD task (Figure 1 c), 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°, ±13.3°, ±26.6° or ±40° 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 seconds 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 seconds, ending the trial. Each participant is asked to complete 10 trials for each target (10*7 = 70 trials).

2.3.5. 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, starting from ±70° ±80° and ±90° from the body midline, on an arc centred on participants' sternum, with a radius equal to their maximum reaching distance, as in the VPD (Figure 1 d). Participants have to report when they feel that the visual cue is aligned with the midline of their body by pressing a response button on the motion controller. The starting positions of the visual cue are randomized across trials. 24 trials in total are collected.

2.3.6. Simultaneity judgement task (SJ)

This task consists of a series of reaching movements in virtual reality towards seven targets (from T1 to T7) arranged at 0°, ±20° or ±40 with respect to participant's sternum (Figure 1 e). 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 from 0 to 700 ms, equally spaced by 100 ms. 10 trials are collected per temporal disparity, with each target repeated twice. The order of targets and delays will be 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.

2.3.7. Visual morphing task (VM)

At the beginning of the task, 3 digital pictures of the participant’s right hand are taken, with 3 different posture 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 will be performed using an automated feature mapping software [50]. 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 degrees 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.

2.3.8. Proprioceptive judgement (PJ) task

The set-up of this task is similar to the VPD (Figure 1 f). 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°, ±15, ±30, ±45° 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 as it follows. At the beginning of each trial, a virtual hand is displayed at +30° (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. Sampling plan

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

There will be no outlier removal in the collected data. If at any point technical issues will arise during an experiment interfering with the experiment's procedure or data-logging, the participant will be 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 will be recognized by the experimenter, who monitors the procedure of all experiments on a separate display. If a participant wants to stop the experiment due to motion sickness or any other discomfort, he/she will be excluded from the experiment. Each task will be considered complete if the participant performed at least 80 % of the trials. Participants who fail to complete more than one task will be excluded from the experiment. In addition, any participant with less than 80 % of the trials in the VPD task will be excluded. All excluded participants due to the above reasons will be replaced with another participant.

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 [25] largest experiment to infer a plausible distribution of the main model parameters, and we simulated behavioural results from 500 surrogate participants, assuming that our Bayesian CI model, described in section 2.3.1, 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 will be computed. This should provide an unbiased estimate of the probability of observing an effect, assuming our model is correct.

The above-described Monte-Carlo approach was used for assessing the correlations between model

parameters extracted from VPD task (σ_p, σ_v, σ_t), and unisensory precisions extracted from the 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 [25] VPD experiment with 22 participants, the largest similar dataset available. The standard deviation of the Gaussians was set to 2 degrees. A lower and upper cut-off of 3 and 8 degrees 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 simultaneity judgement on data extracted from figures in the only previous study that was well adapted to this goal [51] indicates a value of σt of about 85 ms (see supplementary information 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 ms, which is symmetrical around 85 ms. Finally, values of P_π were uniformly distributed between 0.4 and 1, symmetrical around the average reported value of 0.7. Then, the VPD, 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, 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 repeatedly run the pipeline analysis described in the previous section on randomly selected subsets of simulated participants of increasing size and evaluated the probability of observing significant effects as a function of sample size. 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 Bayesian CI model on the multisensory task, and their unisensory counterparts from the other three tasks, with a threshold of $p = .05$. 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 10000 times (with replacement) from our pool of 500 simulated participants, and the fraction of resamples yielding significant results was computed for each of the three hypothesized correlations. 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 minutes. As summarized in Figure 4 a, the analysis shows that a sample of 20 participants should be sufficient to obtain a power above 95% for all the three correlations with a significance threshold of .05 (98.7 % for the MJ task with 15 participants, 99.7 % with 10 participants for OL and 96 % with 20 participants for SJ).

Similarly, we simulated the expected results in our PJ task, assuming that participants have a prior on hand position peaking on the midline and 20±5 degrees wide, with a lower cut-off at 10 degrees. This value was chosen as a plausible range of the natural statistics of hand positions. We then simulated (biased) proprioceptive judgements according to this assumption and tested it with the same procedure described above. We tested the hypothesis of such attractive prior by assessing whether the mean slope of the fit predicting the (simulated) perceived position as a function of the real position was significantly below 1, indicating a bias towards the midline at the population level. With the above determined sample size, power was well above 95 % (99.6 % with 10 subjects).

Concerning the hypothesis that the Bayesian CI model would outperform both the purely spatial Bayesian CI model and forced fusion model, we based ourselves on the sample of the largest experiment performed by Fang on humans, where a sample size of 22 participants was largely

sufficient to select the best model. The same holds for the expected correlation between subjective ratings and observed probability of common cause.

Concerning the hypothesis that subjective ratings will correlate with reaching biases, we based sample size estimation on the previous work by Fang [25], who performed a very similar analysis with fewer trials per participant. Assuming a true R value of this correlation to be equal to 0.82, a 95 % power can be obtained with 13 participants.

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 ~ 0.773. 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 10000 times N surrogated participants and generated n random vectors with an imposed correlation coefficient of 0.773 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 $\alpha = .05$. After computing the fraction of significant correlations, we found that a sample size of 25 would allow to reach a power of 95.5 %.

In summary, based on the hypothesis that requires the largest sample size, we estimate that 25 participants should be sufficient for meeting our statistical power target of 95 %. To be conservative, we will oversample to 40 participants.

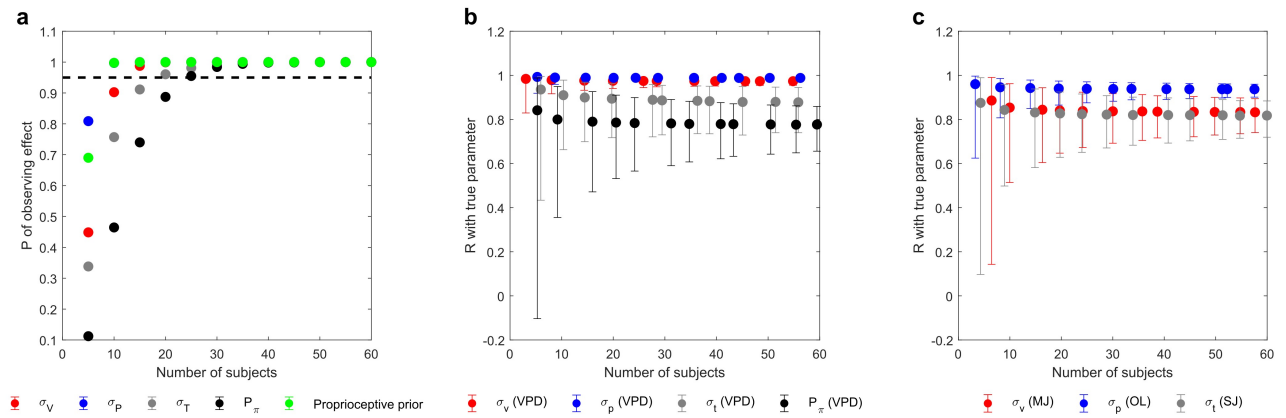


Figure 4. Power analyses. Panel (a) shows the results of the power analysis conducted by Monte-Carlo simulation. We plot the probability of observing a significant effect ($p<0.05$) as a function of the number of participants. The red, blue and grey lines denote respectively the correlations between the values of σv, σp, σt as extracted from the multisensory task (VPD) and its unisensory correspondents. The black line indicates the estimated probability of observing a significant correlation between the slope of the VM task and the value of Pπ extracted by fitting in the VPD task. The green line denotes the probability to detect the presence of an informative prior on proprioception, assuming its true average width is 20 degrees. Probabilities for each sample size were computed over 10000 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 10000 random draws of simulated participants. The color-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.

2.5. Analysis plan

Overall, the main hypothesis of our study is the validity of our Bayesian CI model in both the spatial and temporal domains as potential mechanisms of body ownership. In order to confirm this

hypothesis, results should demonstrate:

a) That the reaching bias in the VPD task is well modelled by a Bayesian CI model of visuo-proprioceptive integration, where the relative weights of vision and proprioception are based both on spatial and temporal congruency between the two cues.

b) That the reaching bias induced by the virtual hand, as resulting from the above processes represents a valid proxy of subjective ownership;

c) That the degree of integration (used as a proxy of ownership) depends on the actual precision of the unisensory channels, as estimated independently through the respective unisensory tasks.

d) That the degree of integration is also constrained by high-level priors which can be independently quantified, namely the ability of recognizing one's own hand via visual features.

e) That biases in static proprioception can be modelled as a spatial Bayesian prior, which is potentially relevant for ownership, and that the same biases do not transfer to dynamic proprioception.

To test these predictions, we will fit our model on our multisensory reaching task (VPD), similarly to what done by Fang and colleagues [25]. For each spatial and temporal disparity, we will simulate 5000 trials, and maximize the likelihood of the data given the simulated model predictions, with respect to the set of fitted parameters (σ_v, σ_p, σ_t, P_π). The fitting will be performed through the BADS Matlab optimization tool (https://github.com/lacerbi/bads)[52]. To avoid convergence problems or poor optimization, the fitting procedure will be repeated 5 times with different randomly selected starting parameters, and the fit with the highest log-likelihood will be selected. No pre-processing

step will be performed on the data, except for the removal of systematic biases in reaching (possibly due to tracking or VR calibration). This will be done by subtracting, for each participant, the mean reaching bias at zero spatial and temporal disparity. In order to assess the quality of fitting, we will test our model's exceedance probability (original code available from: https://github.com/sjgershm/mfit) against the forced fusion model, as described in the next section. Additionally, we will test the specific effect of time by comparing our model with an alternative Bayesian CI model, including only the effect of space, but no temporal manipulation effect. Finally, we will compare model predictions about P_{com} with subjective ratings about ownership. We will fit 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 will be 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 will simulate 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 figure 2d. Since, obviously, negative temporal disparities cannot be sampled, we will mirror the ratings and P_{com} values symmetrically with respect to the spatial axis, so to be able to perform the fits. We will then perform significance tests on Pearson correlation scores. The Gaussian fit will have 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.

As an additional model validation, we proposed to compare individual parameters extracted from the multisensory task with their unisensory correspondents.

Starting from the OL task, if we assume that participants are unbiased and the only source of variability in reaching movements is proprioceptive-motor noise, we expect:

$$x_r = x_t + N(0, \sigma_p) \quad (8)$$

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.

Similarly, in our MJ task, we expect

$$x_j = x_m + N(0, \sigma_v) \quad (9)$$

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

$$P_{sim} = P(\tau_p = \tau_v | t_p, t_v) = \frac{P(t_p, t_v | \tau_p = \tau_v) P(\tau_p = \tau_v)}{P(t_p, t_v | \tau_p = \tau_v) P(\tau_p = \tau_v) + P(x_p, x_v, t_p, t_v | \tau_p \neq \tau_v) P(\tau_p \neq \tau_v)})$$

$$= \frac{P_\sigma \frac{1}{\alpha} e^{-\frac{\delta_t^2}{2\sigma_t^2}}}{P_\sigma \frac{1}{\alpha} e^{-\frac{\delta_t^2}{2\sigma_t^2}} + (1 - P_\sigma) \frac{1}{30^2}} \quad (10)$$

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 information).

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 will perform a logistic fit on the judgements expressed as a function of the percentage of morphing according to the following model.

$$\ln\left(\frac{p_y}{1-p_y}\right) = \beta_0 + \beta_1 x + \varepsilon \quad (11)$$

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 will be given by - β_0/β_1, and the slope at that point by $\beta_1/4$.

We will use 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.

Similarly, we will assess 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.

Finally, our last purely proprioceptive task (PJ) aims at assessing the presence of a Bayesian prior on proprioception. We hypothesize such prior to be a Gaussian, centred around the midline of the body with a standard deviation of 20 degrees, as a reasonable approximation of the marginal distribution of hand positions across time. In that case, proprioceptive judgements are expected to be biased towards zero, with proprioceptive information and the prior weighted according to the inverse of their variance. This follows from the fact that the optimal estimate is the product of two Gaussian distributions, $P(s_p|x_p) = N(0, \sigma_{ps})$, and $P(s) = N(\mu_\pi, \sigma_\pi)$, where by σ_{ps} we indicate the precision of proprioception during the static localization task, as opposed from the σ_p of the dynamic-motor task. Then we can easily derive:

$$\mathrm{x_j} = \frac{\sigma_\pi^2 x_{ps} + \sigma_{ps}^2 \mu_\pi}{\sigma_{ps}^2 + \sigma_\pi^2} + \varepsilon = a + b x_p + \varepsilon \quad (12)$$

Where

$$b = \frac{\sigma_\pi^2}{\sigma_{ps}^2 + \sigma_\pi^2} \quad (13)$$

And

$$\varepsilon = N\left(0, \frac{\sigma_{ps}^2 + \sigma_\pi^2}{\sigma_{ps}^2 \sigma_\pi^2}\right) \quad (14)$$

This would allow extracting the mean and standard deviation of proprioceptive priors through a linear fit

$$\sigma_\pi^2 = \frac{b\sigma_p^2}{1 - b} \quad (15)$$

Then, assuming $\sigma_{ps} \sim 5$ our experimental hypothesis $\sigma_\pi = 20$ translates in an expected value for the slope of the fit ~ 0.94. As a conservative hypothesis, we expect that the average value of the slope

should be significantly below 1. Moreover, a negative correlation between the slope *b* and the static proprioception value σ_{ps} can be predicted as a direct consequence of equation (13). More precise participants are expected to be less influenced by the prior, and therefore have a higher value of the slope (closer to 1). This relates to the general principle in Bayesian inference that priors are more influential when less reliable information is available.

We hypothesize that the prior on static proprioception does not generalize to a motor task, such as the OL. If our hypothesis does not hold, we can assume that OL reaching movements have an equal and opposite bias with respect to proprioceptive judgements. According to the previous point of the power analysis, 10 participants should be sufficient to detect such departure from our hypothesis with a power of 99.7 %. If, counter to our hypothesis, reaching movements are biased compatibly with the presence of a proprioceptive prior on the midline, we will propose additional analyses in which an informative prior is replaced to the fixed, uninformative prior that we assumed.

Custom analysis code and all the data collected for this study will be available from: https://osf.io/azh8p/?view_only=abdd1b1d33c24794b048847432374720.

References

1. Blanke, O., Slater, M. & Serino, A. Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness. *Neuron* **88**, 145–166 (2015).

2. Botvinick, M. & Cohen, J. Rubber hands 'feel' touch that eyes see. *Nature* **391**, 756–756 (1998).

3. Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M. & Haggard, P. What is embodiment? A psychometric approach. *Cognition* **107**, 978–998 (2008).

4. della Gatta, F. *et al.* Decreased motor cortex excitability mirrors own hand disembodiment during the rubber hand illusion. *Elife* **5**, (2016).

5. Armel, K. C. & Ramachandran, V. S. Projecting sensations to external objects: Evidence from skin conductance response. *Proc. R. Soc. B Biol. Sci.* **270**, 1499–1506 (2003).

6. Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429–433 (2002).

7. Makin, T. R., Holmes, N. P. & Ehrsson, H. H. On the other hand: Dummy hands and peripersonal space. *Behav. Brain Res.* **191**, 1–10 (2008).

8. Ehrsson, H. H. Touching a Rubber Hand: Feeling of Body Ownership Is Associated with Activity in Multisensory Brain Areas. *J. Neurosci.* **25**, 10564–10573 (2005).

9. Tsakiris, M. & Haggard, P. The rubber hand illusion revisited: Visuotactile integration and self-attribution. *J. Exp. Psychol. Hum. Percept. Perform.* (2005). doi:10.1037/0096-1523.31.1.80

10. Tsakiris, M. My body in the brain: A neurocognitive model of body-ownership. *Neuropsychologia* **48**, 703–712 (2010).

11. Tsakiris, M., Carpenter, L., James, D. & Fotopoulou, A. Hands only illusion: multisensory integration elicits sense of ownership for body parts but not for non-corporeal objects. *Exp. Brain Res.* **204**, 343–352 (2010).

12. Apps, M. A. J. & Tsakiris, M. The free-energy self: A predictive coding account of self-recognition. *Neurosci. Biobehav. Rev.* **41**, 85–97 (2014).

13. Kilteni, K., Maselli, A., Kording, K. P. & Slater, M. Over my fake body: Body ownership illusions for studying the multisensory basis of own-body perception. *Front. Hum. Neurosci.* **9**, (2015).

14. Tsakiris, M. The multisensory basis of the self: From body to identity to others. *Q. J. Exp. Psychol.* **70**, 597–609 (2017).

15. Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J. & Friston, K. J. Bayesian inferences about the self (and others): A review. *Conscious. Cogn.* **25**, 67–76 (2014).

16. Seth, A. K. Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* **17**, 565–573 (2013).

17. Seth, A. K. & Friston, K. J. Active interoceptive inference and the emotional brain. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20160007 (2016).

18. Seth, A. K. & Tsakiris, M. Being a Beast Machine: The Somatic Basis of Selfhood. *Trends Cogn. Sci.* **22**, 969–981 (2018).

19. Limanowski, J. & Blankenburg, F. Minimal self-models and the free energy principle. *Front. Hum. Neurosci.* **7**, 547 (2013).

20. Alais, D. & Burr, D. The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Curr. Biol.* **14**, 257–262 (2004).

21. Jacobs, R. A. Optimal integration of texture and motion cues to depth. *Vision Res.* **39**, 3621–3629 (1999).

22. van Beers, R. J., Sittig, A. C. & Gon, J. J. D. van der. Integration of Proprioceptive and Visual Position-Information: An Experimentally Supported Model. *J. Neurophysiol.* **81**, 1355–1364 (1999).

23. Litwin, P. Extending Bayesian Models of the Rubber Hand Illusion. *Multisens. Res.* **33**, 127–160 (2020).

24. Noel, J.-P., Blanke, O. & Serino, A. From multisensory integration in peripersonal space to bodily self-consciousness: from statistical regularities to statistical inference. *Ann. N. Y. Acad. Sci.* **1426**, 146–165 (2018).

25. Fang, W. *et al.* Statistical inference of body representation in the macaque brain. *Proc. Natl. Acad. Sci.* **116**, 20151–20157 (2019).

26. Samad, M., Chung, A. J. & Shams, L. Perception of body ownership is driven by Bayesian sensory inference. *PLoS One* **10**, 1–23 (2015).

27. Shams, L. & Beierholm, U. R. Causal inference in perception. *Trends Cogn. Sci.* **14**, 425–432 (2010).

28. Körding, K. P. *et al.* Causal inference in multisensory perception. *PLoS One* **2**, (2007).

29. Krakauer, J. W. Motor Learning and Consolidation: The Case of Visuomotor Rotation. in 405–421 (2009). doi:10.1007/978-0-387-77064-2_21

30. Rohde, M., Van Dam, L. C. J. & Ernst, M. O. Statistically optimal multisensory cue integration: A practical tutorial. *Multisens. Res.* **29**, 279–317 (2016).

31. Costantini, M. *et al.* Temporal limits on rubber hand illusion reflect individuals' temporal resolution in multisensory perception. *Cognition* **157**, 39–48 (2016).

32. Motyka, P. & Litwin, P. Proprioceptive Precision and Degree of Visuo-Proprioceptive Discrepancy Do Not Influence the Strength of the Rubber Hand Illusion. *Perception* **48**, 882–891 (2019).

33. Lush, P. *et al.* Trait phenomenological control predicts experience of mirror synaesthesia and the rubber hand illusion. *Nat. Commun.* **11**, 4853 (2020).

34. Kniestedt, C. & Stamper, R. Visual acuity and its measurement. *Ophthalmol. Clin. North Am.* **16**, 155–170 (2003).

35. Jones, S. A. H., Cressman, E. K. & Henriques, D. Y. P. Proprioceptive localization of the left and right hands. *Exp. Brain Res.* **204**, 373–383 (2010).

36. Buneo, C. A., Jarvis, M. R., Batista, A. P. & Andersen, R. A. Direct visuomotor transformations for reaching. *Nature* **416**, 632–636 (2002).

37. Fuchs, X., Riemer, M., Diers, M., Flor, H. & Trojan, J. Perceptual drifts of real and artificial limbs in the rubber hand illusion. *Sci. Rep.* **6**, 24362 (2016).

38. Erro, R., Marotta, A., Tinazzi, M., Frera, E. & Fiorio, M. Judging the position of the artificial hand induces a "visual" drift towards the real one during the rubber hand illusion. *Sci. Rep.* **8**, 2531 (2018).

39. Galigani, M. *et al.* Effect of tool-use observation on metric body representation and peripersonal space. *Neuropsychologia* **148**, 107622 (2020).

40. Kilbreath, S. L. & Heard, R. C. Frequency of hand use in healthy older persons. *Aust. J. Physiother.* **51**, 119–122 (2005).

41. Kalckert, A., Perera, A. T.-M., Ganesan, Y. & Tan, E. Rubber hands in space: the role of distance and relative position in the rubber hand illusion. *Exp. Brain Res.* **237**, 1821–1832 (2019).

42. Lloyd, D. M. Spatial limits on referred touch to an alien limb may reflect boundaries of visuo-tactile peripersonal space surrounding the hand. *Brain Cogn.* **64**, 104–109 (2007).

43. Garbarini, F. *et al.* When your arm becomes mine: Pathological embodiment of alien limbs using tools modulates own body representation. *Neuropsychologia* **70**, 402–413 (2015).

44. Longo, M. R. & Haggard, P. An implicit body representation underlying human position sense. *Proc. Natl. Acad. Sci.* **107**, 11727–11732 (2010).

45. Haggard, P., Newman, C., Blundell, J. & Andrew, H. The perceived position of the hand in space. *Percept. Psychophys.* **62**, 363–377 (2000).

46. Wozny, D. R., Beierholm, U. R. & Shams, L. Probability Matching as a Computational Strategy Used in Perception. *PLoS Comput. Biol.* **6**, e1000871 (2010).

47. Rohe, T. & Noppeney, U. Cortical Hierarchies Perform Bayesian Causal Inference in Multisensory Perception. *PLOS Biol.* **13**, e1002073 (2015).

48. Jost, T. A., Nelson, B. & Rylander, J. Quantitative analysis of the Oculus Rift S in controlled movement. *Disabil. Rehabil. Assist. Technol.* 1–5 (2019). doi:10.1080/17483107.2019.1688398

49. Spitzley, K. A. & Karduna, A. R. Feasibility of using a fully immersive virtual reality system for kinematic data collection. *J. Biomech.* **87**, 172–176 (2019).

50. Liao, J., Yao, Y., Yuan, L., Hua, G. & Kang, S. B. Visual Attribute Transfer through Deep Image

Analogy. (2017).

51. Farrer, C., Bouchereau, M., Jeannerod, M. & Franck, N. Effect of distorted visual feedback on the sense of agency. *Behav. Neurol.* **19**, 53–57 (2008).

52. Acerbi, L. & Ma, W. J. Practical Bayesian Optimization for Model Fitting with Bayesian Adaptive Direct Search. *Adv. Neural Inf. Process. Syst. 30* 1834–1844 (2017).

Question	**Hypothesis**	**Sampling plan (e.g. power analysis)**	**Analysis Plan**	**Interpretation given to different outcomes**
Can an implicit hand ownership measure be well modelled as emerging from a Bayesian CI process combining sensory information in the temporal and spatial domain?	The proposed spatio-temporal Bayesian CI model, fitted to reaching error in the VPD task, will outperform both the forced fusion model and the purely spatial Bayesian CI model.	Based on Fang's 2019 study, a sample size of 22 participants was sufficient to show strongly significant evidence in favour of the Bayesian CI model.	The model will be tested against the forced fusion model and the purely temporal Bayesian CI model. We will use BIC as an approximation for model evidence, and compute the exceedance probability across participants to select the best model.	If the spatio-temporal CI model outperforms both the forced fusion and the temporal CI models, temporal features contribute to determining the probability of common cause in visuo-proprioceptive integration. This can be seen as an indication that the underlying inference process is unitary.
Does explicit subjective feeling of ownership emerge as a Bayesian CI process combining sensory information in the temporal and in the spatial domain?	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.	Assuming a true R value of 0.82, based on Fang's 2019 study, a 95% power with a significance threshold of 0.05. can be obtained with 13 participants.	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.	If both tests yield the expected results, then the subjective feeling of ownership can be quantitatively and implicitly evaluated using a Bayesian CI model that takes into account both the spatial and temporal dimension of multisensory perception.

Are the weights assigned to visual, proprioceptive and temporal information by the Bayesian CI model predicted by actual unisensory precisions measured in independent tasks?	The parameters estimated from the unisensory tasks (σ_p, σ_v and σ_t) will correlate with the correspondent parameters extracted from the multisensory task.	20 participants will be necessary to have a power larger or equal to 96 % for all the three correlations, with a significance threshold of 0.05.	Significance test on Pearson correlation scores. Scores are expected to be positive.	If the test yields the expected result, then the weights assigned by the model to the parameters are related to the precision of the corresponding unisensory functions, providing a strong evidence for the validity of the Bayesian CI model. Moreover, this would conclusive proof that the inference process leading to body ownership is truly behaviourally optimal, providing evolutionary motivation for the model.
Can cognitive constraints modulating body ownership (e.g. higher order visual features) be quantified as part of a Bayesian CI process?	Cognitive constraints, such as the visual recognition of the hand, have a role in the causal inference process determining ownership that can be quantified as a prior on cue combination.	With the indicated sample size of 25, a standard power analysis indicates a probability of 95.5 % of detecting the effect.	Pearson correlation test between $P\pi$ extracted from the VPD and the slopes of the psychometric function of VM (eq. 11). The score is expected to be significantly positive.	If the test yields the expected result, then so called cognitive constraints can be recast as a quantifiable part a of a Bayesian CI process.
Is the notion of Bayesian priors applicable to proprioception?	There is a systematic bias towards the midline in static hand position sense that can be well modelled as the result of a proprioceptive prior.	10 participants will be necessary to have a power of 99.7 % for all the correlation with a significance threshold of 0.05.	Significance test versus 1 of the slope of the linear regression on PJ (parameter *b*, eq. 12 and 13). The slope is expected to be < 1. Pearson correlation between slope and σ_{ps}. The score is expected to be	If the slope is significantly > 1 then there is a bias towards the midline in static hand position sense. If both tests yield the expected results, then there is a prior on hand proprioception

			significantly positive.	centred around the body midline.
Does this prior play a role in the Bayesian CI process determining ownership during reaching?	Systematic biases in static hand position sense do not affect performance in a motor task, when proprioception is used implicitly.	Assuming that OL reaching movements have an equal and opposite bias with respect to proprioceptive judgements, according to the power analysis of the previous point 10 subjects should be sufficient to detect this bias, if present.	Significance test versus 1 of the slope of the linear regression on the OL as calculated for the PJ (eq. 12 and 13). The slope is expected to be not significantly > 1.	If the test yields the expected result, then there is not a systematic bias in hand reaching movements that could bee explained by a proprioceptive prior. In the opposite case, the Bayesian CI model should be modified to take this bias into account as an informative proprioceptive prior

Supplementary materials

Is body ownership the result of Bayesian causal inference? A pre-registered study

Tommaso Bertoni[1*] & Giulio Mastria[1*], Henri Perrin[2], Boris Zbinden[1], Michela Bassolino[3], Andrea Serino[1]

*Authors equally contributed

[1] MySpace Lab, Department of Clinical Neurosciences, University Hospital of Lausanne, University of Lausanne, Lausanne, Switzerland

[2] School of Medicine, Faculty of Biology and Medicine, Unversity of Lausanne, Lausanne, Switzerland

[3] School of Health Sciences, HES-SO Valais-Wallis, Sion, Switzerland

Effect of the width of a Gaussian prior on the fitted value of P_π

Here we show how the use of a Gaussian prior on hand position and movement timing, even when of very large and fixed width to simulate a uniform distribution, is not desirable as it affects fitted values of the prior probability of common cause P_π.

When using such a Gaussian prior, the likelihood functions defined by our generative model are:

$$P(x_p, x_v, t_p, t_v|C=1) = \frac{1}{4\pi^2\sqrt{\sigma_v^2\sigma_p^2 + \sigma_\pi^2\sigma_p^2 + \sigma_v^2\sigma_\pi^2}\sqrt{\sigma_{vt}^2\sigma_{pt}^2 + \sigma_{\pi t}^2\sigma_{pt}^2 + \sigma_{vt}^2\sigma_{\pi t}^2}}$$

$$exp\left[-\frac{1}{2}\frac{(x_v - x_p)^2\sigma_\pi^2 + (x_v - \mu_\pi)^2\sigma_p^2 + (x_p - \mu_\pi)^2\sigma_v^2}{\sigma_v^2\sigma_p^2 + \sigma_\pi^2\sigma_p^2 + \sigma_v^2\sigma_\pi^2}\right. \quad (16)$$

$$\left. -\frac{1}{2}\frac{(t_v - t_p)^2\sigma_{\pi t}^2 + (t_v - \mu_{\pi t})^2\sigma_{pt}^2 + (t_p - \mu_{\pi t})^2\sigma_{vt}^2}{\sigma_{vt}^2\sigma_{pt}^2 + \sigma_{\pi t}^2\sigma_{pt}^2 + \sigma_{vt}^2\sigma_{\pi t}^2}\right]$$

and

$$P(x_p, x_v, t_p, t_v|C=2)$$

$$= 1\ \frac{1}{4\pi^2\sqrt{(\sigma_v^2 + \sigma_\pi^2)(\sigma_p^2 + \sigma_\pi^2)}\sqrt{(\sigma_{vt}^2 + \sigma_{\pi t}^2)(\sigma_{pt}^2 + \sigma_{\pi t}^2)}} exp\left[-\frac{1}{2}\left(\frac{(x_v - \mu_\pi)^2}{\sigma_v^2 + \sigma_\pi^2}\right.\right.$$

$$\left.\left. + \frac{(x_p - \mu_\pi)^2}{\sigma_p^2 + \sigma_\pi^2}\right) - \frac{1}{2}\left(\frac{(t_v - \mu_{\pi t})^2}{\sigma_{vt}^2 + \sigma_{\pi t}^2} + \frac{(t_p - \mu_{\pi t})^2}{\sigma_{pt}^2 + \sigma_{\pi t}^2}\right)\right] \quad (17)$$

Where σ_π is the width of the spatial prior on hand position and $\sigma_{\pi t}$ is the width of the temporal prior on hand movements.

If, as typically assumed, $\sigma_v, \sigma_p \ll \sigma_\pi$ and $\sigma_t \ll \sigma_{\pi t}$, the above expressions can be approximated as follows

$$P(x_p, x_v, t_p, t_v|C=1) \approx \frac{1}{4\pi^2\sigma_\pi\sigma_{\pi t}\sqrt{\sigma_p^2 + \sigma_v^2}\sqrt{\sigma_{pt}^2 + \sigma_{vt}^2}} exp\left[-\frac{1}{2}\frac{(x_v - x_p)^2}{\sigma_p^2 + \sigma_v^2} - \frac{1}{2}\frac{(t_v - t_p)^2}{\sigma_{pt}^2 + \sigma_{vt}^2}\right]$$

$$\stackrel{\text{def}}{=} \alpha e^{-\frac{\delta_s^2}{2\sigma_s^2} - \frac{\delta_t^2}{2\sigma_t^2}} \quad (18)$$

Where δ_s^2 is the spatial disparity and δ_t^2 the temporal disparity.

$$P(x_p, x_v, t_p, t_v|C=2) \approx \frac{1}{4\pi^2\sigma_\pi^2\sigma_{\pi t}^2} exp\left[-\frac{1}{2}\left(\frac{x_v^2 + x_p^2}{\sigma_\pi^2} + \frac{t_v^2 + t_p^2}{\sigma_{\pi t}^2}\right)\right] \stackrel{\text{def}}{=} \beta e^{-\frac{x_v^2 + x_p^2}{2\sigma_\pi^2} - \frac{t_v^2 + t_p^2}{2\sigma_{\pi t}^2}}$$

$$\approx \beta \quad (19)$$

Since $x_v^2 + x_p^2 \ll 2\sigma_\pi^2$ and $t_v^2 + t_p^2 \ll 2\sigma_{\pi t}^2$.

Then the probability of common cause is given by

$$P(C=1|x_p,x_v,t_p,t_v)=\frac{P_\pi \alpha e^{-\frac{\delta_s^2}{2\sigma_s^2}-\frac{\delta_t^2}{2\sigma_t^2}}}{P_\pi \alpha e^{-\frac{\delta_s^2}{2\sigma_s^2}-\frac{\delta_t^2}{2\sigma_t^2}}+\beta(1-P_\pi)}=\frac{e^{-\frac{\delta_s^2}{2\sigma_s^2}-\frac{\delta_t^2}{2\sigma_t^2}}}{e^{-\frac{\delta_s^2}{2\sigma_s^2}-\frac{\delta_t^2}{2\sigma_t^2}}+\frac{\beta}{\alpha}\frac{(1-P_\pi)}{P_\pi}} \quad (20)$$

Therefore, in model predictions and fitted values, the prior probability of common cause P_π is conflated with the constant β, which only depends on the arbitrary values chosen for the widths of the spatial and temporal priors.

Detailed calculations involved in the approximation shown in equation (4)

The normalization constant α is given by

$$\alpha=\int_{-90}^{90}\int_{0}^{3} exp\left[-\frac{1}{2}\frac{(x_v-x_p)^2}{\sigma_p^2+\sigma_v^2}-\frac{1}{2}\frac{(t_v-t_p)^2}{\sigma_{pt}^2+\sigma_{vt}^2}\right]dx_v dx_p dt_v dt_p$$

$$=\int_{-90}^{90} e^{-\frac{1}{2}\frac{(x_v-x_p)^2}{\sigma_p^2+\sigma_v^2}}dx_v dx_p \int_{0}^{3} e^{-\frac{1}{2}\frac{(t_v-t_p)^2}{\sigma_{pt}^2+\sigma_{vt}^2}}dt_v dt_p \quad (21)$$

The integral can be solved separately for the spatial and the temporal variables. The function is essentially a Gaussian ridge defined on the square domain with $-90< x_v<90$ and $-90<x_p<90$, and directed on the line $x_v=x_p$. For the spatial part, after a change of variables, with u being directed as the Gaussian ridge and v perpendicular to it, we get

$$\int_{-9}^{90} e^{-\frac{1}{2}\frac{(x_v-x_p)^2}{\sigma_p^2+\sigma_v^2}}dx_v dx_p=\int_{-\sqrt{2}\cdot 90}^{\sqrt{2}\cdot 90}\int_{-|\sqrt{2}\cdot 90-v|}^{|\sqrt{2}\cdot 90-v|} e^{-\frac{u^2}{(\sigma_p^2+\sigma_v^2)}}dvdu \approx \int_{-\sqrt{2}\cdot 90}^{\sqrt{2}\cdot 90}\int_{-\infty}^{\infty} e^{-\frac{u^2}{(\sigma_p^2+\sigma_v^2)}}dvdu$$

$$=\sqrt{2}\cdot 180\sqrt{\pi(\sigma_p^2+\sigma_v^2)} \quad (22)$$

The approximation consists in changing the integration bounds in the second integral to minus/plus infinity, which is justified as the width of the Gaussian is much smaller than the integration bounds almost everywhere. The same can be done for the temporal variables

$$\int_{0}^{3} e^{-\frac{1}{2}\frac{(t_v-t_p)^2}{\sigma_{pt}^2+\sigma_{vt}^2}}dt_v dt_p \approx \sqrt{2}\cdot 30\sqrt{\pi(\sigma_{pt}^2+\sigma_{vt}^2)} \quad (23)$$

As a result, we get

$$\alpha=\sqrt{2}\cdot 180\sqrt{\pi(\sigma_p^2+\sigma_v^2)}\cdot\sqrt{2}\cdot 30\sqrt{\pi(\sigma_{pt}^2+\sigma_{vt}^2)}=2\pi\sigma_s\sigma_t\cdot 30\cdot 180 \quad (24)$$

Extraction of σ_t from existing data

In order to obtain a rough estimate of the temporal precision σ_t, we fit the model described in equation (10) to data from Figure 2b in Farrer et al. (2008). This work was chosen as it closely resembles our task for the estimation of the temporal precision. Subjects performed random hand movements while holding a joystick, with their real hand hidden and observing a virtual rendering of the performed movements at various levels of temporal delay. They had to focus on movement onset, and judge whether the observed movements were simultaneous with their own, delayed but still generated by them, or delayed and not generated by them. The two latter responses were grouped as "non-simultaneous" to match our experimental design. The data was graphically extracted from Figure 2b by computing the relative position of data points and y-axis ticks in Inkscape. We show the results of the fit in Figure S1.

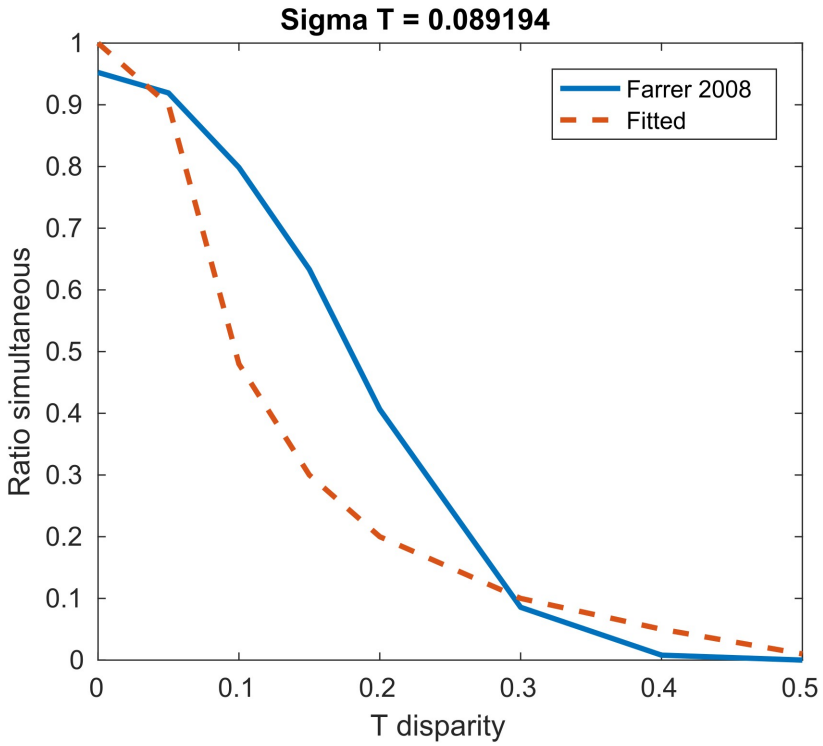


Figure S1. Estimation of σ_t from data from Farrer et al. (2008). The blue line represents the (group averaged) fraction of trials in which the virtual's hand movement perceived as simultaneous. The red dashed line represents a maximum likelihood fit of the data according to the procedure described in the main text (see equation 10 and the following paragraph).

Robustness of the fit of σ_t with respect to model priors and response criterion

Here we demonstrate the robustness of the proposed fitting approach to extract the temporal precision with respect to the two arbitrary parameters contained in the fit, i.e. the response threshold and the prior on simultaneity P_{sim}. To do so, we simulated responses to our task in surrogated subjects with different values of the response threshold and P_{sim}. As evident from Figure S2, changes in the response

threshold only appear to shift the response curve, while changes in σ_t are only reflected in a change of the slope of the response curve. Therefore, fitted values of σ_t should depend very weakly on the (arbitrary) values chosen for the response threshold and P_{sim}.

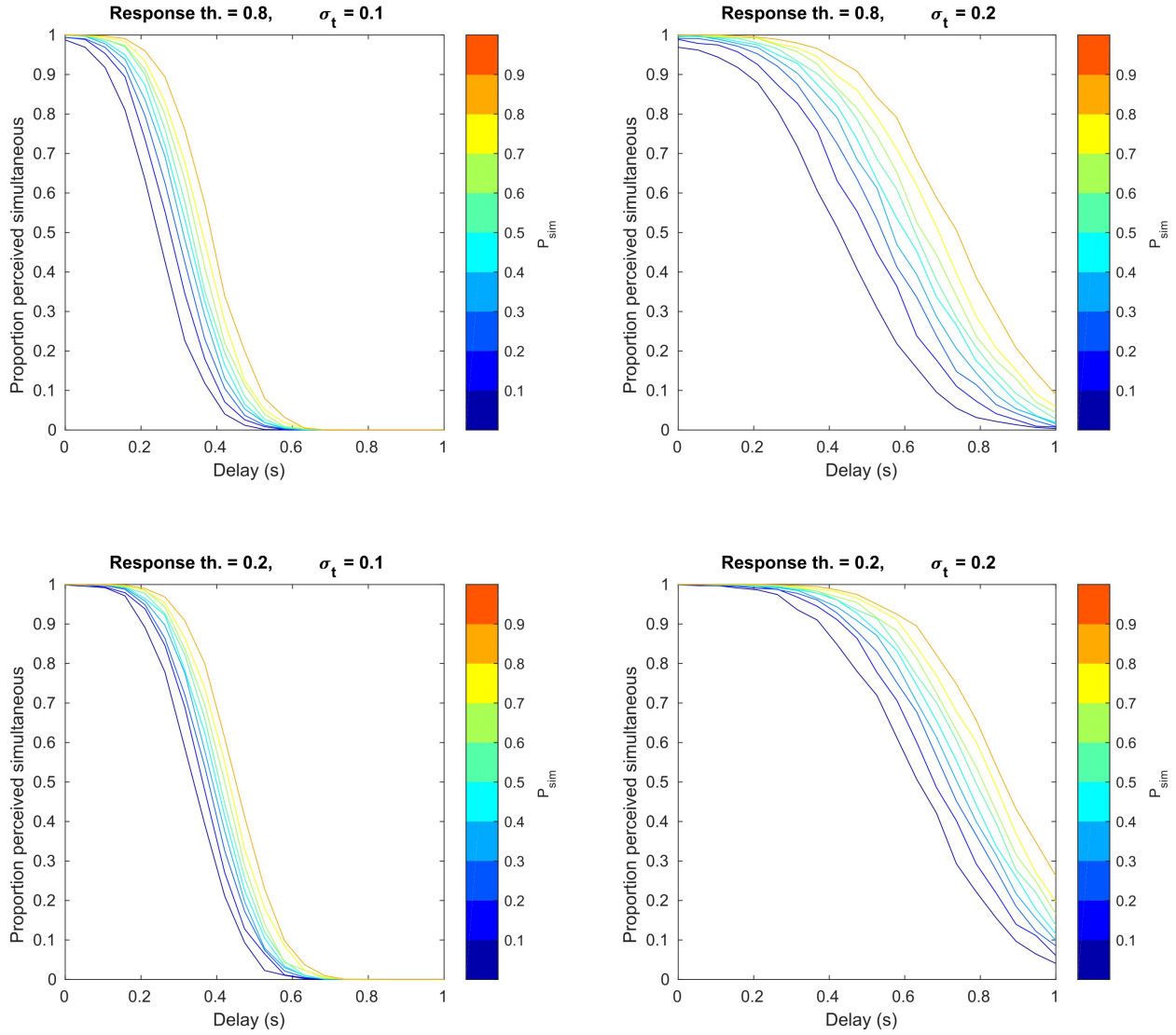


Figure S2. Simulations of the SJ task with different values of the response threshold (top-bottom), σ_t (left-right) and P_{sim} (color code). The slope of the curves does not depend on the response threshold or the prior on simultaneity, hence the value of σ_t can be robustly fitted.

Sense of Agency for intracortical brain machine interfaces

Andrea Serino[*1,2,], Marcie Bockbrader*[*3], Tommaso Bertoni[1], Sam Colachis[4], Marco Solca[2], Collin Dunlap[3], Kaitie Eipel[3], Patrick Ganzer[4], Nick Annetta[4], Gaurav Sharma[4], Pavo Orepic[2], David Friedenberg[4], Per Sederberg[5], Nathan Faivre[2,6], Ali Rezai**[7], Olaf Blanke**[2,8]*

1.MySpace Lab, Department of Clinical Neuroscience, University Hospital Lausanne (CHUV), Lausanne, Switzerland; 2. Laboratory of Cognitive Neuroscience, Brain Mind Institute & Center for Neuroprosthetics, Ecole Polytechnique Fèdèrale de Lausanne (EPFL), Campus Biotech, Geneva, Switzerland; 3.Department of Physical Medicine and Rehabilitation, The Ohio State University, Columbus, Ohio, US: 4 Medical Devices and Neuromodulation, Battelle Memorial Institute, Columbus, Ohio, US; Department of Psychology, University of Virginia, Charlottesville, Virginia, US. 6. Laboratoire de Psychologie et Neurocognition, Universitè Grenoble Alpes, Grenoble, France. 7. Rockefeller Neuroscience Institute, West Virginia University, Morgantown, West Virginia, US. 8. Department of Neurology, University Hospital, Geneva, Switzerland.

* Equal first author contribution; ** equal last author contribution

One Sentence Summary: Subjective sense of agency for decoded actions is processed in primary motor cortex and improves neuroprosthetic proficiency.

Abstract

Intracortical brain machine interfaces (BMI) decode motor commands from neural signals and translate them into actions, enabling movement for paralyzed individuals. The subjective sense of agency associated to actions generated via intracortical BMI, the involved neural mechanisms and its clinical relevance for BMI proficiency are currently unknown. By experimentally manipulating the coherence between decoded motor commands and sensory feedback in a tetraplegic BMI user, we provide evidence that primary motor cortex (M1) activity processes sensory feedback, sensorimotor conflicts and subjective states of BMI actions. Neural signals processing the sense of agency affected the proficiency of the BMI system, underlining the clinical potential of the present approach. These new findings show that M1 encodes information related to action and sensing, but also sensorimotor and subjective agency signals, which in turn are relevant for BMI applications.

Introduction

When performing a voluntary movement, motor commands from the brain activate body effectors, which produce a cascade of reafferent sensory (proprioceptive, tactile, visual) cues. Motor commands are also associated with prediction signals about the sensory consequences of the movement. The congruency between motor commands, reafferent sensory feedback, and sensory predictions is at the basis of the sense of agency, our feeling of being in control of our actions[1–3]. In case of damage to the motor system, motor commands that would trigger actions do not reach body effectors, leading to different types of paralysis, depending on the location and severity of damage. Intracortical brain machine interfaces (BMI) bypass such brain-body disconnection by decoding brain signals from different regions (i.e., primary motor cortex (M1), parietal or premotor cortex) and translating them into motor commands for the control of robots, exoskeletons[4,5], neuromuscular functional electrical stimulation[6,7] or other devices[8], enabling different actions (BMI actions) for patients with severe neuromotor impairments[9].

Despite major advances in intracortical BMIs based on research in human and non-human primates, the sense of agency for BMI actions, its neural mechanisms, and its impact on BMI performance is currently unknown

Although a few recent studies investigated the sense of agency using non-invasive brain computer interfaces (BCI) in humans[10,11], the following questions have never been asked using intracortical BMI. How does it feel to generate movements with a BMI – i.e., what is the sense of agency for BMI actions? Do motor neurons in human M1 encode not only motor commands, but also sensory feedback? Does these signals covary with agency for BMI actions? And does agency affect the efficiency of the BMI system - i.e. is agency of therapeutic benefit?

We applied classic approaches from psychophysics, neurophysiology, neuroengineering and virtual reality (VR) to ask these questions for the first time in a patient suffering from tetraplegia (caused by severe cervical spinal cord injury; C5/C6), who had been a BMI expert for two years before the start of the present study[6]. The patient had no preserved motor function below the C5 level. His sensory functions were

extremely limited and only showed partially preserved function at the C6 level on the left side and at C5 on the right side (there was also residual sensation for pressure on his right thumb). Concerning proprioception, he had preserved perception for shoulder, elbow and wrist joint position, but no proprioception for digits joint position (see Material and methods for more details).

The BMI consisted of a 96-channel array implanted in the hand area of left M1 and actuated a transcutaneous forearm neuromuscular electrical stimulation (NMES) system (see [6] for a full description of the system) to translate decoded cortical signals into right forearm and hand movements. In order to study the sense of agency for BMI actions and to evaluate its clinical impact, we experimentally manipulated the congruency between the decoded actions and the actions actuated by the BMI-NMES system. As illustrated in Figure 1, the participant was instructed to realize a cued action with the BMI and was provided with movement-related sensory feedback using visual (via VR) and/or somatosensory (via NMES) stimulation. Critically, this feedback was either congruent or incongruent with respect to the motor commands decoded from M1: half of the trials, in which the decoded action corresponded to the cued action (e.g., open hand), were associated with congruent feedback (e.g., open hand), while the other half were associated with incongruent feedback (e.g. the opposite action: close hand). For each BMI action, we asked the participant whether he felt in control of that action and to rate his confidence about this judgement, allowing us to (1) gauge the sense of agency for BMI actions and how this was modulated by the congruency between motor commands and sensory feedback. Next, neural data from the M1 implant were analyzed to measure how (2) the sense of agency and (3) sensory feedback were encoded in the activity of M1 neurons, quantified as multi-unit (MU) firing rates and local field potentials (LFP). Finally, we investigated (4) how visual and somatosensory feedback, and the associated sense of agency, affected the performance of the BMI system by changing the pattern of response of M1 neurons. By investigating what it feels like to control actions mediated by an intracortical BMI, our data show neural patterns in M1 activity (MU and LFP) reflecting the processing of agency for BMI actions, as generated by the congruency between intention and sensory feedback. Importantly, we show that the nature of somatosensory feedback (and the related sense of agency) affected the

efficiency of the BMI system by modulating the response properties of M1 neurons, underlining the clinical relevance of sensory feedback and agency for the BMI field.

During the experiment, the participant was cued to execute one of four target actions (hand opening, hand closing, thumb extension, thumb flexion) using a validated BMI neuroprosthesis. Neural activity corresponding to each target movement was recorded via a 96-channel microelectrode array in M1 and a nonlinear support vector machine classifier was applied to decode the participant's chosen action from MU activity (see [6] for full description). On each trial, the classifier provided the likelihood of each target action (on a -1 to +1 range, in 100 ms bins), thus decoding one of the four target actions from the participant's M1 activity. In three different experiments, visual, somatosensory, or visual-somatosensory feedback about the BMI action was provided (Figure 1). In Experiment 1, VR was used to provide visual feedback, consisting of a life-size virtual arm on a monitor superimposed over the participant's right arm, matching the location and dimensions of the participant's real arm, which was occluded from view. In Experiment 2, NMES was used to provide 'somatosensory' feedback: the patient's upper limb muscles were electrically stimulated so he could feel, but not see the selected movement. Experiment 3 combined VR and NMES to provide 'visual-somatosensory feedback' (see below). In half of the trials, sensory feedback was congruent with the cued action, while in the other half it was incongruent (i.e., the opposite, action was executed) (see Figure 1B). At the end of each trial, we gauged the participant's sense of agency (0 or 1; Q1) and confidence (rating between 0 and 100; Q2). Importantly, the amount of sensory information was kept constant across experiments, by providing non-informative sensorimotor feedback in Experiment 1 (i.e., a pattern of NMES triggering no BMI action) and non-informative visual feedback in Experiment 2 (i.e., a static visual hand performing no action).

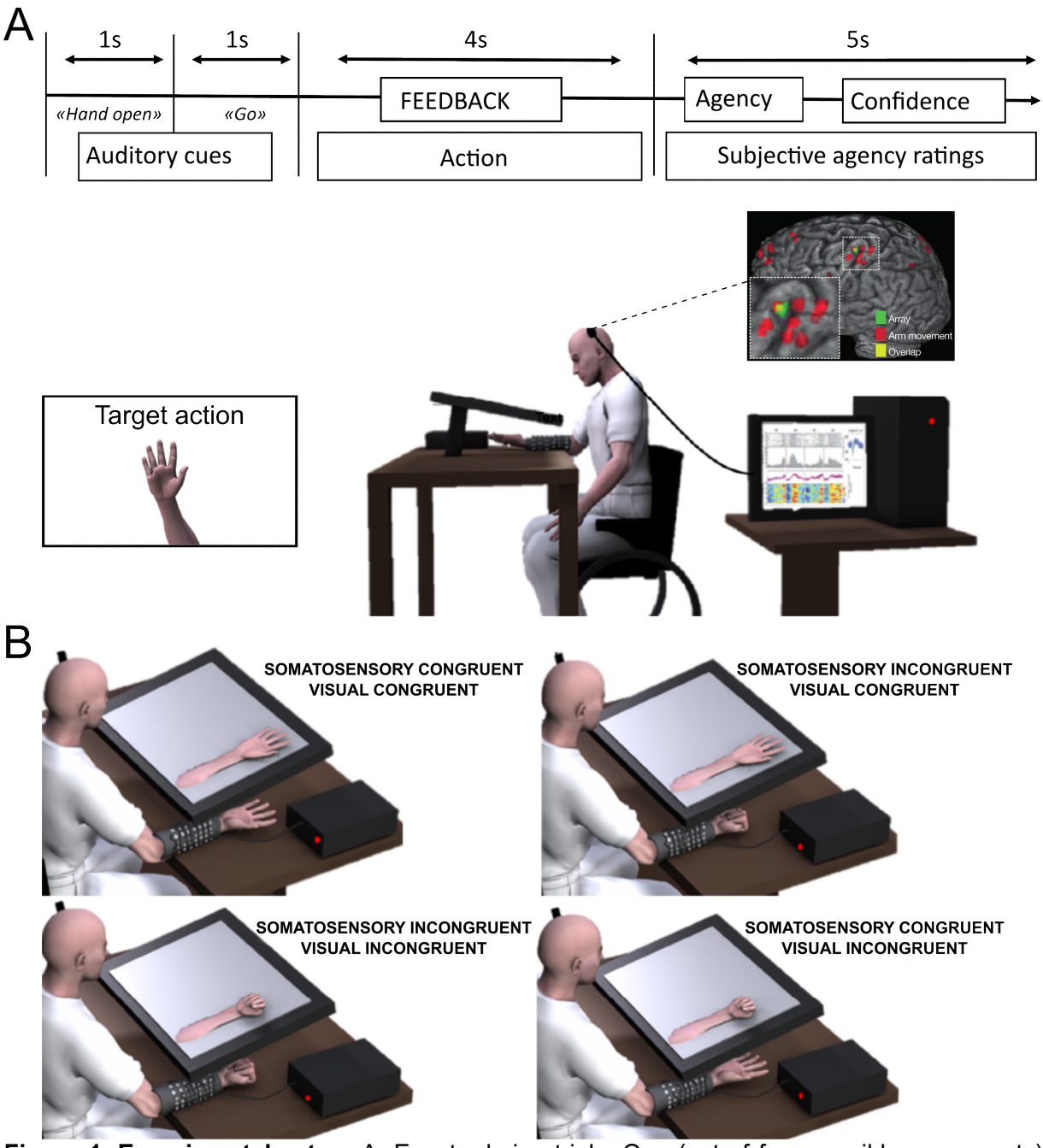


Figure 1. **Experimental setup.** A. Events during trials. One (out of four possible movements) was cued, following a "Go" signal to initiate the movement. The BMI classifier decoded the movement from M1 activity and sensory feedback was given. The patient answered two questions: Q1. "Are you the one who generated the movement?", by saying "Yes" or "No"; and Q2. "How confident are you?", by indicating a number ranging from 0 (absolutely unsure) to 100 (absolutely sure). B: Example of sensory feedback for one type of movement. The chosen movement was realized as a visual feedback, via virtual reality (VR – Experiment 1), as a somatosensory feedback, via NEMS (Experiment 2) or both (Experiment 3). In different congruency conditions, either the cued and correctly decoded movement (Congruent) or the opposite movement (Incongruent) was realized for the different modalities. The location of the

electrodes array in the M1, with respect to the pattern of activity for upper limb attempted movement from fMRI is also shown (from 6).

Results

Sensory feedback determines agency and confidence. Agency ratings were collected in a total of 844 trials (155, 243 and 448 trials for Experiments 1, 2 and 3, respectively; for Experiment 3 see below and supplementary material) and compared across feedback conditions using permutation tests. A null distribution of the mean agency rating was created by shuffling the condition labels over 10'000 iterations. P-values (2-sided) were estimated by counting the proportion of shuffled samples exceeding the observed average difference across conditions. As expected, and as shown in Figure 2A, we were able to manipulate agency and confidence for BMI actions. Thus, congruent visual (Experiment 1, 94% and 5.6% of positive responses to Q1 for congruent and incongruent trials, respectively, $p<.0001$) and congruent somatosensory (Experiment 2, 97.4% and 9% of positive responses for congruent and incongruent trials respectively, $p<.0001$) feedback resulted in more frequent agency responses versus incongruent conditions. Analyzing the role of feedback for confidence ratings (irrespective of the agency ratings), we found that confidence was not modulated by visual congruency (Experiment 1, mean Q2 rating = 70.9 for congruent, 73.6 for incongruent trials; $p = 0.28$), but by somatosensory congruency (Experiment 2, Q ratings were higher for somatosensory congruent [M = 74.1] than incongruent [M = 65] feedback; $p < 0.001$).

In order to disentangle the role of visual and somatosensory cues for agency and confidence, Experiment 3 combined VR and NMES including combinations of congruent and incongruent visual and somatosensory feedback (Figure 1). Most relevant are the comparisons between feedback conditions in which visual (V) and somatosensory (S) signals were both congruent (+) or both incongruent (-) (V+/S+; V-/S-) or when feedback was congruent in one modality and incongruent in the other modality (V+/S-; V- /S+). Results revealed that somatosensory congruency was more effective in driving the sense of agency and the associated confidence: ratings were stronger not only when both feedback signals were congruent (Q1 = 100% "Yes", mean Q2 = 83.8) as compared to both being incongruent (Q1 = 7.8% "Yes", mean Q2 = 72.3) (both p-value

< 0.001), but also in the V- /S-+ (Q1 = 68.5% “Yes”, mean Q2 = 59.4) as compared to the V+/S- condition (Q1 = 52.5% “Yes”, mean Q2 = 54.8; p = 0.0035 and p = 0.036, for agency and confidence respectively) (Figure 2). Collectively, these data from Experiments 1-3 show that the congruency between decoded actions and sensory feedback, especially for the somatosensory modality, alters the sense of agency and confidence for actions mediated by an intracortical BMI.

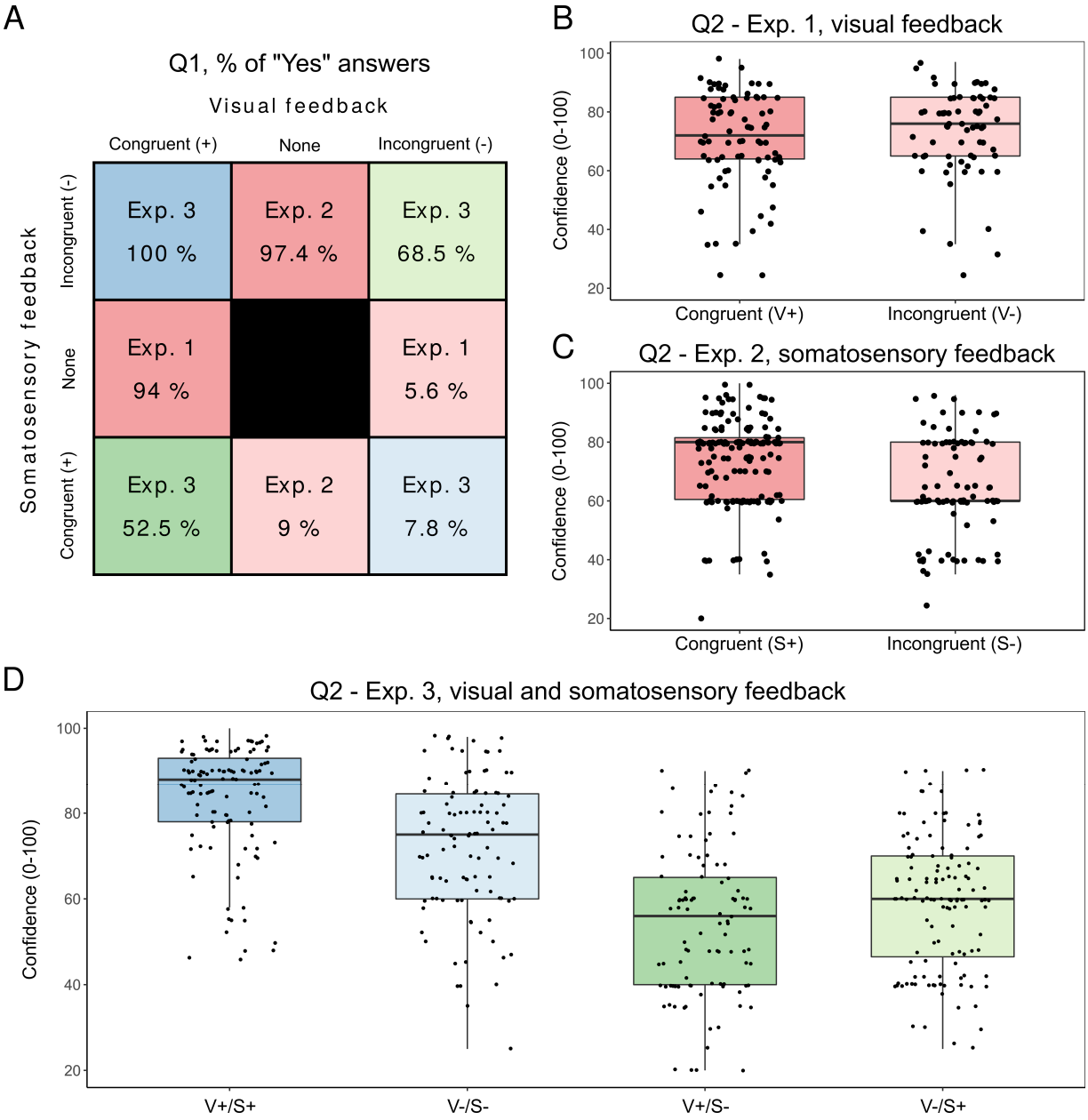


Figure 2. **Agency judgements and confidence depends on sensory feedback.** A. Proportion of “Yes” and “No” answers (Q1) to congruent and incongruent trials for the visual (Experiment 1) somatosensory (Experiment 2) and the combination of the two modalities

(Experiment 3). B-C-D: Confidence about agency judgments. Distribution of Q2 responses as a function of the congruency of visual (B), somatosensory (C) or both (D) sensory feedback.

The sense of agency has been traditionally studied by presenting participants with different visuo-motor couplings[2,12–15]. In comparison, the role of somatosensory signals remains poorly understood[16], notably because it is normally impossible to decouple motor commands, somatosensory feedback and visual feedback, with extremely rare exceptions as in deafferented patients. Here we were able to contrast feedback cues that were congruent in one modality (e.g., visual) and incongruent in the other modality (e.g., somatosensory; and vice versa) with respect to the motor command and demonstrate that somatosensory cues dominate the sense of agency and the associated confidence for BMI-NMES actions. Of note, this effect cannot be due to the presence of somatosensory cues alone, as BMI actions in the visual condition were always associated with non-informative NMES stimulation producing somatosensory sensations without generating any actions (i.e., pseudo random somatosensory feedback, see supplementary material). Collectively these psychophysical data in a BMI expert reveal that agency for BMI actions depends on visual and somatosensory feedback (tactile and proprioceptive input) with somatosensory cues being more relevant.

Cortical signatures of sensory feedback in M1. We next investigated how such sensory feedback, that modulated the sense of agency, was encoded in M1 activity. We first analyzed the LFP amplitude in the different feedback conditions across the three experiments, using a regularized generalized linear model (ridge regression) and input signals from each individual channel at every time point (see Supplementary information). As shown in Figure 3A (left), the analysis distinguished congruent vs. incongruent visual feedback (maximum Cohen's Kappa K=0.42; p-value for the difference from baseline, <0.001) within a single period of a positive potential that lasted from ~700-1200 ms after the BMI action classification onset (Experiment 1). We could also distinguish congruent vs. incongruent somatosensory feedback (maximum Cohen's Kappa K = 0.58; $p < 0.0001$) during two time periods: an early period characterized by a negative potential (stronger for incongruent feedback), starting at ~200 ms after BMI

classification onset, followed by a later persistent differentiation lasting almost until the end of the trial. These results were further corroborated by data from Experiment 3: congruent trials in both modalities were clearly distinguished from incongruent trials in both modalities, lasting from ~250-1900 ms after BMI classification onset (maximum K = 0.66). In addition, V+/S- trials were different from V+/S- trials from ~300-1400 ms from BMI classification onset (maximum K = 0.31) (Figure 3B left). These findings show that visual and somatosensory feedback were both encoded by LFPs in human M1 and that such M1-LFP coding started earlier and was more stable over time for somatosensory feedback.

Applying the same decoding algorithm as for LFPs, we next determined if sensory feedback was also encoded by the spiking rate of MU in M1 (for methods see Supplementary material). As shown in Figure 3A (right), in Experiment 1, MU activity distinguished between congruent and incongruent visual feedback from ~400-900 ms from BMI classification onset (max K value = 0.41, $p < .001$). Extending LFP findings, an earlier and more stable differentiation between congruent and incongruent somatosensory feedback was found in MU activity in Experiment 2, with an effect as early as ~200 ms from the BMI classification onset (max K value = 0.66, $p = < .001$) and then persisted from 800 to 2000 ms. Similar results were found in Experiment 3 (Figure 3B, right), where MU activity distinguished between trials congruent and incongruent in both modalities and between V+/S- and V-/S+ trials from ~160 ms from BMI classification onset. These data show that LFP and MU activity reflects visual and somatosensory feedback during actions driven by a BMI neuroprosthesis, with M1 activity early reflecting somatosensory feedback starting ~200 ms after NMES activation (~150 ms after BMI classification onset, ~200 ms before M1 activity encoding visual feedback) and persisting for a longer period.

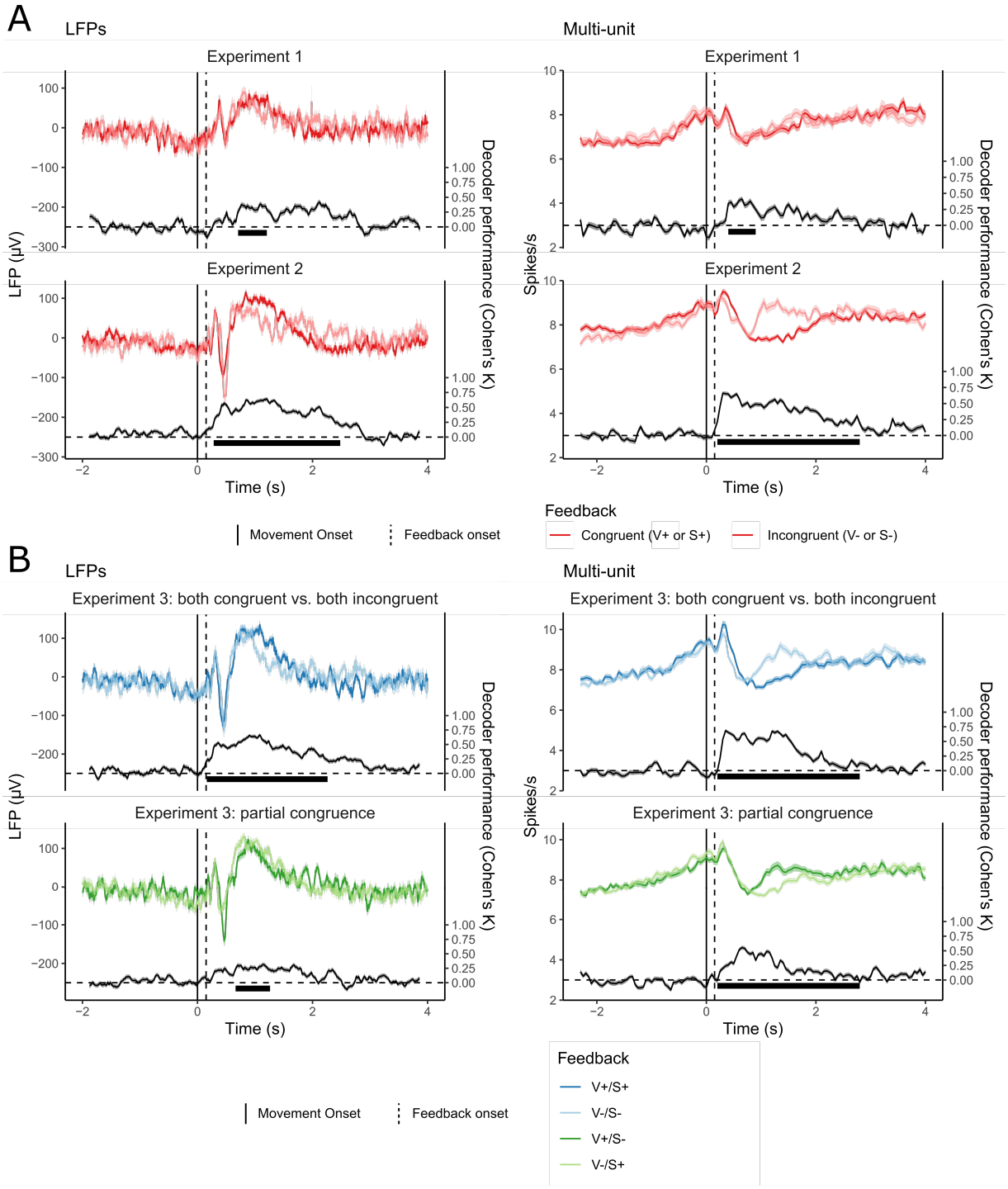


Figure 3. M1 activity depends on sensory feedback. Sensory feedback as encoded by Local field potentials (LFP; left panel) and Multiunit firing rates (MU; right panel). LFP and MU modulation for congruent and incongruent visual (Experiment 1) and somatosensory (Experiment 2) feedback (A) and for the combination of the two (Experiment 3, B). Colored lines represent averaged signal across all channels (shaded areas indicate SEMs); black lines report the time-related k-values of the multivariate decoder distinguishing between congruent and incongruent feedback; the underlying thick segments indicate k-values significantly higher than chance level from cluster-based permutation analyses.

The role of somatosensory and visual information is an important topic in motor control, with robust evidence showing how perturbations of sensory feedback impact motor execution and adaptation[17]. The present data show that the congruency between an intended action and somatosensory/visual feedback is encoded by M1 neurons at different latencies. To our knowledge, comparable data are not available in human or non-human primates, although previous studies in non-human primates described responses in M1 related to tactile and visual input[18,19], during active and passive movements[20] and during visual feedback of a pre-recorded movement[21,22]. The present results are consistent with proposals that suggest that M1 activity codes both for movement types and their sensory consequences, in line with recent proposals describing how M1 neurons encode different movement parameters (see [19,23,24] for reviews). Here we report that, at the population level, human M1 activity in addition discriminates between arm movements that were congruent or incongruent with the motor command, as defined by somatosensory and visual feedback, with higher accuracy, earlier and more consistent processing for the former type of sensory information. Thus, neural coding in M1 contains, at the population level, information not only about the movement itself, but also about sensory consequences of actions, involving somatosensory-motor and visuo-motor loops. These results are important to explain how sensory feedback affects the proficiency of the BMI system as described below.

Cortical signatures of the sense of agency in M1. It is known that sensory-motor congruency is a key mechanism of agency for able-bodied actions[2,3]; here we have shown that this also applies to agency and confidence for BMI-mediated actions and that LFPs and MU activity in human M1 distinguishes congruent vs. incongruent BMI actions. Next, we investigated to what extent LFP and MU activity in M1 also discriminate actions with and without an accompanying sense of agency. For each trial, we sorted LFP responses as a function of whether the participant reported agency or not. As seen in Figure 4A (left), LFP activity starting ~270 ms after BMI classification onset was found to code for agency and reached a maximum information value ($K>.4$) at ~1000 ms after BMI movement onset. Thus, BMI actions for which the participant felt

to be the agent were characterized by a different LPF pattern compared to BMI actions for which he did not. This was corroborated by MU activity analysis (Figure 4A, right). The MU firing rate was higher for trials with versus no agency; this discrimination started at ~300ms after BMI classification onset, until 500 ms, and peaked at ~400 ms (K max=.45). Later on, MU activity also differentiated for agency, with higher firing rate for trials with no agency (800-1600 ms after BMI classification). The same decoding was also able to discriminate trials with high vs. low confidence, based on a median split of Q2, from LFPs (max K = 0.296 at ~1200 ms) and MU (max K = 0.225 at ~400 ms).

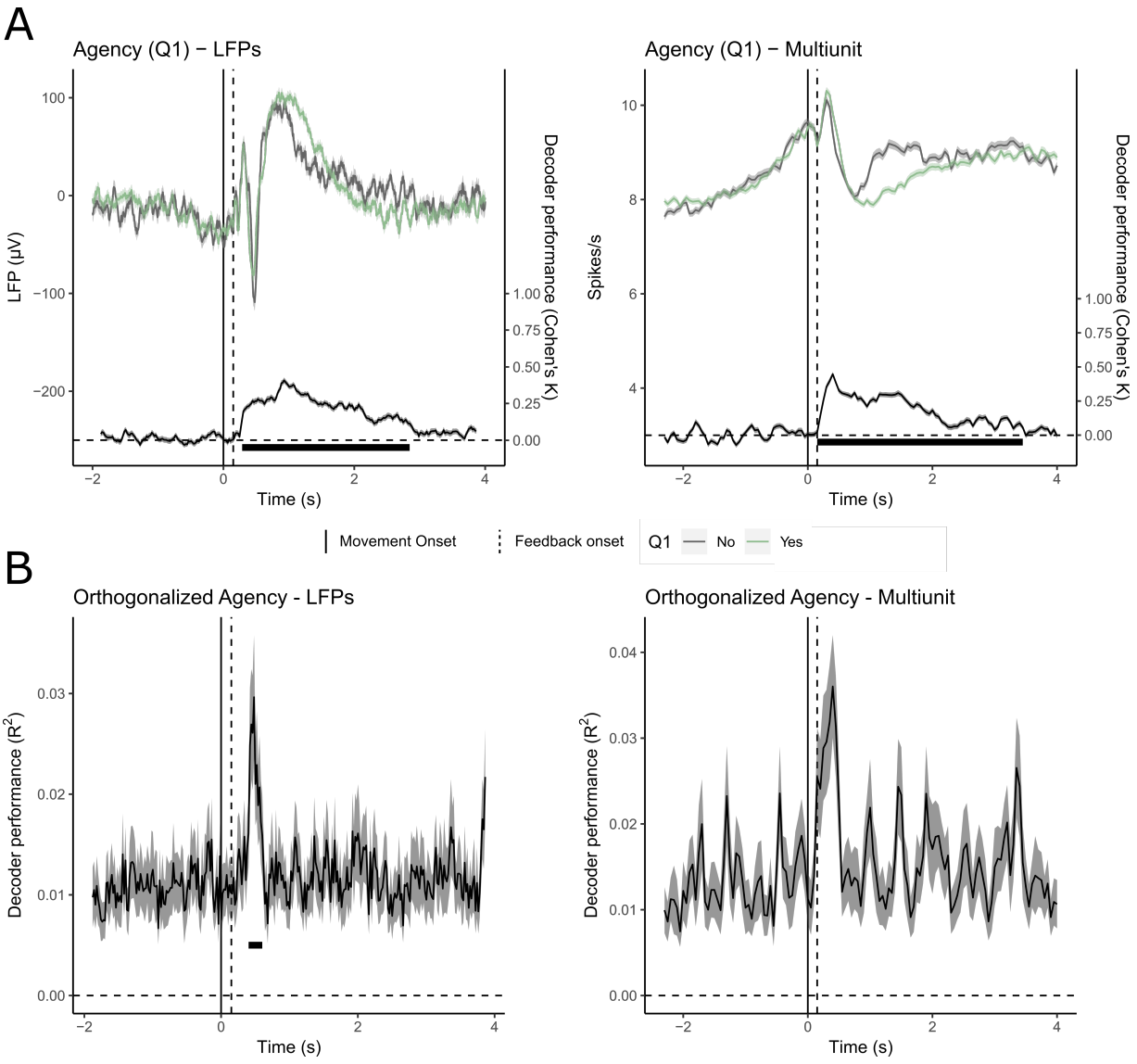


Figure 4. Sense of agency in M1. Sense of agency as coded by LFP (left) and Multi-unit firing rates (right). A. Left and right panels respectively show averaged LFP and Multi-unit modulation for high (green) and no (grey) agency response to Q1 (shaded areas indicate SEMs); black

lines report the time-related k-values of the multivariate decoder distinguishing the two conditions; the underlying thick segments indicate k-values significantly higher than chance level from cluster-based permutation analyses. B: Results of the decoder discriminating between high vs. low orthogonalized agency scores from LFP (left) and MU (right) after regressing out for the effects of the congruency of sensory feedback and type of movements.

In the experimental design, sensory feedback congruency was used to modulate the sense of agency and this may have influenced these agency findings. Accordingly, we next tested whether LFP and MU contained information related to the sense of agency per se, after controlling for the effect of sensory feedback. For this we built a continuous measure of sense of agency and confidence allowing us to regress out the effect of sensory feedback. This new index was computed by recoding confidence ratings (Q2) as -Q2, for trails with no agency (as indicated in Q1) and +Q2 for trials with agency (from Q1). This index was then orthogonalized with respect to congruency in order to regress out this effect from the agency scores. As M1 signals also varied as a function of the different cued actions (see SI), the index was also orthogonalized for the type of action. We then used the same decoder to predict orthogonalized agency scores from LFP and MU activity over time. This analysis shows that LFPs predicted the sense of agency starting at ~450 ms after BMI classification onset ($p < 0.02$ with respect to baseline) (see Figure 4B left). A similar pattern was found when considering MU activity, although the peak failed to reach significance after cluster-based correction for multiple comparisons (Figure 4B right). These data show that M1 activity encodes the sense of agency and associated confidence level and was modulated by the congruency between motor commands and sensory feedback. Thus, subjective mental states associated with BMI actions and control are encoded by M1 activity at the LFP level (and to a minor extent at MU), independent of the neural processing associated with sensory feedback (see supplementary material for single channel analyses).

Somatosensory feedback modulates BMI classifier accuracy. Given the strong role of sensory congruency in determining agency and its coding in M1, we finally asked whether sensory feedback has any impact on the BMI classifier. To this aim, we tested whether the congruency between the decoded motor commands and sensory feedback (visual, somatosensory) affected the accuracy of the BMI classifier, defined as the summed suprathreshold activation values across a 4s window. In Experiments 1 and 2

we found that congruent somatosensory feedback improved classifier accuracy ($t = 9.92$; $p < 0.0001$) (Figure 5B right), while there was no effect due to visual feedback ($p = 0.14$) (Figure 5A). Moreover, incongruent somatosensory feedback was associated with lower classifier accuracy for the cued movement (Figure 5B left), and even increased classifier accuracy for the opposite movement (Figure 5B left). Thus, only somatosensory feedback congruency affected BMI accuracy in the present participant. This was extended by the results of Experiment 3, where we found a significant main effect of sensory feedback condition ($F(3,444 = 15.83$; $p < 0.00001$; Figure 5C). Further post-hoc corrected tests showed that the BMI classifier's accuracy was higher when feedback was congruent, than incongruent, in either modality ($p < 0.0001$). More interestingly, when feedback was congruent for the somatosensory modality and incongruent for the visual modality (V-/S+) BMI accuracy was higher than in the opposite feedback condition (S-/V+) ($p < 0.001$). Figure 5 also shows the modulation of the BMI decoder as function of sensory feedback over time during the trial. Significant change of the decoder's output is visible from 430 ms from somatosensory feedback.

These data from Experiments 1-3 show that BMI performance is affected by the congruency between the decoded motor commands and the somatosensory feedback induced by the action actuated by NMES. This finding is also coherent with the more reliable (i.e., earlier, more long-lasting and better decoded) processing of somatosensory feedback from M1 activity (LFP, MU). The fact that the same action as actuated by NMES (e.g., open hand) increased or decreased the BMI classifier performance, depending on whether somatosensory feedback was congruent (open hand) or incongruent (close hand) with the cued action, excludes that this effect was a generic artifact of NMES stimulation affecting the input to the BMI classifier independently from sensory information. Moreover, the finding that visual feedback did not alter BMI classifier accuracy shows that congruency per se cannot account for changes in BMI performance.

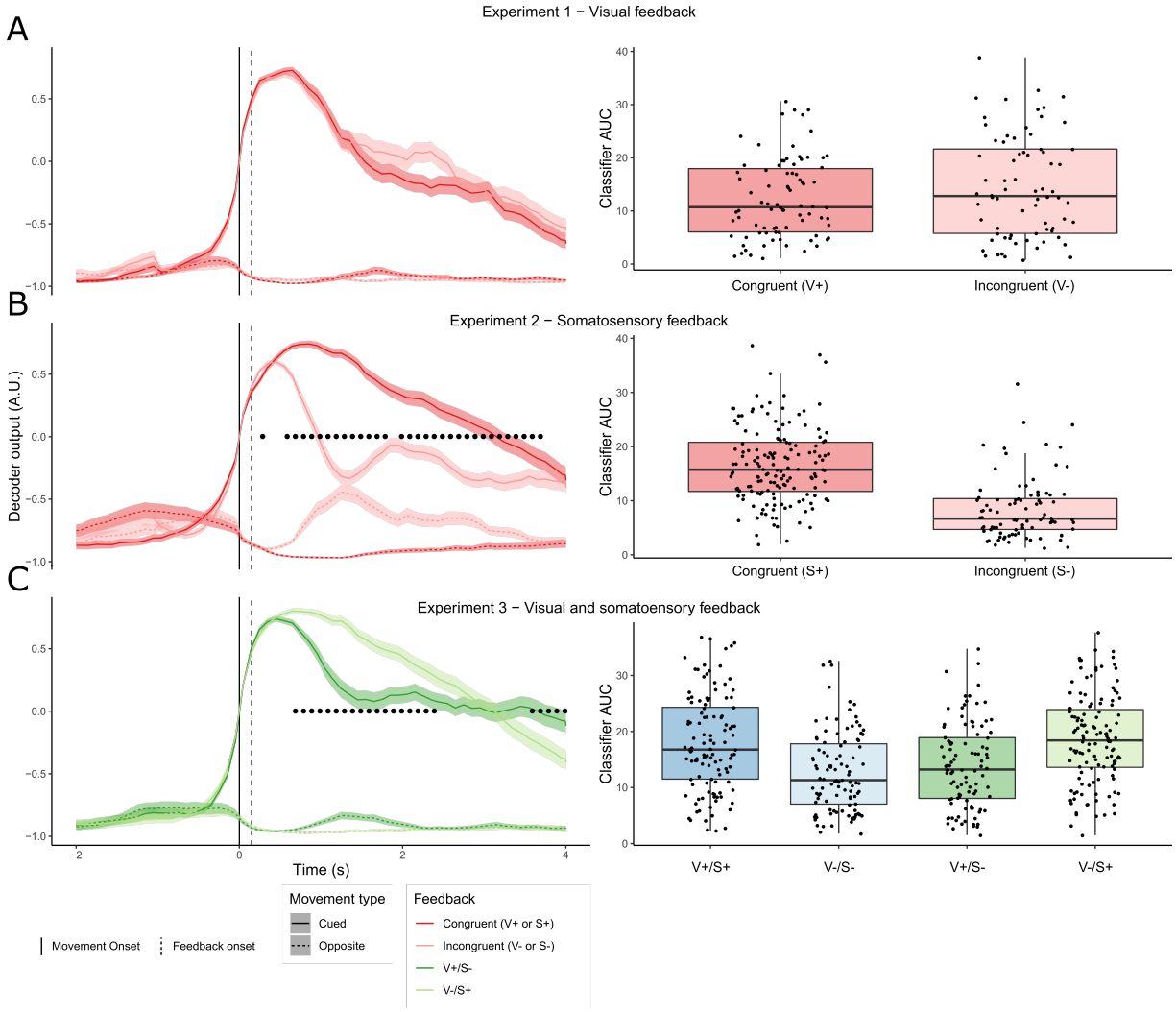
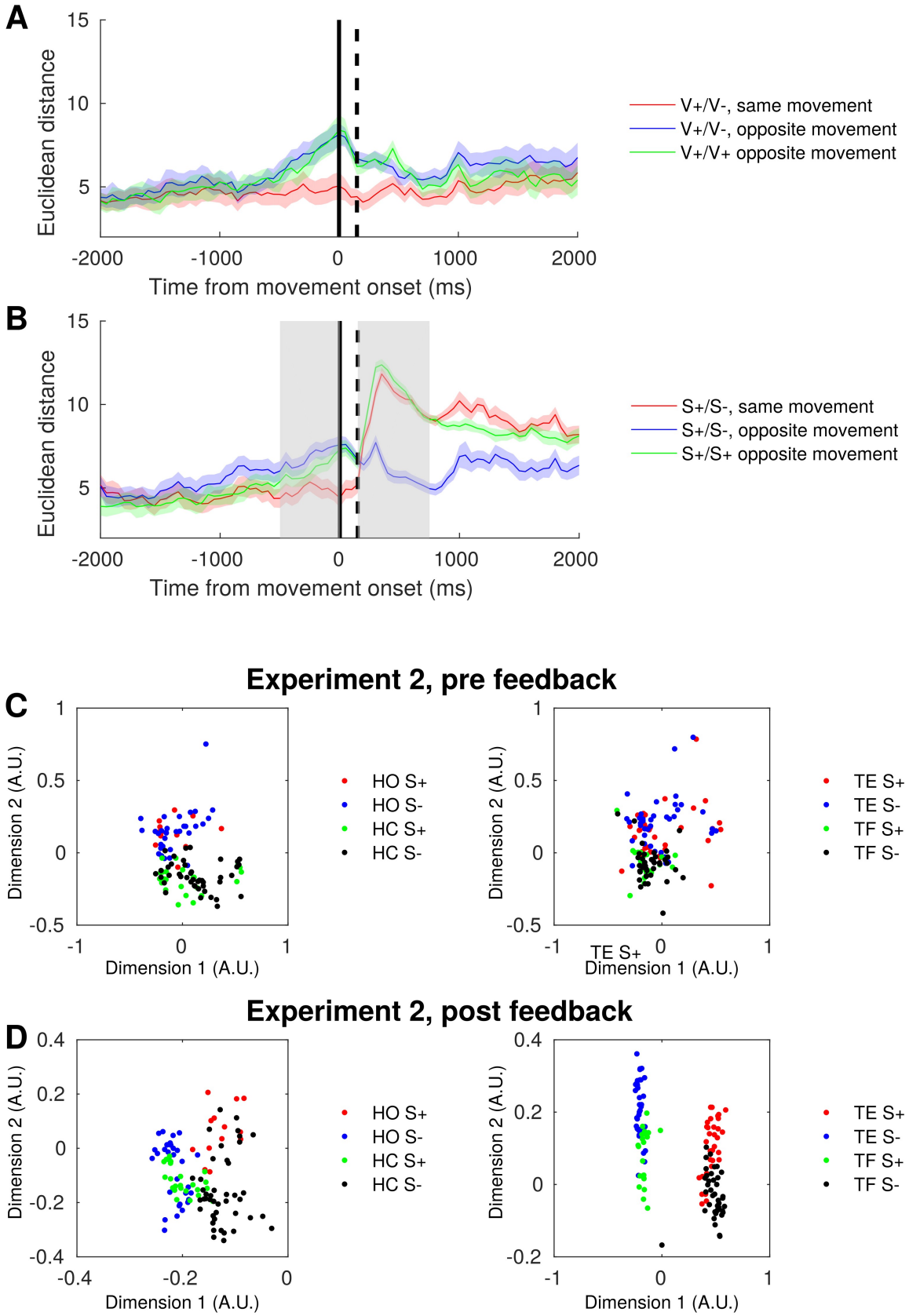


Figure 5. Performance of BMI classifier as a function of sensory feedback. The left panels (A, B, C) show the modulation in time of the performance of the BMI classifier for the 4 types of movements indicated for the cued movement (filled line) and the opposite (dashed line), as a function of feedback (i.e. black dots indicate time points with significant difference). The right panels show the area under the curve taken as an index of global performance of the BMI. The performance of the BMI classifier does vary not as a function of visual feedback (Experiment 1, A), but it is significantly better when somatosensory feedback is congruent both in Experiment 2 (B) and in Experiment 3 (C).

In order to better understand how somatosensory feedback affected the accuracy of the BMI classifier, we analyzed time point by time point changes in multiunit activity for the whole array. We computed the average Euclidean distance between firing patterns of trials with a given cued movement and either congruent or incongruent somatosensory feedback. For a given cued movement (e.g., movement hand open) at congruent

feedback (hand open), we computed its distance either with the same movement (cued: hand open) at incongruent feedback (hand close) or with the opposite movement (hand close), at its relative incongruent feedback (hand open). This way, we compared cases with the same motor intention, but opposite sensory feedback, and trials with the opposite motor intention, but the same sensory feedback. As shown in Figure 6, trials from Experiment 2 with opposite somatosensory feedback, but the same motor intention, diverge after sensory feedback, whereas trials with opposite motor intention, but the same sensory feedback, seem to even converge slightly with respect to baseline. This shows that M1 activity after feedback reflects the movement implemented via NMES more than the intended movement, thus explaining the modulation of somatosensory feedback in BMI proficiency (see Figure 6B). As a control, we also analyzed trials with opposite motor intention and congruent somatosensory feedback. We found the activity patterns to differ only slightly with respect to trials with same somatosensory feedback, but opposite motor intention, further showing that somatosensory feedback is prevailing over motor intention after movement onset. In the case of visual feedback (Experiment 1), instead, there was no divergence of activity patterns after the feedback, while trails with different motor intention clearly diverged before the movement onset (see Figure 6A).


A
Euclidean distance
Time from movement onset (ms)
V+/V-, same movement
V+/V-, opposite movement
V+/V+ opposite movement
B
Euclidean distance
Time from movement onset (ms)
S+/S-, same movement
S+/S-, opposite movement
S+/S+ opposite movement
Experiment 2, pre feedback
C
Dimension 2 (A.U.)
Dimension 1 (A.U.)
HO S+
HO S-
HC S+
HC S-
TE S+
TE S-
TF S+
TF S-
Experiment 2, post feedback
D
Dimension 2 (A.U.)
Dimension 1 (A.U.)
HO S+
HO S-
HC S+
HC S-
TE S+
TE S-
TF S+
TF S-

Fi

gure 6. Somatosensory feedback changes firing rates of M1 neurons. A-B. Euclidean distance in time between trials with same motor intention and opposite feedback (red), same feedback and opposite intention (blue), or opposite congruent feedback and intention (green), for experiment 1 (A) and 2 (B). In Experiment 1, neural activity diverged as a function of motor intention before the movement, as shown by the increase in Euclidean distances between the green and blue curves. In Experiment 2, neural activity diverged as function of sensory feedback after NMES activation. **C-D.** Multidimensional scaling of neural activity before (-650/150 ms; C) and after (0/500 ms; D) sensory feedback. The plots show a 2D dimensionality reduction of population activity in the target period, in order to represent it on a plane. As in a principal component analysis, Dimensions 1 and 2 can be seen as the two abstract coordinates explaining most variance in the data. Movements are separated by classes of hand (open/close; right) and thumb (extension / flection; left) movements.

In order to better display the effect of somatosensory feedback on M1 activity for each type of movements, we computed a 2D multidimensional scaling of neural activity as a function of intended movement and congruency of somatosensory feedback. This technique aims at representing the high dimensional spatio-temporal pattern of neural activity in 2D plane, while maximising the fraction of retained variance. As shown in Figure 6C, both for hand (opening/closing) and thumb (flection/extension) movements, before sensory feedback (in the window between -650 and -150 ms before sensory feedback onset), M1 neural activity is clustered solely as a function of the intended movement. After somatosensory feedback (between 0 and 600 ms from sensory feedback onset, Figure 6D), trials with congruent somatosensory feedback and a given intended movement are clustered more with trials coding for the opposite movement, but receiving the same sensory feedback rather than with trials coding for the same movement.

No prior study in humans and only few studies in monkeys directly tested the effects of sensory feedback on BMI performance[21,25]. Here we show, for the first time, an effect of feedback congruency on BMI performance, and the underlying role of M1 in this process. Our findings indicate that the recorded M1 units processed motor signals for the trained BMI actions, for sensory and sensory-motor signals reflecting the type and congruency of the sensory feedback. Importantly, these processes were found to change across time, as a function of the sensory feedback provided. In particular, our results show that, after somatosensory feedback, the pattern of neural activity from M1

reflected more closely the type of movement realized by the NMES (i.e., the pattern of somatosensory feedback) rather than the intended and decoded movement. This re-writing of the encoded M1 movement as a function of the NMES-implemented movement directly relates to the improvement of BMI efficiency based on congruent somatosensory feedback that we observed and was absent in visual feedback trials.
This effect might be mediated by mutual connections between the primary motor and the primary somatosensory cortices, which have been extensively documented in non-humans primates[26] and in humans[27]. In addition, this effect might also depend on direct somatosensory inputs reaching M1 neurons likely from the dorsal columns via the ventrolateral thalamic nucleus[28]. This is an important finding, considering that original BMI approaches for severely motor-impaired patients generally provide visual feedback only[5,29] or somatosensory feedback by directly stimulating primary somatosensory cortex[30–32] (see [31] for a review). Although from a single tetraplegic participant, the present data show that non-invasive somatosensory feedback via NMES not only enables higher subjective feeling of being in control (agency and confidence), but also leads to better actual control of the patient's BMI actions.

Agency covaries with BMI classifier. We finally investigated whether agency has an impact on BMI efficiency and thus tested whether the sense of agency covaried with BMI classifier accuracy. We found that trials with agency versus trials without agency were associated with higher classifier accuracy. However, this was only the case when somatosensory (Experiment 2; $F(1,239=4.23$; $p<.05$), but not visual feedback was modulated (Experiment 1; $p=.14$), as also confirmed from analysis of data from Experiment 3 ($F(1,441=6.94$; $p<.001$). In addition, there was a significant correlation across all three experiments between BMI classifier accuracy and confidence (Q2, $F=46.95$; $p<.001$; $r^2=.10$; See Supplementary Table 1 for multiple regression analyses). Thus, agency and confidence were both directly related to the performance of the present BMI system, but only when somatosensory feedback was involved. In order to confirm the role of agency on BMI performance, while controlling for other potential factors, we compared the BMI performance between trials in which the BMI user reported high and low agency, within conditions at equivalent sensory feedback, that is

V-/S+ and V+/S- from Experiment 3 (which resulted in a balanced and sufficient numbers of trials with “Yes” and “No” responses to Q1). As shown in Figure 7, BMI accuracy varied as a function of subjective agency judgments, in conditions of equivalent sensory feedback. BMI accuracy was significantly higher in trials with high agency as compared to trials with low agency from 300 ms in the V-/S+ condition. The same pattern is visible in the V-/S+ condition, although the comparison was not significant (i.e. did not survive to correction for multiple comparisons). The same analysis run on confidence ratings (by sorting high and low confidence ratings by means of a median split) did not show any significant difference in BMI accuracy due to confidence at equivalent conditions of sensory feedback (see Supplementary material, Figure S6). These results suggest that the sense of agency, and not confidence (see supplementary Table 1 for further analyses), has an effect on BMI accuracy beyond the prominent role of sensory feedback, and impacts BMI accuracy at a later time point. Since agency judgments and confidence ratings reflect two different processes of subjective experience, the present data suggest that pre-reflexive, rather than post-decisional agency components more strongly affect the proficiency of a BMI decoder in M1.

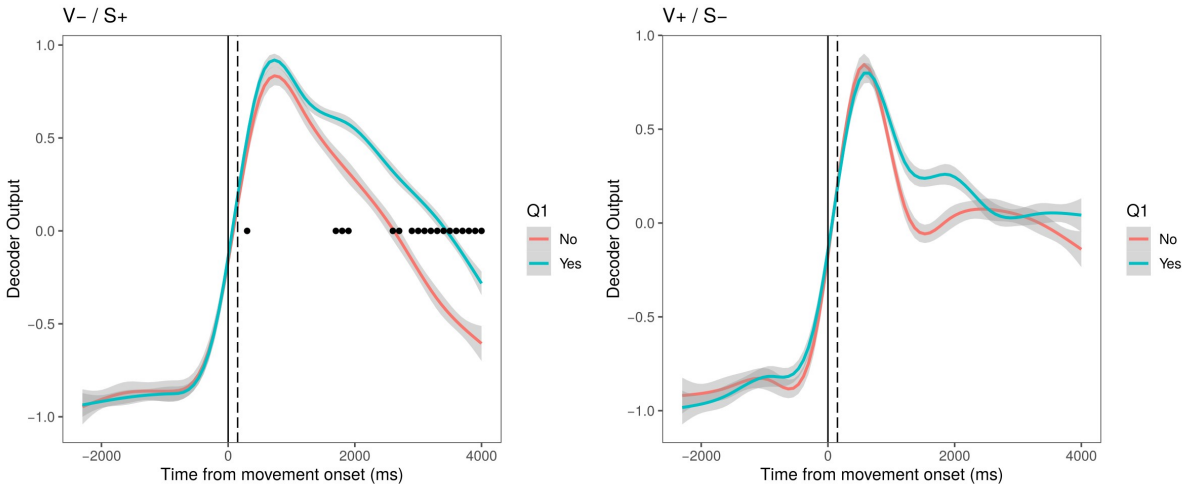


Figure 7. BMI accuracy in time as a function of sense of agency. Blue/red curves represent the BMI classifier output for the cued movement as a function of agency judgements (Q1: 1=high agency; 0=low agency) in conditions of equal sensory feedback.

Discussion

By combining techniques from neurophysiology, neuroengineering, and VR with psychophysics of agency, we were able to study for the first time the sense of agency for actions enabled by a BMI-based neuroprosthesis and found that congruent sensory feedback boosted agency and confidence when controlling BMI actions. Moreover, we showed that human M1 processes not only motor and sensory information, but also different levels of congruency between sensory and motor signals and the resulting sense of agency. The present data are also of clinical relevance, because our NMES-based BMI approach, by providing congruent somatosensory feedback (without direct S1 stimulation) to a tetraplegic patient, improved the ability of the BMI classifier in decoding the patient's motor commands. Interestingly, such higher BMI proficiency was associated with a stronger sense of agency, suggesting that, beyond supporting close-loop systems and M1 feedback in general, somatosensory feedback and signals related to subjective aspects of motor control (i.e. agency) are important input for improving BMI proficiency. Quantifying subjective action-related mental states and including controlled motor and sensory feedback may therefore provide new levels of comfort and personalization and should be considered for the design of future BMIs.

The present data demonstrate that M1 activity contains information specifically linked to subjective aspects of motor control, in particular the sense of agency and confidence that our participant associated with his BMI actions. It is known that agency likely involves a network of multiple brain areas from which we did not record in the present study (e.g., posterior parietal cortex[33] and angular gyrus; anterior insula[34,35]; supplementary motor cortex[36]; premotor cortex[37]; for reviews see [3,38]). However, our findings – even if coming from a single tetraplegic patient - directly demonstrate that M1 activity contains sufficient information to decode actions for which a human participant feels to be in control.

The present BMI findings extend previous research that investigated the sense of agency for non-invasive BCI, as based on scalp electroencephalography (EEG). They add important new information about the underlying neural underpinnings based on M1 multiunit activity of the sense of agency in humans. In line with a prominent line of research on the role of visuo-motor (and visuo-tactile) cues in boosting or modulating

body ownership for artificial and real limbs[39–41], previous BCI studies demonstrated that coherent visual feedback results in higher sense of agency for BCI actions[14]. This effect is associated with stronger activations in a cortical-subcortical network, recruited during motor imagery used to control the BCI. This network consisted of regions in the posterior parietal cortex, the insula, lateral occipital cortex and the basal ganglia[13]. A recent study[10] further demonstrated that a stronger sense of agency for BCI-mediated actions is associated with stronger activity in sensorimotor areas during motor imagery based BCI (however, not for BCI using different signals), compatible with a link between sensorimotor activity and the associated sense of agency in humans. The present data on the sense of agency when using an intracortical BMI, although from a single, highly proficient BMI user (see below), demonstrate that this relationship can be tracked down even at the level of multi-unit activity from M1 neurons, and it is further associated with higher BMI proficiency.

Moreover, the present findings offer a mechanistic explanation for the relationship between sensorimotor activity, sensory feedback and the resulting sense of agency, by showing that M1 activity before movement execution codes for the intended movement, while activity after movement execution encodes the sensory feedback associated with the implemented movement. By showing that somatosensory feedback in particular affects the performance of the BMI classifier, these analyses further provide novel insights into the sensorimotor mechanisms of BMI proficiency. Note that this last finding was possible only due to the combination of a SCI lesion and an NMES-based BMI, which allowed us manipulating not only visual reproductions of body movements (via VR, as in previous studies), but also physical movements of the real body (via NMES). In order to highlight the dynamic, multiscale brain mechanisms underlying the sense of agency in humans, future studies should combine insights that can be gained from invasive BMI - with ultra-high spatial resolution, but limited coverage in a handful of subjects – and non-invasive BCI – with limited resolution, but recording from the entire brain in larger subject samples.

Finally, our results are important not just for the field of neuroprosthetics and its clinical goals, but also for basic neuroscience as well as current ethical and legal debates about

the subjective sense of agency and responsibility when applying neurotechnology solutions for human repair or enhancement[38,42,43].

Limitations of the study

Because of the uniqueness of the present experimental setup, generalizations from the present findings to the general population should be done carefully. First, we tested a single participant, who is an extremely trained BMI user, who could have thus developed extraordinary capacity of controlling his BMI system. This could have in turn impacted the associated sense of agency and the discovered links with BMI proficiency. Second, in order to enable movements of his upper limb, we used an NMES system that provides a series of somatosensory cues, which are only partially comparable to those associated with natural movements. For example, the intensity and temporal activation of cutaneous stimulation, as well as of motor fibers (antidromic) differs from sensorimotor stimulation during normal movement. We also note that although our participant suffered severe somatosensory loss (following damage at the C5-C6 level), he may have "learned" to associate some of the patterns of cutaneous sensations with the specific type of NMES stimulation used to enable specific movements. Indeed, in some circumstances, beyond the experimental sessions, he was able to identify a type of movement implemented via NMES even without seeing his arm. Finally, given the chronic spinal cord lesion suffered by our participant, we cannot exclude that changes related to sensory or motor plasticity have occurred in M1, S1, or the functional and anatomical connections between the two. In general, there is still no consensus about plasticity following SCI, with some evidence of preserved network organization, and some possible changes in grey matter density[44,45] or activation in the sensorimotor cortices[46]. Although important improvements in upper limb function have been documented in this participant following his extensive usage of the NMES-BMI system and concurrent rehabilitation[47], there are no available data about plasticity in his sensorimotor cortices.

Materials and Methods

Participant

The participant in this study was enrolled in a pilot clinical trial (NCT01997125, Date: November 22, 2013) of a custom neural bridging system (Battelle Memorial Institute) to reanimate paralyzed upper limbs after C4-6 spinal cord injury. The system consisted of a Neuroport data acquisition system (Blackrock Micro, Salt Lake, Utah), custom signal processing and decoding algorithms (Battelle), and a NeuroLife Neuromuscular Stimulation System (Battelle). The trial received investigational device exemption (IDE) approval by the US Food and Drug Administration and Institutional Review Board approval through the Ohio State University (Columbus, Ohio). The study conformed to institutional research requirements for the conduct of human subjects. The site of the experiments was the Ohio State University NeuroRehabLab (Bockbrader, PI) and data was analyzed at Ohio State (Columbus, Ohio) and École polytechnique fédérale de Lausanne (EPFL, Switzerland). The participant provided informed consent at the time of enrollment and also provided written permission for photographs and videos.

The study participant was a 22 year-old male at the time of study enrollment. He had complete C5 ASIA A, non-spastic tetraparesis from cervical spinal cord injury associated with a diving accident 3 years prior. On neurological exam, he had full motor function bilaterally for C5 level muscles (e.g., biceps and shoulder girdle muscles), but no motor function below the C6 level. He had 1/5 strength on the right and 2/5 strength on the left for wrist extension (C6 level) on manual muscle testing. His sensory level was C6 on the left and C5 on the right, although he had sensation for pressure on his right thumb. He had preserved proprioception for shoulder, elbow and wrist joint position, but was at chance level for distinguishing digit joint positions (flexion/neutral/extension) for the thumb and fingers. He had mild finger flexor contractures bilaterally, limiting finger extension at the proximal and distal interphalangeal joints of digits 2-5.

He was implanted with a 4.4 x 4.2mm intracortical silicon Neuroport microelectrode array (Blackrock Microsystems) in the dominant hand/arm area of his motor cortex on 4/22/2014, as previously described[6]. The implant site was determined by preoperative functional neuroimaging obtained while the participant visualized movements of his right

hand and forearm. He began using cortically-controlled transcutaneous neuromuscular electrical stimulation (NMES) on his right forearm on 5/23/14, participating in sessions to practice device use for up to 3.5 hrs/day and 3 days/week. In 7/2015, his practice with the device was reduced to 2 days/week. Data for this study was collected over 13 sessions (45 hours) from 11/16/2016 - 2/20/2017, corresponding to post-implant days 939-1035. One session with visual and NMES feedback was used for practice (5 blocks of 32 trials on post-implant day 939). At the time of data collection, the participant was an expert brain-machine interface (BMI) user with over 800 hours of study participation. Of note, the participant underwent cognitive testing of attention, memory and processing ability (without the BMI) approximately one year after Utah array implantation (January – July, 2015). He scored in the gifted range with superior verbal abilities, attention, and working memory (ranging between 92nd - 99th percentile for his age), and no significant differences between auditory or visual memory. However, his processing speed and performance scores were significantly affected by his upper limb impairment (ranging between 27th – 39th percentile for his age).

Cortical Signal Acquisition And Classification

Neural data (96 channels) were acquired from the left motor cortex Utah array through the Neuroport data acquisition system (Blackrock Micro). Raw data were processed using analog hardware with 0.3Hz 1st order high-pass and 7.5kHz 3rd order Butterworth low-pass filters, then digitized at 30,000 Hz. Data were divided into 100ms bins and passed into Matlab (version 2014b), where signal artifact was removed by blanking over 3.5ms around artifacts (defined as signal amplitude >500μV at the same time on 4 of 12 randomly-selected channels). Signals were decomposed into mean wavelet power (MWP) using the 'db4' wavelet over 100ms[48]. Coefficients within the multiunit frequency bands (234–3,750Hz, coeficients of scales 3, 4, 5, 6) were averaged across the 100ms window and normalized by channel (by subtracting the mean and dividing by the standard deviation of each channel and scale, respectively). Normallized coefficients for each channel were averaged across scales 3-6, creating 96 MWP values (one for each channel) per each 100ms. MWP values were fed as features into a real-time, nonlinear support vector machine (SVM) classifier[49] with five classes (hand open, hand closed,

thumb extension, thumb flexion, and rest). Classifier activation values were computed for each 100ms bin and ranged from -1 to 1. Classifier output represented the movement pattern (hand open, hand closed, thumb extension, thumb flexion) with the highest activation greater than threshold (zero). If no movement classes had activation greater than zero, the classifier was in the "rest" state. If multiple output classes exceeded threshold, only the one with the highest score was used to provide feedback.
Signal quality was stable[50] during the interval of data collection; but represented about a 30% decline in MWP normalized to post-implant 87[51]. (See below for single unit statistics.) Average impedance was approximately 200 kΩ, a decline of 40% of its initial value. Average signal-to-noise was approximately 17.5dB, a decrease of about 10% of its initial value[52]. Most of the decline in signal quality occurred in the first 400 days post-implantation.

Classifier Training And Neurally Controlled Hand Movements

Before each session, the SVM classifier was trained in an adaptive manner over 5 blocks. Each block consisted of 3 repetitions of 4 movements (hand open, hand closed, thumb extension, thumb flexion) presented in a random order. Movements were cued for 3-4s (4-5s inter-cue interval) using a small, animated hand in the corner of the video display. Feedback was given with both NMES and the feedback hand on the video screen. During the first training block, scripted feedback was provided simultaneously with the cued movements. In subsequent blocks, appropriate movements were activated when an output class for a given movement exceeded threshold (>0). Training took approximately 10-15 minutes per session.

Neuromuscular Electrical Stimulation

The NMES system was used to evoke hand and finger movements by stimulating forearm muscles. The system consisted of a multi-channel stimulator and a flexible, 130-electrode, circumferential forearm cuff. Coated copper electrodes with hydrogel interfaces (Axelgaard, Fallbrook, CA) were 12mm in diameter, spaced at regular intervals in an array (22mm longitudinally X 15mm transversely), and delivered current in monophasic, rectangular pulses at 50Hz (pulse width 500μs, amplitude 0-20mA).

Desired hand/finger movements were calibrated at the beginning of each session by determining/confirming the intensity and pattern of electrodes required to stimulate intended movements. This took 5-10 minutes per session.

During the experiment, the participant's view of NMES-evoked movements was obscured from view by the video display. During Experiment 1, non-informative NMES feedback was given (current at an intensity equivalent to what was used for movement calibration patterns, but that did not evoke movement). During Experiments 2 and 3, NMES feedback was provided that evoked hand and finger movements.

Virtual Reality Animation

A non-immersive virtual reality system (i.e. without a head-mounted display or head-tracking) was used to provide visual feedback. This was done in order to adopt a previous setup that the participant was already familiar with to the present experiments and also facilitated the calibration procedure to train the BMI classifier. A physics-based animated hand was used to provide visual feedback of classifier activation. During training, two animated hands were displayed, a small cue hand at the bottom left and a larger centrally-placed feedback hand (Figure 1 main text). During the experiment, the display was oriented over the participant's forearm, a single, centrally-placed feedback hand was displayed to match the size and location of the participant's right hand (the cue hand was not displayed). During Experiments 1 and 3, feedback was provided using the virtual hand. During Experiment 2, non-informative visual feedback was given (the feedback hand remained in a neutral, rest position).

Feedback Congruency

In half of the trials across Experiments, the visual and/or somatosensory feedback was covertly manipulated to be incongruent with the cue. In incongruent trials, when the participant correctly activated the classifier associated with the cue, he received feedback opposite to the cue (i.e., hand closed for "hand open", thumb extension for "thumb flex", etc.). In congruent trials, he received feedback consistent with the cue (i.e., hand open for "hand open", thumb flexion for "thumb flex", etc.).

Agency Assessment

All experimental trials began with a verbal cue (“hand open”, “hand closed”, “thumb extend”, “thumb flex”), followed by a 2 second delay, then a verbal cue (“go”). During the next 4s, the participant was given feedback based on classifier activation levels, and then was told to “stop”). Over the next 5-5.5s, the participant reported whether he felt in control of the movement (“yes” or “no”) and his degree of certainty (0-100). The next trial began at the end of this 5-5.5s interval. There were 32 trials per block in Experiments 1 and 2 and 26 trials per block in Experiment 3.

Trial Selection and Time-locking

To ensure that the participant is succesfully activating the classifier for the cued movement, and the signal can be meaningfully time-locked to movement onset, we applied the following selection criteria on the trials. We consider it as a correct imagined movement when the participant is able to maintain the classifier of the cued movement above the threshold for at least 600 ms (6 classifier output bins). We retain trials in which at least one correct movement happens between the GO cue and 1.5 seconds before the STOP cue. Epochs are then constructed by time-locking every trial with respect to the onset of such imagined movements. In case several correct movements occurred during the same trial, the time-locking is relative to the first movement. Furthermore, we excluded 128 trials from the session on which the participant systematically reported problems with controlling the BMI system and absent subjective agency. Globally, we retained 846 out of 1408 trials (60%).

Note that, since we define the onset as the beginning of the 100 ms bin of neural activity that is fed to the classifier, and around 50 ms are required to compute the output, the corresponding feedback is received about 150 ms after the onset of the imagined movement.

Experiment 1: Agency Assessment with Virtual Hand Feedback and Non-informative NMES

Twelve blocks of 32 trials were collected on post-implant days 953 (4 blocks), 988 (4 blocks), and 1035 (4 blocks). In each trial, the participant received a verbal cue to perform a movement (“hand open”, “hand closed”, “thumb extend”, “thumb flex”). When

a classifier crossed threshold during the 4 second feedback window, feedback was given by showing movement of the virtual hand and by activating non-informative NMES (radial wrist electrode activation that did not elicit movement, did not vary from trial to trial, and that the participant could feel and distinguish from real NMES feedback). Feedback on half of the trials was randomly selected to be incongruent with the cue. His subjective sense of agency and level of certainty were recorded for each trial.

A total of 384 trials were collected across three days. After removing trials where the cued action could not be correctly decoded and the session on post-implant day 1035 (see trial selection paragraph), 83 congruent and 72 incongruent trials remained for behavioral and neural activity analysis.

Experiment 2: Agency Assessment with NMES Feedback and Non-informative Virtual Hand

Twelve blocks of 32 trials were collected on post-implant days 941 (5 blocks), 960 (3 blocks), and 967 (4 blocks). In each trial, the participant received a verbal cue to perform a movement (“hand open”, “hand closed”, “thumb extend”, “thumb flex”). When a classifier crossed threshold during the 4 second feedback window, feedback was given by activating movement of the participant’s hand and wrist through NMES and showing non-informative visual feedback (non-moving hand). The participant could not see his own hand/wrist, but could distinguish his hand state based what the stimulation patterns felt like to him. Feedback on half of the trials was randomly selected to be incongruent with the cue. His subjective sense of agency and level of certainty were recorded for each trial.

A total of 384 trials were collected across three days. After removing trials where the participant did not respond correctly by activating the classifier associated with the cue, 154 congruent and 89 incongruent trials remained for behavioral and neural activity analysis.

Experiment 3: Agency Assessment with Virtual Hand and NMES Feedback

Twenty blocks of 32 trials were collected on post-implant days 993 (3 blocks), 990 (5 blocks), 1007 (4 blocks), 1014 (3 blocks), and 1021 (5 blocks). In each trial, the participant received a verbal cue to perform a movement (“hand open”, “hand closed”,

“thumb extend”, “thumb flex”). When a classifier crossed threshold during the 4 second feedback window, feedback was given by activating movement of the participant’s hand and wrist through NMES and showing movement of the virtual hand. The participant could not see his own hand/wrist, but could distinguish his hand state based what the stimulation patterns felt like to him. Congruency with respect to the cue was manipulated independently in the visual and somatosensory modalities such that 25% of the trials were each: congruent for both visual and NMES feedback, incongruent for both visual and NMES feedback, congruent for visual but incongruent for NMES feedback, congruent for NMES but incongruent for visual feedback. His subjective sense of agency and level of certainty were recorded for each trial.

A total of 520 trials were collected across five days. After removing trials where the participant did not respond correctly by activating the classifier associated with the cue, the number of trials that remained for behavioral and neural activity analysis were: 117 congruent for both visual and NMES feedback, 103 incongruent for both visual and NMES feedback, 101 congruent for visual and incongruent for NMES feedback, and 127 congruent for NMES and incongruent for visual feedback.

Firing Rate Calculation and Single Unit Analyses

Single units were identified through offline data processing. For each block, raw voltage recordings at each channel were processed in a series of steps. First, FES stimulation artifact was removed using a 500μV threshold and 3.5ms artifact removal time window. The removed window was replaced with an interpolated segment to retain temporal information. Then, the raw signal with FES artifact removed was processed with a 300-3000Hz bandpass filter. The filtered data was fed into an automated spike detection and sorting algorithm, wave_clus[53] using the default optimization settings. A threshold was set to four times the standard deviation of the noise and used to detect spike locations. A wavelet decomposition was performed on the spikes to extract features and a superparamagnetic clustering algorithm was used to cluster the spikes into groups, representative of individual single units. The superparamagnetic clustering algorithm was used to eliminate spikes that were considered noise to ensure only single units were analyzed. As spike sorting was not performed before data collection, there was no way to match single units across days. Additionally, the number of single units detected

at a given channel fluctuated between days, possibly due to micro-movement of the array and brain state changes. For this reason, all single units detected at a given channel were considered the same, and pooled at the single channel level as multiunit activity in subsequent analysis.

Offline neural decoding

Sensory feedback congruency and subjective ratings (Q1 and Q2) were decoded offline both from LFPs and from multiunit activity. For LFP analysis, the signal amplitude for each channel was downsampled to 500 Hz, band-passed between 0.1 and 12 Hz with an IIR filter, and smoothed using sliding averaging windows of 250 ms. Following multiunit spike times calculation (see above), multiunit firing rate was estimated at 20 Hz over a 250 ms sliding window.

We fed each channel's signal amplitude (LFP) or firing rate (multiunit) as predictors to a penalized linear decoder based on ridge regressions[54]. A separate model was trained to decode congruency (Q1) or confidence (Q2) on each signal timepoint, with a sampling rate of 20 and 500 Hz for multiunit and LFPs respectively. Decoding performance was evaluated by computing and averaging Cohen's k (logistic regression; Q1) or R2 (linear regression; Q2) values over 10 independent 10-fold cross validation runs. The regression was performed through the "train" function of the R "caret" package[55]. To evaluate the statistical significance of the decoding, we generated a null decoding performance distribution by applying the same decoding methods on the data after randomly shuffling Q1 and Q2 values. 1000 permutations were generated, and the decoding performance was evaluated for each of them. Then, a t-value was assigned to every time-point both in real and permuted data, by comparing its decoding performance to the null distribution of permuted data. Finally, the t-values were used to define significant decoding time windows based on a cluster-based permutation test on each epoch's largest cluster[56]. After checking that the t-value threshold used to define clusters was not significantly affecting the results, its value was set at 2.

Computation of distance between neural activity patterns

Since the neural activity recorded by the microelectrode array can change significantly between experimental sessions (i.e. days of recording) spanned by our analysis, Euclidean distances per each pair of conditions were computed separately within each day of recording and then averaged to obtain the final results. Confidence intervals were obtained through a bootstrapping technique, again applied within sessions. For each session and condition, we extracted n random trials with replacement, where n is the number of trials for that condition/session, and the final Euclidean distance was obtained by averaging across sessions as described above. The procedure was repeated 100 times, and 95 % confidence intervals were obtained as 1.96 times the standard deviation of the surrogated distribution obtained as explained here.

Multidimensional scaling

In order to graphically represent the spatio-temporal patterns of neural activity, we performed a multidimensional scaling (MATLAB function mds) on correlation distances computed between spatio-temporal patterns of neural activity. Also in this case, to avoid including sources of variances due to the change in signal between experimental sessions, the procedure was run within experimental sessions. In order to obtain correlation distances between trials we started by concatenating, for each trial, data from all channels and timepoints within the selected temporal window. Then, we computed the correlation coefficient of the resulting vector with the equivalent vector from all other trials within the same session, and subtraced the obtained values to 1 to obtain values of the correlation distance. The first two dimensions of the multidimensional scaling were then aligned across sessions via the Procrustes analysis (MATLAB function Procrustes), using the means by conditions (combinations of movement/somatosensory feedback) in the first session as a reference.

Supplementary Materials

References

1. Blakemore, S. J., Wolpert, D. M. & Frith, C. D. Abnormalities in the awareness of action. *Trends Cogn. Sci.* **6**, 237–242 (2002).
2. Jeannerod, M. *Motor cognition : what actions tell the self*. (Oxford University Press, 2006).
3. Haggard, P. Sense of agency in the human brain. *Nat. Rev. Neurosci.* **18**, 196–207 (2017).
4. Hochberg, L. R. *et al.* Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* **485**, 372–375 (2012).
5. Collinger, J. L. *et al.* High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* **381**, 557–564 (2013).
6. Bouton, C. E. *et al.* Restoring cortical control of functional movement in a human with quadriplegia. *Nature* **533**, 247–250 (2016).
7. Ajiboye, A. B. *et al.* Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration. *Lancet* **389**, 1821–1830 (2017).
8. Lebedev, M. A. & Nicolelis, M. A. L. Brain-Machine Interfaces: From Basic Science to Neuroprostheses and Neurorehabilitation. *Physiol. Rev.* **97**, 767–837 (2017).
9. Donoghue, J. P. Connecting cortex to machines: recent advances in brain interfaces. *Nat. Neurosci.* **5**, 1085–1088 (2002).
10. Nierula, B. *et al.* Agency and responsibility over virtual movements controlled through different paradigms of brain−computer interface. *J. Physiol.* **0**, 1–16 (2019).
11. Nierula, B. & Sanchez-Vives, M. V. Can BCI Paradigms Induce Feelings of Agency and Responsibility Over Movements? in (2019). doi:10.1007/978-3-030-05668-1_10.
12. Sato, A. & Yasuda, A. Illusion of sense of self-agency: Discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition* **94**, 241–255 (2005).
13. Marchesotti, S. *et al.* Cortical and subcortical mechanisms of brain-machine interfaces. *Hum. Brain Mapp.* **38**, 2971–2989 (2017).
14. Evans, N., Gale, S., Schurger, A. & Blanke, O. Visual Feedback Dominates the Sense of Agency for Brain-Machine Actions. *PLoS One* **10**, e0130019 (2015).
15. Knoblich, G. & Sebanz, N. Agency in the face of error. *Trends Cogn. Sci.* **9**, 259–261 (2005).
16. Tsakiris, M., Prabhu, G. & Haggard, P. Having a body versus moving your body: How agency structures body-ownership. *Conscious. Cogn.* **15**, 423–432 (2006).
17. Scott, S. H., Cluff, T., Lowrey, C. R. & Takei, T. Feedback control during voluntary motor actions. *Curr. Opin. Neurobiol.* **33**, 85–94 (2015).

18. Shokur, S. *et al.* Expanding the primate body schema in sensorimotor cortex by virtual touches of an avatar. *Proc. Natl. Acad. Sci.* **110**, 15121–15126 (2013).

19. Hatsopoulos, N. G. & Suminski, A. J. Sensing with the motor cortex. *Neuron* **72**, 477–487 (2011).

20. Hatsopoulos, N. G., Xu, Q. & Amit, Y. Encoding of movement fragments in the motor cortex. *J. Neurosci.* **27**, 5105–14 (2007).

21. Suminski, A. J., Tkach, D. C., Fagg, A. H. & Hatsopoulos, N. G. Incorporating Feedback from Multiple Sensory Modalities Enhances Brain-Machine Interface Control. *J. Neurosci.* **30**, 16777–16787 (2010).

22. Tkach, D., Reimer, J. & Hatsopoulos, N. G. Congruent activity during action and action observation in motor cortex. *J. Neurosci.* **27**, 13241–50 (2007).

23. Churchland, M. M. & Shenoy, K. V. Temporal Complexity and Heterogeneity of Single-Neuron Activity in Premotor and Motor Cortex. *J. Neurophysiol.* **97**, 4235–4257 (2007).

24. Schwartz, A. B. Movement: How the Brain Communicates with the World. *Cell* **164**, 1122–1135 (2016).

25. O'Doherty, J. E. *et al.* Active tactile exploration using a brain–machine–brain interface. *Nature* **479**, 228–231 (2011).

26. Stepniewska, I., Preuss, T. M. & Kaas, J. H. Architectionis, somatotopic organization, and ipsilateral cortical connections of the primary motor area (M1) of owl monkeys. *J. Comp. Neurol.* **330**, (1993).

27. Eickhoff, S. B. *et al.* Anatomical and functional connectivity of cytoarchitectonic areas within the human parietal operculum. *J. Neurosci.* **30**, (2010).

28. Fetz, E. E., Finocchio, D. V., Baker, M. A. & Soso, M. J. Sensory and motor responses of precentral cortex cells during comparable passive and active joint movements. *J. Neurophysiol.* **43**, (1980).

29. Hochberg, L. R. *et al.* Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* **442**, 164–171 (2006).

30. Tabot, G. A. *et al.* Restoring the sense of touch with a prosthetic hand through a brain interface. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18279–84 (2013).

31. Flesher, S. N. *et al.* Intracortical microstimulation of human somatosensory cortex. *Sci. Transl. Med.* **8**, 361ra141 LP-361ra141 (2016).

32. Bensmaia, S. J. & Miller, L. E. Restoring sensorimotor function through intracortical interfaces: progress and looming challenges. *Nat. Rev. Neurosci.* **15**, 313–325 (2014).

33. Desmurget, M. *et al.* Movement intention after parietal cortex stimulation in humans. *Science* **324**, 811–3 (2009).

34. Farrer, C. & Frith, C. D. Experiencing Oneself vs Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency. *Neuroimage* **15**, 596–603 (2002).

35. Chambon, V., Wenke, D., Fleming, S. M., Prinz, W. & Haggard, P. An online neural substrate for sense of agency. *Cereb. Cortex* **23**, 1031–1037 (2013).

36. Fried, I., Mukamel, R. & Kreiman, G. Internally Generated Preactivation of Single Neurons in Human Medial Frontal Cortex Predicts Volition. *Neuron* **69**, 548–562 (2011).

37. Fornia, L. *et al.* Direct electrical stimulation of the premotor cortex shuts down awareness of voluntary actions. *Nat. Commun.* **11**, 1–11 (2020).

38. Sperduti, M., Delaveau, P., Fossati, P. & Nadel, J. Different brain structures related to self- and external-agency attribution: A brief review and meta-analysis. *Brain Struct. Funct.* **216**, 151–157 (2011).

39. Blanke, O., Slater, M. & Serino, A. Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness. *Neuron* **88**, 145–166 (2015).

40. Makin, T. R., Holmes, N. P. & Ehrsson, H. H. On the other hand: Dummy hands and peripersonal space. *Behav. Brain Res.* **191**, 1–10 (2008).

41. Rognini, G. *et al.* Multisensory bionic limb to achieve prosthesis embodiment and reduce distorted phantom limb perceptions. *J. Neurol. Neurosurg. Psychiatry* **90**, 833–836 (2019).

42. Yuste, R. *et al.* Four ethical priorities for neurotechnologies and AI. *Nature* **551**, 159–163 (2017).

43. Fried, I., Haggard, P., He, B. J. & Schurger, A. Volition and Action in the Human Brain: Processes, Pathologies, and Reasons. *J. Neurosci.* **37**, 10842–10847 (2017).

44. Wang, W. *et al.* Specific Brain Morphometric Changes in Spinal Cord Injury: A Voxel-Based Meta-Analysis of White and Gray Matter Volume. *J. Neurotrauma* **36**, (2019).

45. Melo, M. C., Macedo, D. R. & Soares, A. B. Divergent Findings in Brain Reorganization After Spinal Cord Injury: A Review. *Journal of Neuroimaging* vol. 30 (2020).

46. Jurkiewicz, M. T., Mikulis, D. J., McIlroy, W. E., Fehlings, M. G. & Verrier, M. C. Sensorimotor cortical plasticity during recovery following spinal cord injury: A longitudinal fMRI study. *Neurorehabil. Neural Repair* **21**, (2007).

47. Bockbrader, M. *et al.* Clinically Significant Gains in Skillful Grasp Coordination by an Individual With Tetraplegia Using an Implanted Brain-Computer Interface With Forearm Transcutaneous Muscle Stimulation. *Arch. Phys. Med. Rehabil.* **100**, (2019).

48. Mallat, S. *A Wavelet Tour of Signal Processing*. *A Wavelet Tour of Signal Processing* (2009). doi:10.1016/B978-0-12-374370-1.X0001-8.

49. Humber, C., Ito, K. & Bouton, C. Nonsmooth Formulation of the Support Vector Machine for a Neural Decoding Problem. *arXiv* (2010).

50. Colachis, S. C. *et al.* Dexterous control of seven functional hand movements using cortically-controlled transcutaneous muscle stimulation in a person with

tetraplegia. *Front. Neurosci.* **12**, 1–14 (2018).

51. Colachis, S. C. I. Optimizing the Brain-Computer Interface for Spinal Cord Injury Rehabilitation. (2018).
52. Zhang, M. *et al.* Extracting wavelet based neural features from human intracortical recordings for neuroprosthetics applications. *Bioelectron. Med.* **4**, 11 (2018).
53. Quiroga, R. Q., Nadasdy, Z. & Ben-Shaul, Y. Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. *Neural Comput.* **16**, 1661–1687 (2004).
54. Hoerl, A. E. & Kennard, R. W. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics* **12**, 55–67 (1970).
55. Kuhn, M. R Package: caret, Ver. 6.0-80. *CRAN* (2018).
56. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).

Acknowledgments

The authors thank Ian for his dedication to the study and insightful conversations.

Funding: AS is supported by the Swiss National Science Foundation (grant (PP00P3_163951 / 1), OB is supported by the Swiss National Science Foundation and the Bertarelli Foundation.

Author contributions AS: Conceptualization, Formal Analysis, Methodology, Writing; *MB:* Methodology, Investigation, Project Administration, Review & Editing; SC: Methodology, Data Curation, Formal analysis, Investigation, Software; MS: Formal analysis, Investigation, Visualization, Review & editing; TB: Data curation, Formal analysis, Software, Visualization, Review & editing; CD, KE: Investigation, Data collection PG: Methodology, Review & Editing; GS: Methodology, Software and Hardware development; NA: Methodology, Review & editing; DF: Investigation, Software and Hardware development, PS: Methodology, Review & editing; NF: Formal analysis, Methodology, Visualization, Review & editing; AR: Funding acquisition, Resources, Supervision, Review & editing; OB: Conceptualization, Funding acquisition, Methodology, Supervision, Writing.

Competing interests: AS, MS, TB, NF, PS, OB: no conflict of interest. CD, KE, PG, GS, NA, DF have patents for the BMI system.

Data and materials availability

Behavioral data and processed data necessary to reproduce the figures in the main text can be found in the OSF repository accessible at:
https://osf.io/7rma5/?view_only=9928bd8e32a748828f7ecfdbeb1f8baa

Supplementary Materials

Sense of Agency for intracortical brain machine interfaces

Andrea Serino[1,2], Marcie Bockbrader*[3], Tommaso Bertoni[1], Sam Colachis[4], Marco Solca[2], Collin Dunlap[3], Kaitie Eipel[3], Patrick Ganzer[4], Nick Annetta[4], Gaurav Sharma[4], Pavo Orepic[2], David Friedenberg[4], Per Sederberg[5], Nathan Faivre[2,6], Ali Rezai**[7], Olaf Blanke**[2,8]*

[1].MySpace Lab, Department of Clinical Neuroscience, University Hospital Lausanne (CHUV), Lausanne, Switzerland; [2]. Laboratory of Cognitive Neuroscience, Brain Mind Institute & Center for Neuroprosthetics, Ecole Polytechnique Fèdèrale de Lausanne (EPFL), Campus Biotech, Geneva, Switzerland; [3].Department of Physical Medicine and Rehabilitation, The Ohio State University, Columbus, Ohio, US: [4] Medical Devices and Neuromodulation, Battelle Memorial Institute, Columbus, Ohio, US; [5] Department of Psychology, University of Virginia, Charlottesville, Virginia, US. [6]. Laboratoire de Psychologie et Neurocognition, Universitè Grenoble Alpes, Grenoble, France. [7]. Rockefeller Neuroscience Institute, West Virginia University, Morgantown, West Virginia, US. [8]. Department of Neurology, University Hospital, Geneva, Switzerland.

* Equal first author contribution; ** equal last author contribution

Correspondence to: andrea.serino@unil.ch

This file includes:

Supplementary Text

Whole Sample Results

To ensure that our trial selection criteria is not biasing the behavioral results on reported agency and confidence, we run the same analyses as in the main text on the whole sample of 1408 trials. In this sample, we still exclude 128 trials from session of post-implant day 1035 (see trial-selection paragraph). All results go in the same direction as in the main text, therefore we succinctly report them without further discussion.

In Experiment 1, Q1 and Q2 are significantly higher in the congruent condition $p < 0.001$ and $p=0.003$ respectively. In Experiment 2, Q1 and Q2 are significantly higher in the congruent condition $p < 0.001$ and $p<0.001$ respectively. In experiment 3, when contrasting Vis. congruent/NMES incongruent and Vis. incongruent/NMES congruent, we found Q1 and Q2 to be higher in the Vis. incongruent/NMES congruent condition ($P=0.009$ and $p=0.02$ respectively).

Sensory congruency and agency at the level of single channels.

We analysed whether sensory feedback and agency was more specifically processed in any of the 96 channels from the M1 implant. To this aim, we firstly identified the channels where the decoder's coefficients for MU activity more strongly and significantly contributed to the decoding of the of visual and somatosensory congruency (see Figure S1). Two of the 6 significant channels for visual and somatosensory feedback overlapped (channels 80 and 76 in figure S1). Sense of agency was more strongly decoded from 7 channels, two of them overlapping with both visual and somatosensory congruency decoding (channels 80 and 76), and 3 others with somatosensory congruency decoding only, confirming the stronger interdependency between somatosensory signals of agency judgments. All these electrodes were mainly located in the rostral part of the array. Interestingly, the electrodes more strongly decoding sensory congruency and agency were clearly dissociated from those more strongly decoding for the intended BMI action, since the spatial distributions of the decoder coefficients for the type of intended movements highlighted significant electrodes in the caudal part of array, not overlapping with sensory congruency nor agency electrodes (e.g., channels 17 and 67 in figure S1). Thus, M1 activity, also at the single channels level, codes not only for type of movement, but also for the congruency between selected movement and sensory feedback, and the associated sense of agency. Despite stronger contribution from specific electrodes, additional analysis suggests that both sensory congruency and agency are more likely to be encoded at the population level as the power of the decoder in classifying congruent vs. incongruent movements or high vs. low agency actions was higher at the population level than at any of the best 20 channels (see Figure S2).

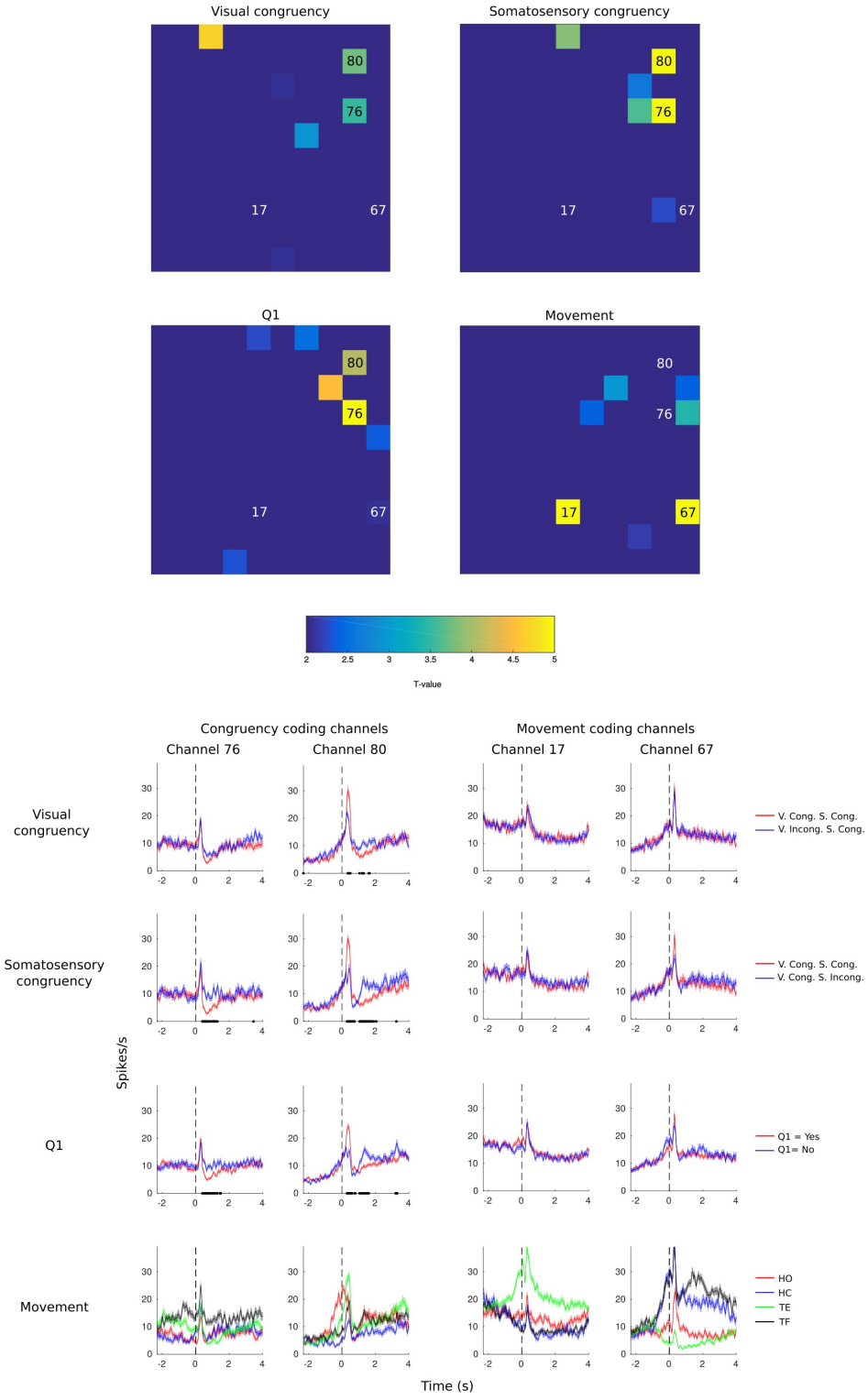

Visual congruency
Somatosensory congruency
Q1
Movement
80
76
17
67
T-value
Congruency coding channels
Movement coding channels
Channel 76
Channel 80
Channel 17
Channel 67
Visual congruency
Somatosensory congruency
Q1
Movement
Spikes/s
Time (s)
V. Cong. S. Cong.
V. Incong. S. Cong.
V. Cong. S. Incong.
Q1 = Yes
Q1= No
HO
HC
TE
TF

Fig. S1.

Contribution of individual channels to neural decoding. In the top panel we show a map of individual channels that contributed most to decoding visual congruency, somatosensory congruency, agency and movement type (going left to right and then top to bottom). To isolate the role of visual and somatosensory modalities in feedback congruency, while minimizing the signal variability between sessions, we trained another ridge regression based decoder on experiment 3, so that we can compare the two modalities in the same set of trials. For visual congruency, we trained the decoder on trials with congruent somatosensory feedback, by contrasting congruent and incongruent visual feedback. For visual congruency, we contrasted congruent and incongruent somatosensory feedback in trials with congruent visual feedback. For movement, to maintain a 2 class decoding schema, we contrasted extension movement (hand open, thumb extension) and flexion movements (hand close, thumb flexion), in trials with congruent visual and somatosensory feedback. To evaluate the contribution of each channel we compared its coefficient in the Ridge regression in a 1 second window starting at movement onset (where decoding of all features is significant), with the distribution of coefficients over all the 96 channels on a 1 second window preceding movement onset, used as a null distribution. T-values are extracted and thresholded at 2. Then, their absolute value is color-coded and displayed on the array grid. Note that this method aims at setting a cut-off on each channel's contribution to the neural decoding, for easier visualization, not at providing a statistically rigorous estimate of decoding significance. In the lower panel we show exemplary channels' response to different conditions. In the two columns on the left we show the response of "congruency coding" channels 76 and 80. In the two columns on the right the same is done for "movement coding" channels 17 and 67. Going from the top to the bottom row, we contrast visual and somatosensory congruencies, positive and negative agency ratings, and the four different movements. After movement onset, the two congruency coding channels clearly differentiate feedback congruency and agency, but show no big difference with respect to the movement. Conversely, two movement coding channels show large differences with respect to the movement even prior to movement onset, suggesting motor intention coding, but no modulation from feedback congruency or agency. Shaded areas indicate standard errors, and black dots indicate significant differences after FDR correction (only where two conditions are contrasted).

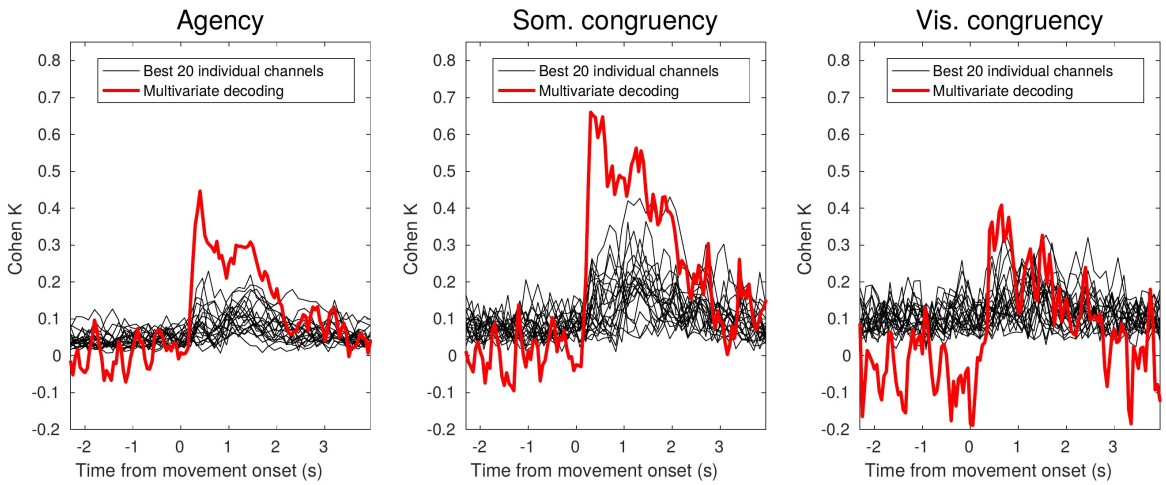


Fig. S2.

Agency and congruency decoding. Comparison between multivariate and single-channel decoding, for Agency, somatosensory and visual congruency (from left to right). Red lines represent cross-validated Cohen's K values for the multivariate Ridge regression presented in the main text. Black lines represent Cohen's K of univaried decoding based on the 20 channels giving the highest mean K value. Note that, to be more conservative, the single channel decoding is not cross-validated, and therefore its performance is slightly overestimated. For the same reason, chance level is higher than 0 and pre-movement decoding is slightly above 0 in the univariate case. Nevertheless, multivariate decoding is greatly outperforming univariate decoding in the case of agency and somatosensory congruency, and only slightly better in the case of visual congruency.

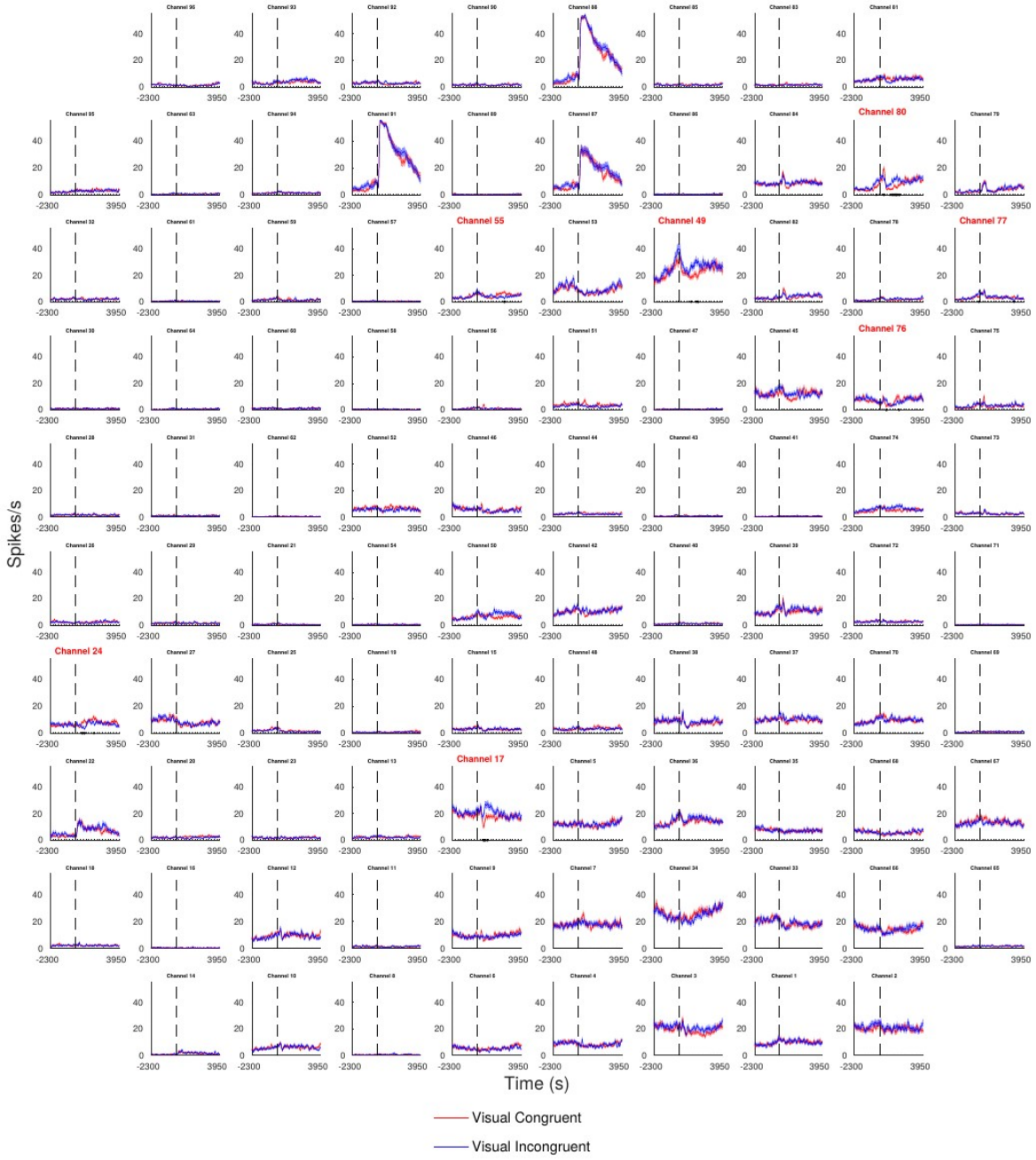


Fig. S3.

Visual feedback in individual channels. Time-locked multiunit response of individual channels contrasting visual congruent and visual incongruent feedback in Experiment 1. The black lines indicate significantly different responses between the two conditions (FDR corrected across timepoints), and subplots with red titles indicate channels with at least one significant timepoint.

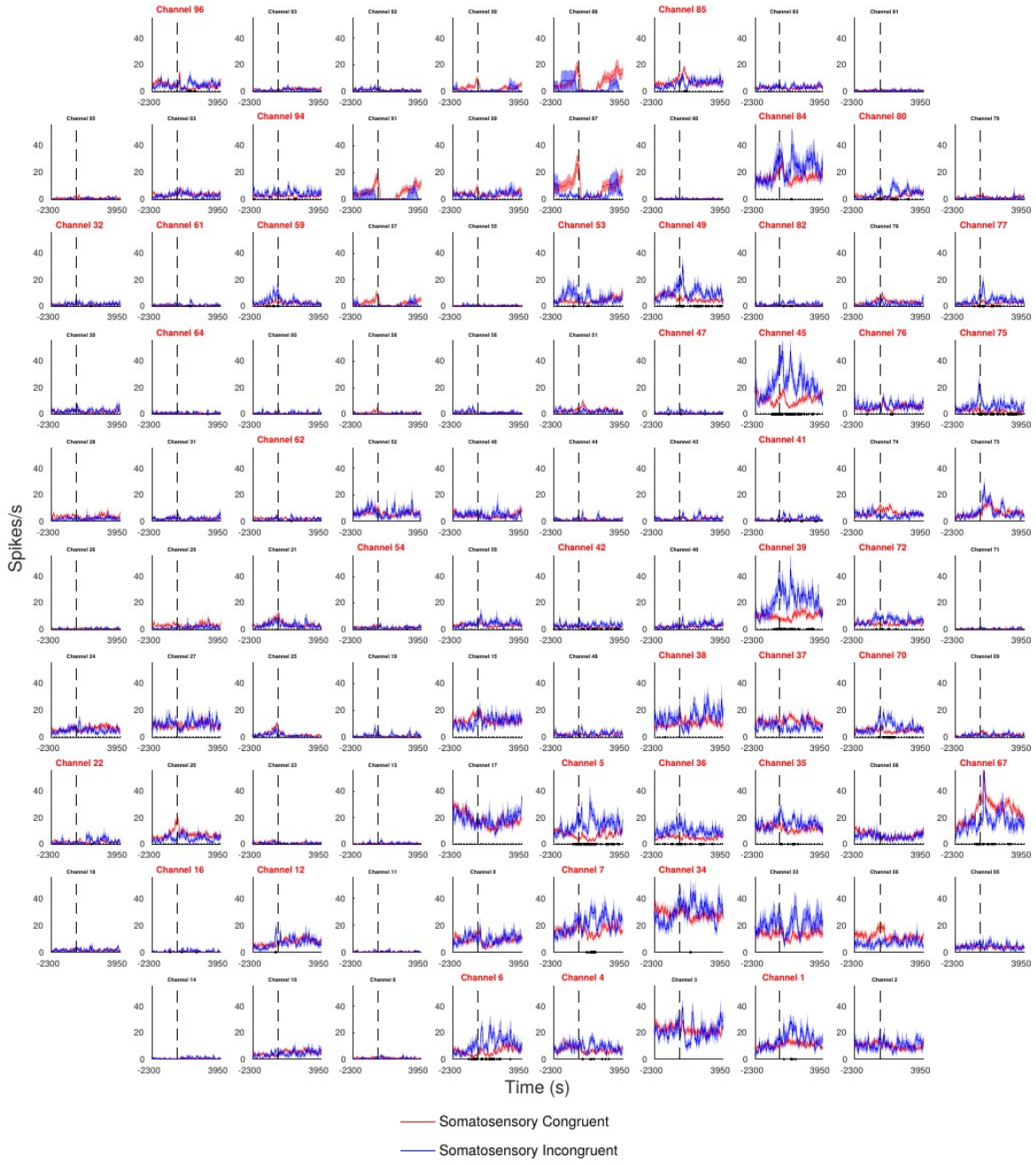


Fig. S4
Somatosensory feedback in individual channels. Time-locked multiunit response of individual channels contrasting somatosensory congruent and somatosensory incongruent feedback in Experiment 2. The black lines indicate significantly different responses between the two conditions (FDR corrected across timepoints), and subplots with red titles indicate channels with at least one significant timepoint.

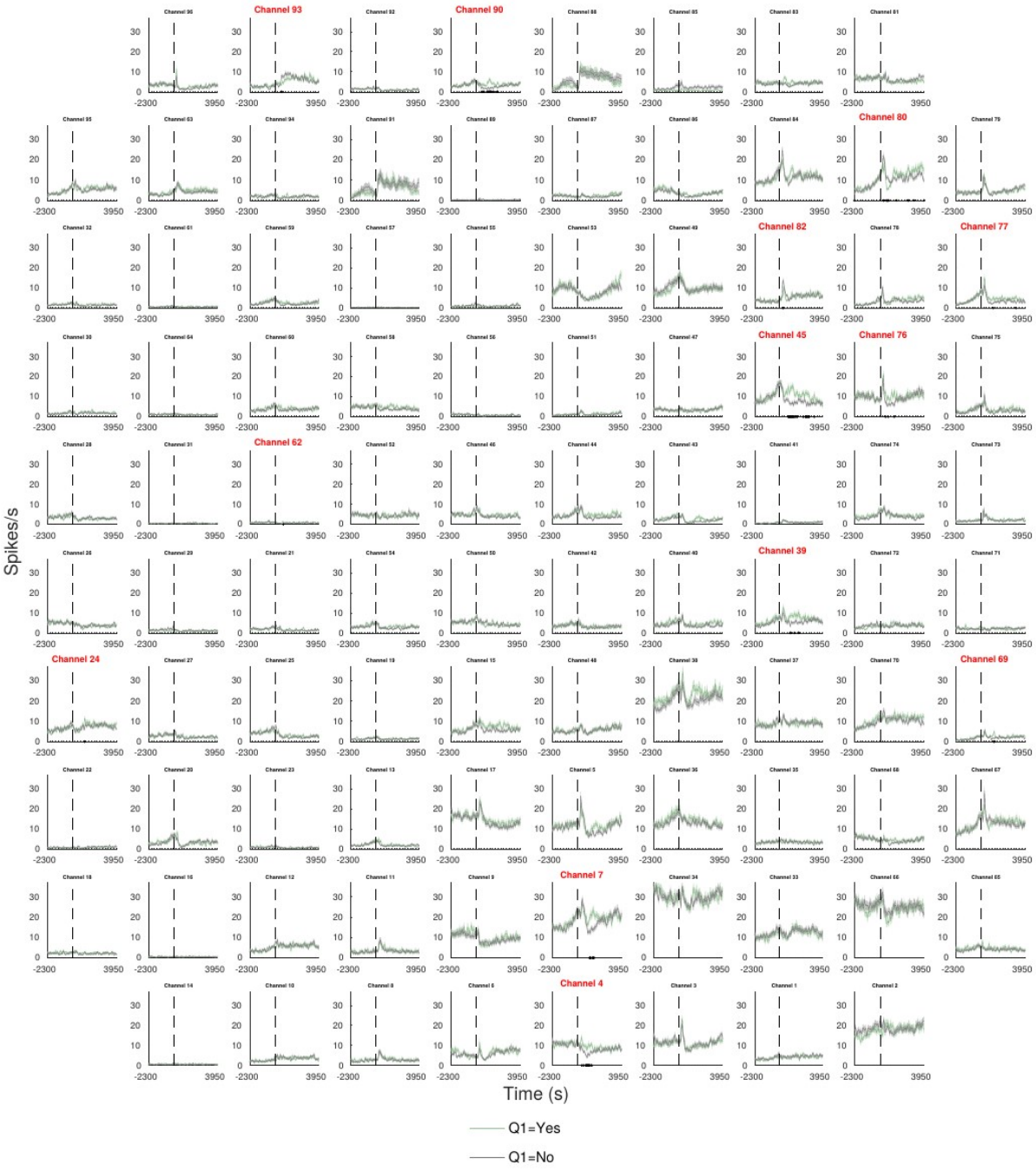


Fig. S5
Sense of agency in individual channels. Time-locked multiunit response of individual channels contrasting positive and negative sense of agency in all experiments. The black lines indicate significantly different responses between the two conditions (FDR corrected across timepoints), and subplots with red titles indicate channels with at least one significant timepoint.

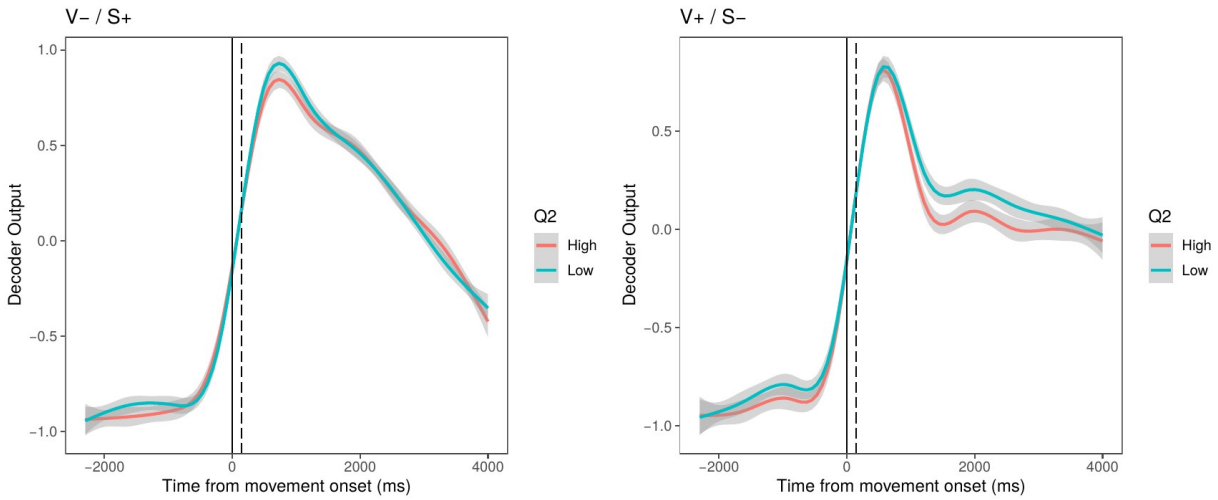


Fig. S6. Effect of confidence on BMI accuracy at fixed feedback. Output of the BMI classifier as a function of Q2 ratings.at fixed sensory feedback (left, V-/S+; right. V+/S-). Blue/red curves represent average values of the BMI classifier output for the cued movement when Q2 ratings were lower/higher than the median rating.

	Experiment 1			Experiment 2			Experiment 3		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
Intercept	0.96	-4.81 – 6.73	0.744	1.05	-3.07 – 5.17	0.618	6.75	1.84 – 11.66	**0.007**
Confidence	9.84	3.36 – 16.32	**0.003**	7.77	2.77 – 12.78	**0.003**	5.59	1.16 – 10.03	**0.014**
Q1=Yes	1.09	-3.23 – 5.42	0.621	1.24	-2.42 – 4.91	0.508	2.42	0.50 – 4.33	**0.014**
Visual Congruent	-1.58	-5.91 – 2.75	0.476						
Somatosensory Congruent				6.53	2.83 – 10.22	**0.001**			
Visual Congruent / Somatosensory Incongruent							-1.45	-4.02 – 1.12	0.271
Visual Incongruent / Somatosensory Congruent							2.81	0.55 – 5.06	**0.015**
Visual Incongruent / Somatosensory Incongruent							-2.89	-5.62 – -0.15	**0.039**
Cue HC	1.89	-0.95 – 4.72	0.194	1.42	-0.87 – 3.70	0.226	2.81	0.68 – 4.95	**0.010**
Cue TE	12.92	10.20 – 15.64	**<0.001**	3.39	1.13 – 5.66	**0.004**	6.20	4.25 – 8.16	**<0.001**
Cue TF	1.32	-1.89 – 4.54	0.421	6.33	3.98 – 8.67	**<0.001**	7.41	5.43 – 9.38	**<0.001**
Trial number	0.03	-0.08 – 0.13	0.617	-0.05	-0.13 – 0.03	0.204	-0.02	-0.09 – 0.05	0.600
Observations	155			243			448		
R^2 / adjusted	0.500 / 0.476			0.394 / 0.376			0.241 / 0.225		

Supplementary Table 1. Sense of agency covaries with the BMI classifier. Multiple regression coefficients predicting agency scores, while regressing out the effects of sensory feedback and movement type. In order to test the role of agency on BMI performance, 140while controlling for other potential factors, we modelled classifier performance based on a multiple regression including agency, confidence, feedback type, feedback congruency, and movement type as regressors (Table 1 and see supplemental information). The results show that for all the three experiments confidence covaried significantly with the performance of the classifier (p<.01; <.01; and <.05, respectively), even when the 145variability explained by the other factors was taken into account. As expected from the previous analyses, the congruency of the somatosensory (p<.001), but not of the visual (p=.47), feedback predicted the classifier's accuracy. Classifier performance also varied as a function of movement type (all p-values<.01). These findings show that movements with higher sense of agency and confidence are associated with higher BMI proficiency, 150suggesting that subjective feelings associated to the control of a BMI-based neuroprosthesis is an important element to take into account to improve their effectiveness

The phase of pre-movement mu oscillations predicts sense of agency for an intracortical brain machine interface

Tommaso Bertoni[1], Marcia Bockbrader[2], Sam Colachis[3], Marco Solca[3], Jean Paul Noel[4], Nathan Faivre[5], Ali Rezai[2], Olaf Blanke[3], Andrea Serino[1]

[1]*MySpace Lab, Department of Clinical Neuroscience, Centre Hospitalier Universitaire Vaudois (CHUV), University of Lausanne, Lausanne, Switzerland*

[2]*Center for Neuromodulation, The Ohio State University, Columbus, Ohio, USA*

[3]*Medical Devices* and *Neuromodulation, Battelle Memorial Institute, Columbus, Ohio, US;*

[4]*Center for Neural Science, New York University, New York, NY 10003;*

[5]*Laboratoire de Psychologie et Neurocognition, Universitè Grenoble Alpes, Grenoble, France.*

[6]*Laboratory of Cognitive Neuroscience (LNCO), Brain Mind Institute, Faculty of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

Abstract

The sense of agency, the subjective experience of self-generation that arises when our actions match our intentions, is a crucial component of self-awareness, and of the development of causal reasoning. We investigated its neural bases in a tetraplegic individual who is a proficient user of an intracranial brain machine interface. His motor commands were recorded from the primary motor cortex, decoded, and translated into functional hand movements through a neuromuscular electrical stimulation system. In a first experiment, we coupled neuromuscular stimulation with virtual reality, in order to provide both somatosensory and visual feedback. We manipulated the congruency of sensory feedback with motor commands, and assessed the participant's sense of agency via explicit judgements. In the second experiment, we developed an adapted version of the Libet task to implicitly measure agency. Sense of agency is the result of the integration of afferent and efferent information over a broad network of brain regions. Due to the importance of slow neural oscillations for long-range communication in the brain, we investigated their effect on sense of agency in our setup. We found that the phase (and not the power) of pre-movement mu oscillations (7.5 Hz) consistently predicted sense of agency both for explicit judgements and implicit measures. The optimal phase for agency was compatible with a facilitation window for spiking activity in the motor cortex, but did not affect patterns of population activity. For the first time, we identify a predictive, endogenous neural marker of sense of agency, potentially connected to general oscillatory mechanisms of information integration in the brain.

Introduction

The sense of agency is the subjective, pre-reflexive feeling of causing and controlling our actions, and the consequent events in the external world (Gallagher, 2000). In humans, agency plays a central role in the processes of self-recognition and self-identification, grounding the awareness of the self as an independent agent in the world (Blanke and Metzinger, 2009; Gallagher, 2000; Synofzik et al., 2008). Influential theories posit that the sense of agency arises from the congruency between efferent motor commands and afferent sensory feedback. Such congruency may be detected by comparing sensory information with predictions based on intended actions and efference copies (i.e., predictive evaluation, Frith et al., 2000), or by the post-hoc comparison between intended actions and their observed sensory consequences (i.e., postdictive evaluation, Wegner, 2002). In order to perform such comparisons, the brain must integrate sensory and motor information at a large scale, gating the information flow between different functional areas with precise timing. Neural oscillations have a key role in orchestrating long-range information integration (Fries, 2015, 2005), with a crucial role of the pre-stimulus phase at low frequencies in determining subsequent perception (Ai and Ro, 2014; Busch et al., 2009), behavioural responses (Landau and Fries, 2012) and neural connectivity (Hanslmayr et al., 2013). Here, we hypothesize that neural oscillations are also involved in the comparisons between motor prediction and sensory feedback underlying sense of agency. In order to address this question, we investigated agency in a tetraplegic individual who is a proficient user of an experimental brain machine interface (BMI) for upper limb control. Motor commands are decoded from the primary motor cortex (M1) based on signals from a chronically implanted microelectrode array, and translated to functional hand movements through a neuromuscular electrical stimulation system (NMES). Within this experimental setting, we sought to uncover the role of neural oscillations in sense of agency in two ways. First, in Experiment 1, the BMI-NMES system was coupled with virtual reality based visual feedback, in order to experimentally manipulate the congruency between motor commands and both somatosensory and visual feedback. After the participant executed a given motor command (e.g., hand close), decoded by the BMI, either congruent (hand close) or incongruent (hand open) feedback was provided via NMES and virtual reality, and he was asked to explicitly rate his sense of agency for that specific action. In the second experiment, we realized a BMI version of the Libet experiment (Libet et al., 1983) to gather an implicit measure of sense of agency. The participant was asked either to perform a BMI hand movement at his will, or his hand was passively moved by the NMES system. In a Libet-like setting, he was then asked to report the perceived timing of his hand movement when this was self-generated (high agency condition) or passively induced (low agency control condition). In line with the Intentional binding effect (Haggard et al., 2002), self-initiated movements were perceived to happen earlier (i.e., closer to the time of intention) than passive movements, and this effect was taken as an indirect behavioural index of sense of agency. We studied the role of neural oscillations in local field potentials for sense of agency in both experiments, in the time window of 1 second preceding movement execution. Specifically, we tested whether movements associated to high vs.

low sense of agency – based on explicit ratings (experiment 1) or implicit temporal binding (experiment 2) - were differentiated by the phase of LFP oscillations in M1 activity in a given frequency band. We observed phase opposition in the mu range (7.5 Hz) differentiating between high vs. low agency trials up to 640 ms before movement onset. Furthermore, we analysed the relation between the mu phase and multiunit activity, and found that the optimal phase for sense of agency coincided with the phase where spikes are more likely to happen. This finding suggests a link between pre-movement endogenous oscillations in M1, neural activity and the subsequent sense of agency for that movement.

Results

Experiment 1 – The phase of mu neural oscillations correlates with explicit agency ratings

The participant is a tetraplegic individual who is an experienced user of a BMI neuroprosthesis restoring his control of right hand movements. A chronically implanted 96 channels microelectrode array reads neural activity from a region of M1 controlling the right hand, and motor commands are decoded from multiunit activity by a nonlinear support vector machine (Fig. 1a, 1b). The neural decoding system is coupled with a custom built high-resolution neuromuscular electrical stimulation system (NMES, Fig. 1c) able to implement the decoded action, thus allowing the participant to control his right arm and perform dextrous manual tasks (Fig. 1).

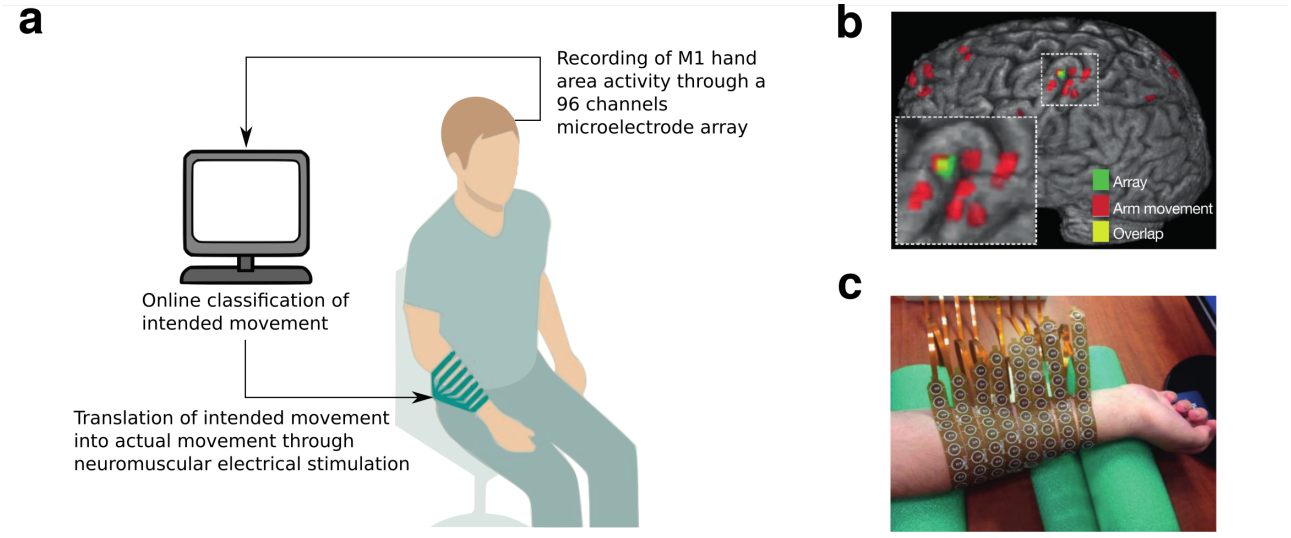


Figure 1: *BMI setup. (a) Neural activity is recorded from the region controlling hand movements in the participant's left motor cortex through a 96 channels Utah array. Mean wavelet power in the multiunit range (234-3750 Hz) is extracted for each channel and fed to a nonlinear support vector machine to decode motor intention every 100 ms. When the neural decoder reaches the threshold of 0 (on a -1/1) scale for a given movement, the selected movement is executed through a custom*

NMES sleeve (panel c). (b) fMRI scan showing areas coding for hand movement (red), array position (green) and their overlap (yellow). (c) Custom NMES system fitted on the participant's hand.

In Experiment 1, the participant was cued to execute one of 4 possible hand movements (hand closing/opening, thumb flexion/extension) through the BMI prosthesis. When the neural classifier's threshold for a given movement was reached, the neuromuscular stimulator generated a hand movement, providing somatosensory feedback. Visual feedback was provided by displaying a virtual hand, while the participant's real hand was hidden from view (Fig. 2a). The congruency of sensory feedback was manipulated to alter the participant's sense of agency, by presenting either the cued and correctly decoded action (congruent feedback), or the opposite action (incongruent feedback, i.e.: flexion instead of extension, and vice versa). In different conditions, all the possible four combinations of congruent and incongruent visual and somatosensory feedback were provided, in randomized order, with 160 trials for each condition over 5 experimental sessions (visual and somatosensory congruent: V+/S+; visual and somatosensory incongruent: V-/S-; visual congruent, somatosensory incongruent: V+/S-; visual incongruent, somatosensory congruent: V-/S+). At the end of each trial, the subject was asked about his feeling of agency for the performed movement (Q1: "Was it you who generated the movement?") and his confidence about his answer ("Q2: how sure are you about your answer?"). Experiment 1 is described and included in another study (Serino et al., under consideration). Here we applied a novel approach and run novel analyses to the data.

We started our analysis from one of the main results of our previous study (Serino et al., under consideration) focusing on the effect of sensory feedback on the percentage of trials with a positive agency judgement. Results showed that the participant almost always reported positive or negative sense of agency when visual and somatosensory stimuli were both congruent or both incongruent, respectively, with extremely high confidence ratings (see Fig. 2b-c). Cleary, more variable agency ratings were obtained in the case of conflicting sensory feedback, reflected in significantly lower confidence ratings (see Fig. 2c).

In the full experimental setup, a significant part of the variability in agency ratings is explained by exogenous sensory feedback congruency. Therefore, in order to assess the role of endogenous neural oscillations in the participant's sense of agency, we restricted our analysis to trials with conflicting sensory feedback, where confidence was lower and agency ratings weakly correlated with sensory feedback (McFadden's R^2 in a logistic regression Q1~feedback = 0.02). In other words, by considering only trials in which exogenous factor, such as sensory feedback congruency, were essentially decoupled from sense of agency, we maximise the likelihood of uncovering the role of pre-stimulus endogenous factors, such as neural oscillations.

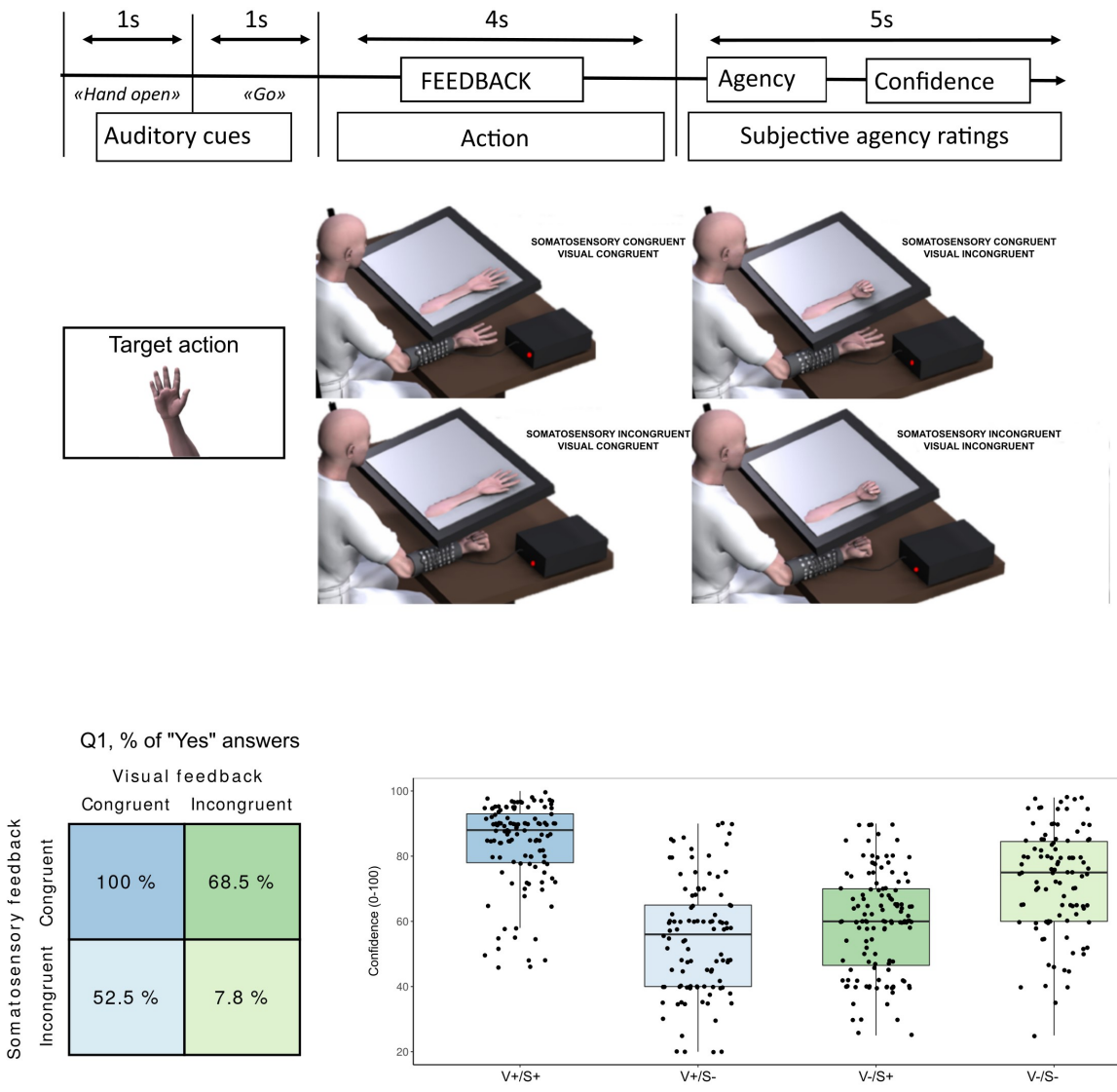


Figure 2: *Experiment 1 timeline and behavioural results. (a) Timeline of the experiment. The participant is cued to perform one of 4 possible hand movements. For 4 seconds after the "go" cue, visual and somatosensory feedback about the movement can be either congruent or incongruent, with the opposite movement executed/displayed in the case of incongruent feedback (hand open for hand close, thumb flexion for thumb extension). (b) Percentage of high agency trials ("yes" answers to Q1) by visual and somatosensory feedback. (c) Confidence ratings on Q1 answers (Q2) by visual and somatosensory feedback.*

Following previous reports on the role of slow oscillations in long-range connectivity for multisensory and sensorimotor integration (Hanslmayr et al., 2013), we analysed oscillations in the 4-16 Hz range in local field potentials recorded from the microelectrode array during the task, and tested their ability to predict the participant's agency ratings. Preliminary analysis showed an extremely high coherence of low frequency oscillations across the whole array, therefore we analysed the average signal from all electrodes, reducing the number of comparisons with negligible information loss (see Fig. S4).

We then extracted the instantaneous phase and power for 10 frequencies in the 4-16 Hz range through a Morelet wavelet transform, and contrasted them between trials with positive and negative agency judgements. To quantify phase opposition between high and low agency trials, we computed the Phase Opposition Product (VanRullen, 2016), measuring the amount of simultaneous clustering of phase angles for high and low agency trials around opposite s. We corrected for multiple comparisons across time and frequency through a cluster-based permutation test with 10000 iterations. We found a significant ($p = .0029$) phase opposition in the mu range, peaking at 7.5 Hz and 210 ms before movement onset, and extending roughly between 570 and 60 ms (Fig. 3a-b) before movement. The histogram of phase angles at the time-frequency point with the highest phase opposition are shown in Fig. 3d. In order to estimate the phase angles at movement onset that elicit the highest and lowest sense of agency, we extrapolated phase angles at movement onset from the time-frequency point with maximal phase opposition (-210 ms, 7.5 Hz). This analysis showed that phase angles at movement onset are clustered slightly after π for high agency trials, and slightly after 0 for low agency trials (Fig. 3e). As an additional control, we explored whether agency ratings depended continuously on phase angles and compared this dependence separately in the V-/S+ and V+/S- trials. We found a similar and continuous relationship between phase and agency in the two conditions, with negative agency judgements becoming increasingly frequent as phase angles at movement onset approached π (Fig. 3f).

Additionally, we investigated whether the power of such oscillations, and not only the phase, modulated sense of agency. First, we noted that the mu rhythm is relatively low in our participant, peaking at 6.2 Hz (see Fig. S1a). We observed the expected movement related de-synchronization (ERD), with a pronounced decrease in power in the mu range starting around 700 ms prior to movement onset (see Fig. S1b). However, the ERD (and power in the 4-16 Hz range in general) was comparable in magnitude in high and low agency trials, with no significant difference in power between the two conditions (see Fig. S1c).

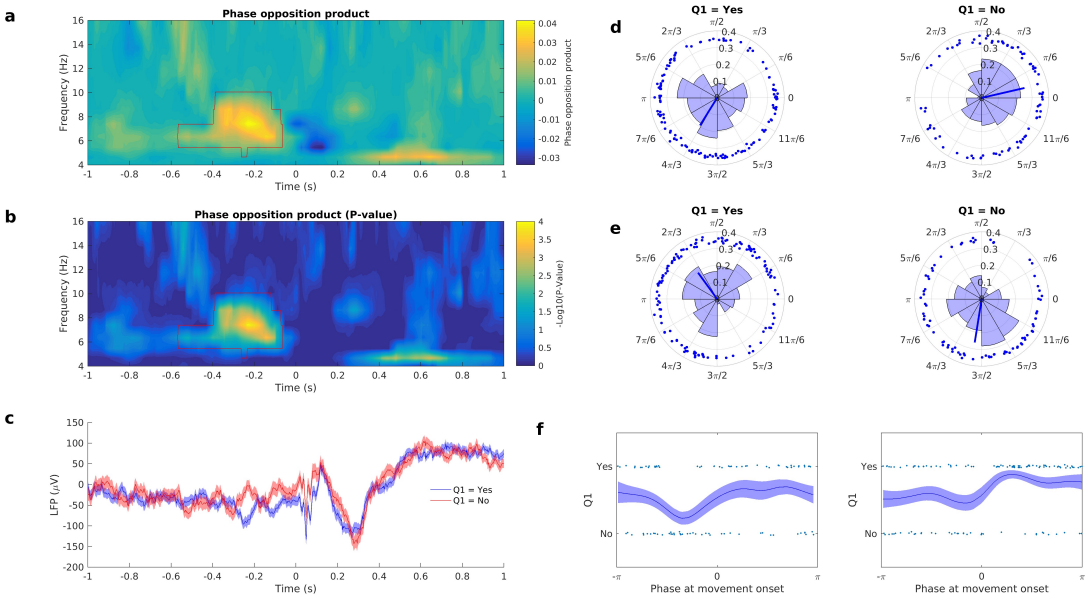


Figure 3: *phase opposition analysis for Experiment 1. (a) Phase opposition product in the time-frequency domain between trials with yes and no answers for Q1. (b) P-values for the phase opposition product. The red contour delimits the significant cluster after a cluster-based permutation test (p = 0.0029). (c) Average LFP for high (red) and low (blue) agency trials. The phase opposition is visible between -400 and 0 ms. Shades denote standard errors. (d) Histograms of phase angles for individual trials at the time-frequency point of maximal phase opposition (-210 ms, 7.5 Hz). High agency trials are displayed on the left, low agency trials on the right. (e) Same histograms, in which the phase angles have been extrapolated at movement onset by transposing phase angles by 210 ms. (f) Dependency of agency judgements (Q1) on phase angles (extrapolated at movement onset) for the V+/S- (left) and V-/S+ conditions separately. The blue line indicates a rolling mean obtained through a Gaussian smoothing with a bandwidth of 0.5 radians, and the shade its standard error.*

Experiment 2 – The phase of mu neural oscillations correlates with an implicit index of sense of agency

To further demonstrate the link between the phase of mu oscillations before the onset of an action and the sense of agency associated to that action, we analysed data from a second experiment, using a BMI version of the Libet experiment (Libet et al., 1983) as an implicit measure of sense of agency. The experiment consisted in two sessions, one with high and one with low sense of agency. In the high agency session, the participant was asked to perform one of two possible movements (hand opening/closing) through the BMI system. During each trial, a rotating clock was displayed on a screen, and the participant was asked to report the position of the clock at the onset of the movement to measure his perceived movement time. In the low agency session, acting as a control, the participant performed the same task, but the two possible movements were randomly generated

via the NMES system. In both sessions, one of the two movements (operant movement) elicited a sound that was produced 300 ms after movement onset, the other (baseline movement) had no additional sensory consequence. This was to investigate intentional binding, which goes beyond the scope of the present study. Therefore, as a first approach, the two movements were analysed together. We compared the perceived time of movement between voluntary, high agency, movements, and involuntary, low agency, movements, with respect to the real time of the movement. Due to the non-normal distribution of responses with frequent outliers, we used the median as a robust indicator of the subject's perceived movement timing, and performed a Wilcoxon rank sum test. The subject was generally biased towards perceiving the movement earlier than in reality, but this bias was significantly stronger for active than for passive movements (median active = -497.8 ms, median passive = -384 ms, p = 0.033). Thus, the presence of intention induced an anticipation in the perceived timing of the movement. We then used this index to differentiate, within active movements, actions with higher sense of agency – i.e., in which the onset of the movement was perceived earlier in time – from trials with lower sense of agency – in which movement onset was perceived later – by performing a median split on the time of perceived action.

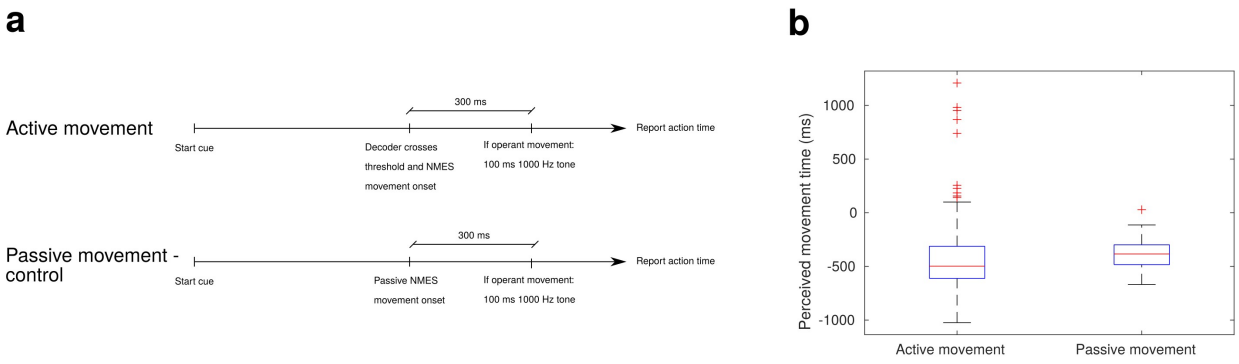


Figure 4: *Timeline and behavioural results of Experiment 2. (a) Timeline. In the active movement condition, the participant spontaneously initiates hand opening or hand closing movements, that is produced upon the decoder crossing threshold through the NMES system. 300 ms after the onset of hand closing movements (operant condition) a "beep" sound is produced, no sound is produced after hand opening movements. In the passive movement condition, no motor intention is formulated and the same movements are generated by activating the NMES system at a random delay from trial onset. After each trial, the participant reports the position of a rotating clock displayed on a screen at movement onset, to indicate his perceived movement timing. (b) Timing of perceived movements (both operant and non-operant condition) in the active (left) and passive (right) movement conditions. Boxes contain the central 75 % of trials, and red lines indicate medians.*

Based on results from Experiment 1, we expect the phase angles of trials with an early perception of the movement to be clustered around π at movement onset, similarly to what observed for higher agency trials in the previous experiment. Conversely, we expect an opposite phase clustering

(around 0) for trials in which the action was perceived late, corresponding to lower agency trials. Indeed, we found a significant phase opposition (p = 0.0096 over 10000 permutations) peaking at 8.5 Hz and -310 ms relative to movement onset, and extending between -640 and -190 ms (Fig. 5a-b). As shown in Fig. 5f, the perceived timing of the action depended smoothly and continuously on the phase at movement onset.

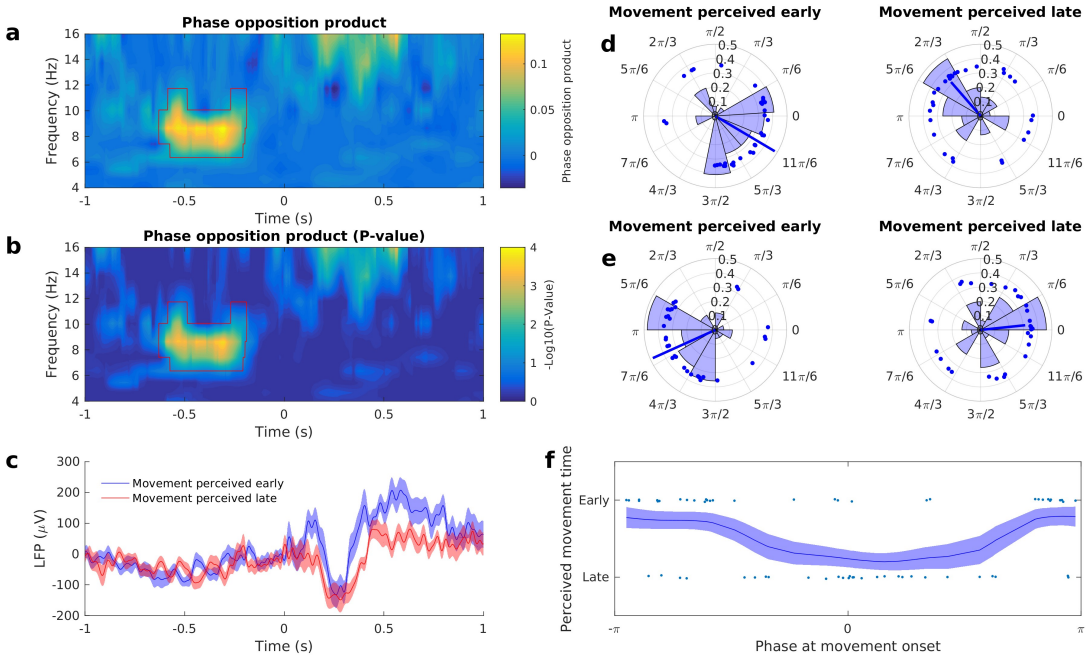


Figure 5: *phase opposition analysis for Experiment 2, in the active movement condition. (a) Time-frequency plot of the phase opposition product between trials with early and late perception of movement onset. (b) P-values for the phase opposition product. The red contour delimits the significant cluster after a cluster-based permutation test (p = 0.0096). (c) Average LFP for trials with early (red) and late (blue) perception of movement, as defined through a median split. Shades denote standard errors. (d) Histograms of phase angles for individual trials at the time-frequency point of maximal phase opposition (-310 ms, 8.5 Hz). Trials in which the movement was perceived early are displayed on the left and vice versa. (e) Same histograms, in which the phase angles have been extrapolated at movement onset by transposing phase angles by 310 ms. (f) Dependency of movement perception on phase angles (extrapolated at movement onset). The blue line indicates a rolling mean obtained through a Gaussian smoothing with a bandwidth of 0.5 radians, and the shade its standard error.*

We then investigated whether the observed effect was related to intentional binding between the movement and its auditory effect (Haggard et al., 2002), by splitting the trials according to whether the performed movement elicited a further sensory consequence (the sound) or not. This analysis

revealed that the observed phase opposition was only significant in trials in which the movement was followed by an effect, with a significant phase opposition in the usual time-frequency region (Fig. S2, left panels). However, a weaker, non-significant phase opposition was observed as well in the second set of trials (Fig. S2, right panels). Moreover, the fact that when grouping the trials the phase opposition remains significant suggests that, besides statistical significance, the phase relation between timing of the perceived movement and neural oscillations is the same in both groups of trials. Generally, low frequency oscillations have been reported to play a role in attention and in temporal perception (Busch et al., 2009; Hanslmayr et al., 2011; Landau and Fries, 2012; VanRullen et al., 2007). Therefore, to rule out attentional or perceptual confounds, we ran the same analysis on the passive control condition (Fig. S3), in which agency levels are constantly very low. We found no modulation of the perceived action timing in the passive condition, suggesting that, indeed, the observed relationship between phase and agency is specific to the binding between intention and action.

Finally, in order to confirm that the relationship between oscillatory phase and sense of agency is the same for explicit and implicit agency measures, we analysed together data from Experiment 1 and Experiment 2. We contrasted all trials with “high agency” (“Yes” answers in Experiment 1 and “early” perceived movements in Experiment 2) with all trials with “low agency” (“No” answers in Experiment 1 and “late” perceived movements in Experiment 2). We found a significant phase opposition ($p = .0033$), again peaking at 8.5 Hz and -250 ms before movement onset (Fig. 6a). As a control, we tested the opposite hypothesis, grouping “Yes” and “late” trials and contrasting them to “No” and “early” trials (Fig. 6d). No significant phase opposition emerged in this case (p of largest cluster 0.07). These results confirm that, when the oscillatory phase before movement onset was close to π, the probability of feeling higher sense of agency for that movement is higher, as measured both in terms of higher explicit agency ratings or early perception of voluntary movements (Fig. 6e, left). Conversely, when the mu oscillatory phase at movement onset is close to 0, actions are less likely attributed to oneself, both explicitly and implicitly (Fig. 6e, right).

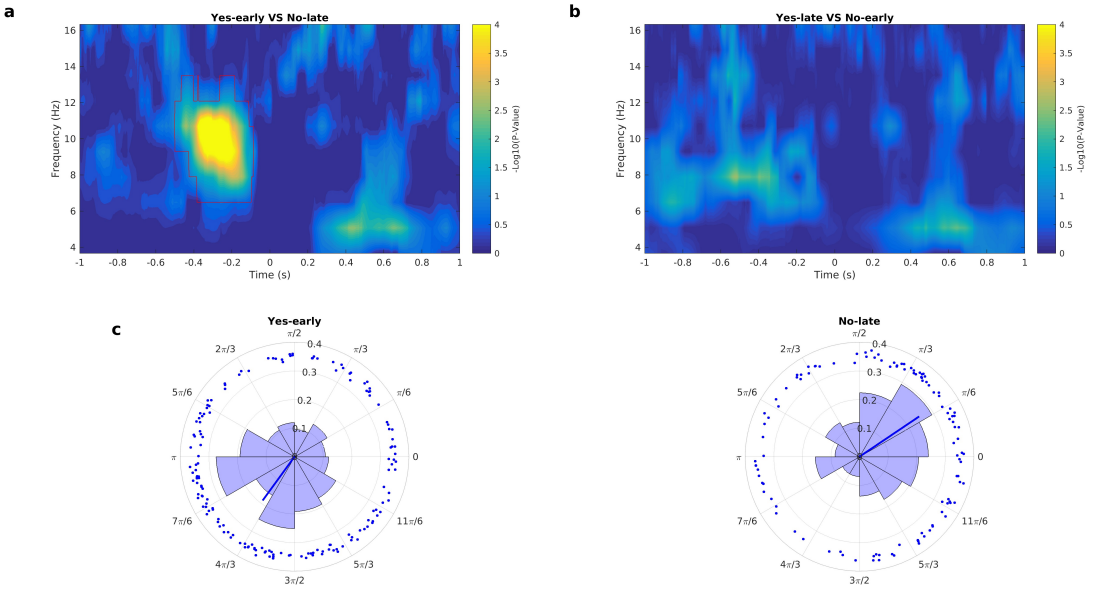


Figure 6: *grouping phase opposition results across Experiments 1 and 2. (a) Phase opposition (Q1="yes" + movement perceived early) vs (Q1="no" + movement perceived late). The red contour indicates a significant cluster (p = 0.0033), and confirms the same phase relation holds for high agency trials and trials with early perception of movement, and vice versa. (b) Control contrast, (Q1="yes" + movement perceived late) vs (Q1="no" + movement perceived early). (c) Phase histograms (extrapolated at movement onset) for (Q1="yes" + movement perceived early), on the left, and (Q1="no" + movement perceived late), on the right.*

The high-agency LFP phase is associated with facilitation of spiking activity, but does not affect population activity

The LFP oscillations that predict sense of agency in our setup might be driven by both local and long-range neural activity. To gather a deeper insight into the relationship between the two, we analysed the coherence between spiking activity in M1 and LFP phase in the mu range. Trials from the two experiments were pooled together. Across 5 experimental sessions, 1458 single units were identified. We quantified the coherence between spikes and LFP phase through the spike phase histogram. For each spike of each unit, the phase vector of the mu LFP oscillation at the time the spike occurred was extracted. For each unit, the length of the vector obtained by summing all the spike-locked phase vectors was used as a quantification of the coherence between spikes and the phase of LFP oscillations. We found 366 units (around 25.1 %) whose spike activity was significantly time-locked to mu LFP oscillations (Fig. 7a). Moreover, the histogram of preferred LFP phase angles across all units shows a clear clustering around 5π/6 (Fig. 7b). This phase angle is remarkably similar to the preferred phase angle extrapolated at movement onset for high agency trials. Therefore, the phase of LFP at which M1 spikes tend to occur is similar to the one that enhances sense of agency

in conflicting feedback sessions. We then asked whether, at the population level, patterns of multiunit activity were affected by the mu phase at movement onset. In order to quantify the impact of the mu phase on spiking activity, we computed the mean Euclidean distance between vectors of multiunit activity across time, between pairs of trials where the movement occurred close to the optimal phase ($\varphi < \pi/2$, $\varphi > 3\pi/2$) or far from it ($\pi/2 < \varphi < 3\pi/2$). We compared such distance to the mean distance between pairs of trials with random phase at movement onset, and found the time course of the distance to be virtually identical in the two conditions (Fig. 7c). A multidimensional scaling of activity in the 0-600 ms range, used to graphically compare activity in the optimal vs non-optimal phase, also shows overlapping results between the two conditions (Fig. 7d). This suggests that, while on average the mu phase has some excitatory effect on spiking activity, it weakly affects population activity in M1.

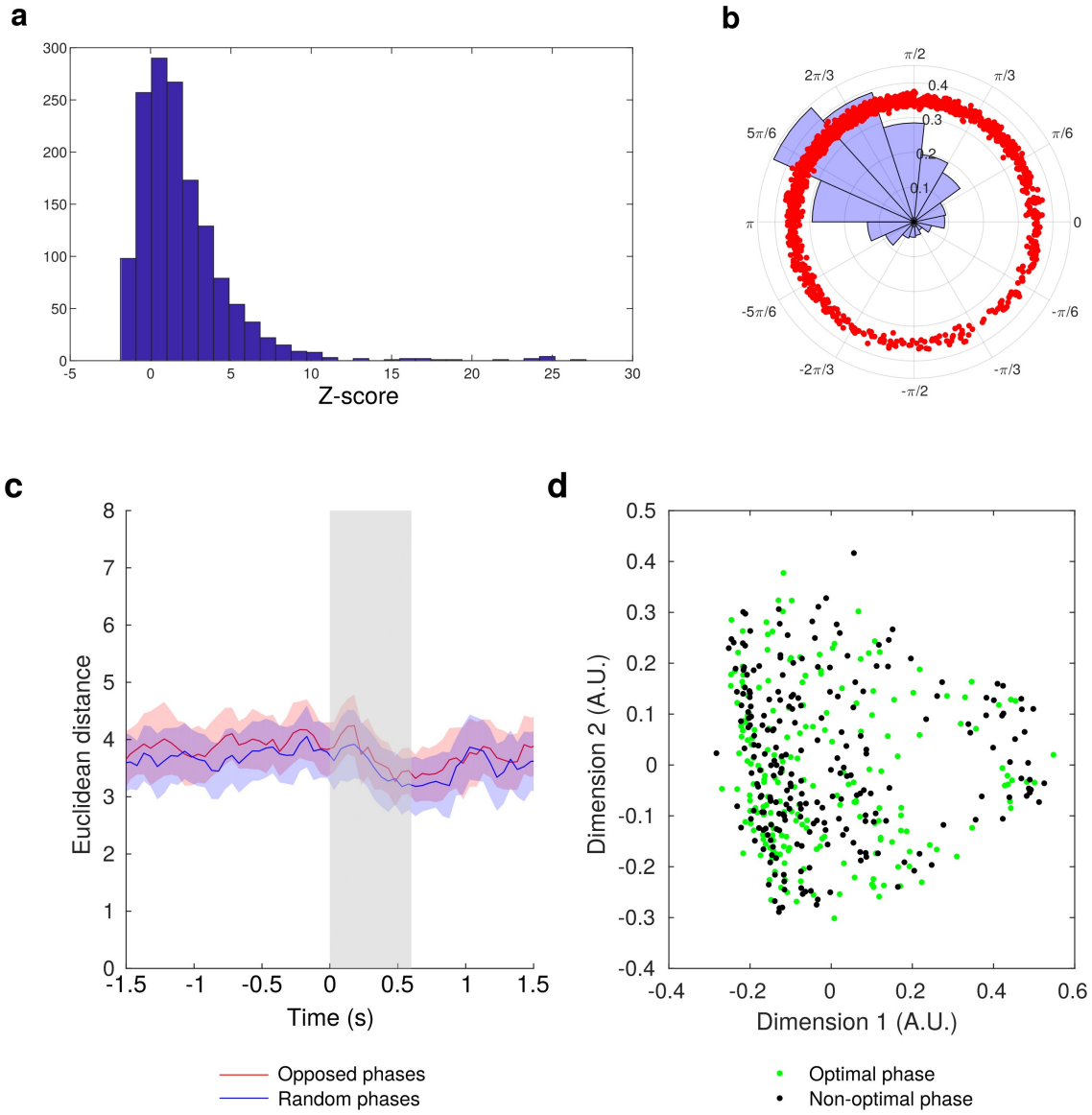


Figure 7: *spike-field coherence (SFC). (a) Distribution of Z-scores for the spike-field coherence at 7.5 Hz across 1458 units. Data from two additional studies was added here to increase statistical*

power. (b) Distribution of the preferred angle of the same 1458 units, defined as the sum of phase vectors of 7.5 Hz oscillations taken at each spike. (c) Average Euclidean distance between pairs of trials with optimal vs non-optimal phase (red), compared to the distance between pairs of trials with random phase relations. The shaded area denotes the time period used in panel d. (d) Multidimensional scaling of neural activity in the 0-600 ms period. The two axes represent the two main components of a dimensionality reduction of 96 channels of multiunit activity. Each dot represents a trial, with green and black trials being those in which the movement occurred respectively close to or far from the optimal phase.

Discussion

We investigated the role of neural oscillations in determining sense of agency for hand movements in a tetraplegic human participant, who is a proficient user of an intracortical brain machine interface. The natural cortico-spinal communication pathway was replaced by the artificial decoding-stimulation system, offering us the unique opportunity to fully control the congruency between desired actions and their implementation in upper limb movements in order to modulate and assess the participant's sense of agency explicitly (Experiment 1) and implicitly (Experiment 2). In Experiment 1, we modulated the visual and somatosensory congruency between intended actions and sensory feedback and asked explicit agency judgements. As expected, the participant's sense of agency was strongly modulated by sensory feedback congruency. To uncover the endogenous contribution of neural oscillations to sense of agency, we then focused on the residual variability in agency judgements that was not explained by sensory feedback. The phase of mu (~8 Hz) LFP oscillations 570 ms to 60 ms before movement onset reliably predicted the participant's agency judgements, with positive answers more likely to occur when the BMI generated movement happened close to π, corresponding to the negative trough of an oscillation. This result was confirmed when agency was implicitly assessed by means of a modified version of the Libet experiment. The participant either freely performed active hand movements, or we implemented the same hand movements passively via the NMES system, and we asked him to report the perceived timing of the movement. Voluntary movements were perceived as occurring significantly earlier than involuntary movements. Thus, we used such binding effect in the temporal perception of actions towards the time of intention as an implicit measure of agency. Within the voluntary condition, we therefore contrasted trials with an early vs late perception of the movement as a proxy of higher vs lower agency. Such implicit index of sense of agency was also modulated by the mu phase at movement onset, meaning that trials with early perception/higher agency were characterized by a phase angle at movement onset around π, whereas late perception/lower agency trials showed the opposite phase clustering. Furthermore, the correlation between phase and perceived movement timing was only observed for active movements, movements, and not when the arm was passively moved, while the participant was asked to perform the same temporal judgement. This suggests that

our results cannot be explained simply in terms of attentional or perceptual effects. Indeed, if the pre-movement phase was only affecting the participant's time perception (either directly, or through attention), we would observe the same effect in the passive condition. Importantly, by combining data from both experiments, we showed that the specific relation between mu phase and sense of agency was the same in Experiment 1 and Experiment 2.

To our knowledge, this study provides the first demonstration of a mechanistic link between endogenous pre-movement neural states and the subsequent subjective experience of agency, likely reflecting the neural signature of predictive components of sense of agency. The famous study by Libet and colleagues (1983) showed that conscious intentions are preceded by a slow negative deflection above motor areas, the so-called readiness potential. More recently, this effect has been interpreted as a sign that endogenous, stochastic fluctuations of neural activity contribute to determining the timing of spontaneous movements (Schurger et al., 2012). Here, we show that an oscillatory component of such endogenous activity also relates to subjective aspects of self-causation for movements. This mechanism likely complements the postdictive computations relying on the comparison between intended and observed actions, which explain the modulation of sense of agency as a function of feedback congruency. Several hypotheses can be formulated on the brain-level mechanism that connect local oscillations in the pre-movement phase with the feeling of agency arising after movement onset. One simple explanation may be that LFP oscillations carry relevant temporal cues about the probability of onset of self-generated movements. This idea would be in line with observations that, in able-bodied humans, spontaneous movements' onsets (Popovych et al., 2016; Tomassini et al., 2017), as well as the excitability of the cortico-spinal tract (Desideri et al., 2019), tend to be phase-locked to mu oscillations. Under this perspective, the phase of mu waves would modulate sense of agency by encoding and signalling to other brain areas the probability for a movement to occur at a given moment. It is worth saying that such property would be hard to assess in healthy subjects, as the natural relation between neural oscillations and movement onsets, and between intended and executed actions, would be extremely difficult to manipulate. Instead, in our setup, the timing and sensory consequences of intended movements are independent from respectively neural oscillations and the intended action. Another alternative, but not necessarily exclusive explanation can be proposed considering the role of neural oscillations in modulating the functional connectivity between distant cortical areas. In our case, the phase of mu oscillations recorded from M1 neurons may modulate the long-range neural connectivity in the wide fronto-parietal networks involved in sensorimotor and cognitive computations which are thought to underlie sense of agency (Chambon et al., 2013; Sperduti et al., 2011). Then, we can hypothesize that when movements occur at the optimal phase angle for connectivity, stronger binding between intentions and sensory events occurs, leading to higher sense of agency. In this respect, it is worth noting that in Experiment 2 we found that mu oscillatory phase affects the perceived timing of movement onset only when movements are preceded by intention. This suggests that the observed role of oscillatory phase is specific for the binding of actions and intentions. Conversely, the set of conditions analysed

in Experiment 1 contained a mixture of trials with either visual or somatosensory feedback being congruent, and the other modality being incongruent, and the same relation between agency and oscillatory phase held for both conditions (V+/S- and V-/S+). We can therefore exclude the possibility that, in our setting, oscillatory phase modulates the relative weight of sensory modalities (as has been noted in other cases, see Thézé et al., 2020), or their overall saliency. Instead, it seems that it relates to the binding of action and sensory feedback independently from its modality, and even when partially incongruent with intentions. In any case, this second hypothesis does not necessarily exclude the first hypothesis about movement timing, but rather encompasses it, as oscillation-gated connectivity may carry information about the expected timing of actions, besides their sensory content.

In order to dig deeper in the underlying physiological mechanisms, we investigated the link between neural oscillations and actual spiking activity, by focusing on the spike-field coherence in the mu band. The main result was that a significant fraction of units had a greater tendency to fire in the phase of the LFP that was associated with higher feeling of agency. This seems to speak in favour of the movement timing hypothesis, as bursts of M1 activity that generate movements may be more likely to happen during the favourable phase of the LFP. This would also be in line with the already mentioned fact that spontaneous movements tend to be phase locked to neural oscillations (Popovych et al., 2016; Tomassini et al., 2017). When BMI generated movements happen during the optimal phase window, agency ratings would then be higher simply because self-generated movements are more likely to occur at that time in an able-bodied human (see Schurger et al., 2012 for a similar account for readiness potential). On the other hand, it is interesting to note that the LFP phase at movement onset did not seem to significantly affect the subsequent spiking activity at the population level. This suggests that its effect on brain activity, which ultimately leads to a different experience of agency, may be more strongly reflected in other areas, and possibly related to transient changes in effective connectivity, as mentioned previously. In this respect, it is worth noting that a previous well controlled MEG study found higher beta and alpha connectivity between the contralateral motor cortex and premotor, insular and temporal regions to correlate with (overtly cued) sense of agency (Buchholz et al., 2019). In the context of visual perception, another elegant EEG-fMRI study showed that the phase of low-frequency oscillations at stimulus onset affects perceptual performance, by modulating the subsequent functional connectivity between the lateral occipital complex and the contralateral intraparietal sulcus (Hanslmayr et al., 2013). The same mechanism might apply to sensorimotor areas, acting as a primer for the subsequent connectivity between M1 and premotor or insular regions and affecting sense of agency as a result. In order to shed more light on such hypotheses, it would be necessary to record brain activity at a larger scale, as in the present intracortical BMI setting, designed for clinical purposes, recording sites were limited to the primary motor cortex.

The presence and the importance of predictive components in sense of agency have been extensively investigated and demonstrated behaviourally, but their neural mechanisms have remained largely unexplored. fMRI studies allowed pinpointing a set of key regions, mainly the TPJ, pre-supplementary motor area (pre-SMA), precuneus, and dorsomedial prefrontal cortex (Chambon et al., 2013; Sperduti et al., 2011; Yomogida et al., 2010; Zito et al., 2020). However, results largely varied across studies. More importantly, fMRI lacks the temporal resolution needed to disentangle pre-movement neural correlates, likely connected to predictive computations, from post movement signals, more likely linked to postdictive computations. Other studies using perturbative techniques such as TMS and tDCS, allowed investigating causality, specifically to gather further evidence in favour of the link between premotor and parietal areas and sense of agency (Chambon et al., 2015; Moore et al., 2010). However, these techniques lack the ability to elucidate the fine structure of the neural mechanisms involved. Most importantly, no previous study could manipulate the natural relation between intentions and body movements, allowing introducing nuisances in the normally perfect feeling of control for body movements, and searching for the neural source of such variability in the experience of agency.

Sense of agency is a crucial component of self-awareness, and it has important implications for healthy and pathological cognition. Voluntary actions and the underlying sense of agency allow infants to develop causal reasoning from the detection of sensorimotor congruencies (Zaadnoordijk et al., 2015). At the same time, neuropsychiatric disorders implying deficits of self experience, such as schizophrenia (Daprati et al., 1997; Hur et al., 2014; Mellor, 1970; Moore and Obhi, 2012) and autism (Sperduti et al., 2014) are accompanied by distorted feeling of agency, and it has been suggested that impaired predictive abilities may be a key pathogenic element (Alloy and Abramson, 1979; Fletcher and Frith, 2009; Sinha et al., 2014). However, little is known about the physiological bases of sensorimotor predictions in determining sense of agency. The present experimental setup allowed us to provide novel insights into the potential importance of pre-movement neural oscillations for such predictive mechanisms. The phase of pre-stimulus neural oscillations is already known to affect perceptual (Ai and Ro, 2014; Busch et al., 2009; Rice and Hagstrom, 1989), multisensory (Ikumi et al., 2019; Keil and Senkowski, 2018; Thézé et al., 2020) and sensorimotor (Tomassini et al., 2017) processing. More generally, their role in orchestrating information exchange in the brain is relatively well understood and provides a solid interpretative framework of brain functioning. Therefore, our study takes a decisive step towards explaining sense of agency, as well as its cognitive and clinical implications, through the general and well understood phenomenon of neural oscillations.

Materials and Methods

Participant

The participant was a 27-year-old male with quadriplegia at the C5/C6 level originating from a cervical spinal cord injury (SCI) dating to 8 years prior to data collection. He had a full range of motion

in both shoulders and elbow flection and could perform twitches of wrist extension (1/5 and 2/5 strength on left and right wrists respectively). He had no motor function below C6. His proprioception was intact in the right upper limb/shoulder for internal through external rotation, forearm pronation through supination, and wrist flexion through extension. Proprioception at the level of metacarpal-phalangeal joints for all right hand digits was impaired. He was enrolled in a pilot clinical trial (NCT01997125, Date: November 22, 2013) of a custom BMI system (Battelle Memorial Institute) to restore motor functionality of the upper limb following SCI. The BMI system required the implantation of a Utah microelectrode array (96 channels, 4.4 x 4.2 mm, 1.5 mm depth) in the hand region of the left primary motor cortex. Reference wires were placed subdurally. The target region was identified via pre-operative functional Magnetic Resonance Imaging as the patient was asked to attempt performing right hand movements. See the first description of the BMI system by Bouton et al. (2016) for further details about the participant and surgical process.

BMI system

Neural data from the Utah array was sampled at 30kHz and band-pass filtered between 0.3Hz and 7.5kHz at the hardware level (3^{rd} order Butterworth). The data were digitized in 100ms bins and analysed through custom MATLAB code. Before decoding, artefacts due to NMES were removed by blanking the signal over 3.5ms around the artefact, defined as a signal amplitude exceeding 500μV in at least 4 out of 12 randomly selected channels. Neural decoding was based on a non-linear Support Vector Machine (SMV; e.g., Cortes and Vapnik, 1995). The SVM used 96 input features consisting of the mean wavelet power (MWP) for each channel and 100 ms bin. To obtain the MWPs, neural activity was decomposed into 11 wavelet scales (Daubechies wavelet, MATLAB), and the coefficients of wavelets 3-6, corresponding to the multi-unit frequency band spanning from 235 to 3.75kHz, were averaged for each channel. The decoder was re-trained before each experimental session, by asking the participant to attempt performing one of the four hand movements (HO, HC, TE, TF) in order to generate the training data. The subject performed 7 blocks consisting of 3 repetitions per movement type each. The decoder output consists of four scalar numbers in the -1/1 range, indicating the relative probability for each movement. A threshold of 0 was set for the selection of an intended movement, with the movement with the highest score prevailing if two or more classes exceeded the threshold. A custom-built Neuromuscular Electrical Stimulation (NMES) system was used to translate the decoded intentions into actual hand movements, by stimulating forearm muscles. The NMES system consisted of a circumferential forearm sleeve with 130 copper-coated electrodes, 12mm in diameter. The electrodes were disposed in an array, spaced at regular intervals (22mm longitudinally x 15mm transversely). Stimulation was delivered through rectangular pulses of 50Hz monophasic current (pulse width 500μs, amplitude 0-20mA). The stimulation patterns and intensity were re-calibrated at the beginning of each session in order to optimize the match with the participant's intentions. See the paper by Bouton et al. (2016) for further details about the neural decoder and NMES system.

Experiment 1 – protocol

In Experiment 1, we manipulated the congruency between the participant's motor intentions and sensory feedback and assess how this affected his sense of agency. Each trial started with a verbal cue about the hand movement to be performed ("hand open", "hand close", "thumb extension", "thumb flexion"), followed after a 2 second delay by a verbal "go" cue. The participant was instructed to start attempting the cued movement only at the "go", without anticipating. During the following 4 seconds, the participant received visual and somatosensory feedback according to the decoded movement and the feedback congruency for that trial and sensory modality. Somatosensory feedback was delivered by eliciting the target movement through the NMES sleeve. Visual feedback was constituted by an animation of a virtual hand performing the target movement, displayed on a screen placed horizontally to cover the participant's right hand. The hand model and the animation corresponded to the ones routinely used by the participant during BMI training sessions, and its size and position were adjusted to match the participant's real hand. In trials with congruent somatosensory (and/or visual) feedback, the decoded movement was executed through NMES (or displayed in a virtual animation). In incongruent trials, the opposite movement was executed and/or displayed, replacing hand opening with hand closing, thumb extension with thumb flexion, and vice versa. Sensory feedback was only delivered when one of the output classes of the neural decoder reached the threshold of 0. In the 5-6 seconds after the sensory feedback phase, the participant answered two questions, Q1 and Q2, about his feeling of agency for the movement. The whole experiment consisted of five experimental sessions performed over different days, each consisting of four blocks of BMI training and four blocks of experiment, each lasting around 15 minutes. Each experimental block consisted of 32 trials, where each combination of V/S feedback and cued movement (2X2X4=16) was repeated twice. Therefore, the grand total of trials was 640, 160 for each feedback condition.

Experiment 2 – protocol

The second experiment was part of a broader study (currently under consideration), aiming at investigating the effects of manipulating the intentionality chain composed of intentions, motor acts, and their consequences on the external world, based on a modification of Libet's 1983 intentionality experiment. Here, we focused on two experimental sessions within that study to establish an implicit measure of sense of agency. In the first session (high agency), the participant was cued to perform one of two possible movements through the BMI system, HO and HC. While performing the movements, the participant observed a single hand clock on a computer screen, with numbers from 5 to 60, completing a full rotation in 2.56 seconds. Again, movements were triggered by the activation of the neural decoder, but the NMES was always activated congruently. Additionally, 300 ms after HC was executed, a 1000 Hz "beep" was produced, lasting 100 ms. No additional consequence followed HO execution. The participant was instructed to pay attention to the location of the clock

hand at the time of movement onset, and to report it at the end of the trial allowing us to measure the perceived timing of the action. Differently from Experiment 1, the movements were self-paced, meaning the participant was instructed to freely initiate the movement and encouraged to vary his waiting time, which should in any case exceed one full clock rotation. In the second session (low agency), the only difference was that the same movements were executed passively, by randomly activating the NMES for HO or HC while the participant was instructed to remain at rest. Each session consisted of 80 trials, 40 per movement.

Data pre-processing

For LFPs, the main source of, data pre-processing consisted essentially of four steps: trial selection, artefact removal, down sampling, and epoching. Trial selection had the main goal of discarding trials in which the participant failed to generate any movement, or to activate the correct decoder. Therefore, we only kept trials in which the participant managed to keep the cued decoder above threshold for at list 600 ms (6 classifier bins). In Experiment 1, we additionally required that such movement happens after the “go” cue, and at least 1.5 seconds before the “stop”, in order to ensure a sufficient time window for epoching. Additionally, 5 HC trials from the high agency session had to be removed due to technical issues with the recording. After trial parsing, we retained 422 out of 640 trials for Experiment 1 (66 %), and 61 out of 80 (76 %) trials from the high agency session in Experiment 2 (all trials were retained in the low agency session as the participant did not need to activate the decoder in this session). The parsing was relatively even across conditions of interest, with 114/160 for V+/S+, 93/160 for V+/S-, 117/160 for V-/S+, and 98/160 for V-/S-. Similarly, in Experiment 2, we retained 26/35 (HC) trials for the movement eliciting the sound, and 35/40 for the movement not eliciting the sound (HO). Artefact removal was performed before epoching, as done online for BMI decoding, with the difference that we applied a 8.7 ms blanking window, in order to be more conservative on oscillatory analyses. Then, the data was down sampled to 1000 Hz, using a Kaiser anti-aliasing kernel. Spiking activity was extracted through the wave_clus spike detection and sorting algorithm with default settings. For the detection, a threshold was set at four times the standard deviation of baseline noise. Spikes were clustered through the superparamagnetic clustering algorithm, allowing to remove spurious signals. Since data collection was done in different sessions spanning several weeks, and the number of units in each channel fluctuated across sessions, we did not attempt to match units across recording sessions. Instead, we pool the spikes at each channel as multiunit activity. For both LFP and multiunit activity, the data was epoched by time-locking to the onset of sensory feedback.

Oscillatory analyses

Since in our analyses we focus on low frequencies, which are expected to be highly coherent on the small spatial scales of an Utah array, all analyses were performed on the mean LFP across all channels. The high coherence of oscillations in the 4-16 Hz range was furthermore confirmed by

analyses shown in the Supplementary Information (Figure S4). Instantaneous values for power and oscillatory phase were obtained by convolving the signal with Morelet wavelets over 10 logarithmically spaced frequencies between 4 and 16 Hz, setting the number of cycles at 2π. Our main analysis focuses on quantifying phase opposition in time and frequency between conditions of interest (high-low levels of explicitly or implicitly assessed agency). As a measure of phase opposition, we use the Phase Opposition Product (POP VanRullen, 2016). To compute the POP, first we compute the inter-trial phase coherence (ITC) for all the trials pooled together, and for the two conditions separately.

$$ITC_{ALL} = |\sum_{i=1:n} \omega_i/|\omega_i||/n \ (1)$$

$$ITC_A = |\sum_{conditionA} \omega_i/|\omega_i||/n_A \ (2)$$

$$ITC_{ALL} = |\sum_{conditionB} \omega_i/|\omega_i||/n_B \ (3)$$

Then, the POP is simply obtained as follows

$$POP = ITC_A ITC_B - ITC_{ALL}^2 \ (4)$$

The underlying idea is that, if trials within a condition are clustered around some angle, and trials in the other condition are clustered around an opposed angle, then the inter-trial coherence within conditions is going to be higher than when pooling the trials together. Therefore, higher values of POP indicate a stronger phase opposition between conditions. Instantaneous values of oscillatory power were simply computed as the absolute value of the wavelet convolution. MATLAB code for the wavelet convolution was adapted from Mike Cohen's website (http://mikexcohen.com/lectures.html).

Statistical analysis

Statistical analyses of the time-frequency distribution of POP values were performed through cluster based permutation tests to address the multiple comparison problem (Maris and Oostenveld, 2007). In order to run the permutations, a suitable statistics to define clusters needs to be defined for POP values, as with a single subject it is not possible to simply run a T-test on POP values across subjects. To the best of our knowledge, the analytical form for the null distribution of POP values is not known, and running permutations to define a P-value for each time-frequency point, nested in the main cluster correction permutation, would be too computationally demanding. To overcome this problem, we used an heuristic method, by generating randomly distributed phase angles, splitting them in two groups to compute their POP, and then fitting such null distribution of POP values as a function of the number of trials in each condition (see Fig S5 for additional details). We found this surrogated POP null distribution to be well fitted by the formula

$$P(POP > x) \approx \left[1 - \frac{n_A n_B}{(n_A+n_B)^2+(n_A-n_B)^2}\right] e^{-2x\sqrt{n_A n_B}} \ (5)$$

Where n_A and n_B denote the number of trials in each condition. The factor in square brackets denotes the approximated probability that the POP is positive, since our method was only applied to fit

positive values, which are of interest for statistical analyses. We set the P-value of negative POP values to 0.75 by default, as this will not affect the results of cluster correction in any case. The approximated P-values were then transformed into T-values, and a threshold of 2 was set to define the clusters. The total value of each cluster was then defined as the sum of the T-values of all time-frequency points composing it. Importantly, the exact nature of the statistics used at this stage to define cluster scores does not influence the test's ability to appropriately control for type I errors, as this is addressed by the permutations performed subsequently (Maris and Oostenveld, 2007). Therefore, the approximation we use allows to save computational time while not affecting the final result. The final P-value for each cluster was defined as the probability of finding a cluster with a larger score over 10000 permutations, obtained by randomly reassigning trials to the high or low agency condition. The analysis was performed over a 1s window ending at the time of movement onset.

Multiunit analyses

Analyses on multiunit activity consisted essentially of the evaluation of spike-field coherence with mu oscillations, and of the analysis of similarity of activity depending on the mu phase at movement onset. To define the spike-field coherence, units were not pooled for each channel, and each unit was analysed separately for each session, for a total of 1408 units. For each unit, the spike field coherence was calculated by extracting the phase vectors for LFP oscillations at 7.5 Hz at each spike location, and then applying formula (1) to the ensemble of phase vectors. In order to compute statistics for the values of SFC, we start from the null distribution for the length of the resulting vector in (1), which for a large number of spikes is well approximated by a Gaussian (see Fig S6)

$$ITC = N(\mu, \sigma) \quad (6)$$

$$\mu = \sqrt{\frac{\pi}{4n}} \quad (7)$$

$$\sigma = \sqrt{\frac{4-\pi}{4n}} \quad (8)$$

This allows to quickly compute Z scores and p-values from the ITC of spike-locked phase vectors.

Since the neural activity recorded by the microelectrode array varied significantly between experimental sessions (i.e., days of recording) and movements, Euclidean distances were computed separately within each session and movement, and then averaged to obtain the final results. To estimate confidence intervals, we used a bootstrapping technique, again applied within sessions and movements. For each session and movement, n random trials were extracted replacement, where n is the number of trials for that condition. The procedure was repeated 100 times, and 95 % confidence intervals were obtained as 1.96 times the standard deviation of the distances obtained from the bootstrapped trials.

References

Ai, L., Ro, T., 2014. The phase of prestimulus alpha oscillations affects tactile perception. J. Neurophysiol. 111, 1300–1307. https://doi.org/10.1152/jn.00125.2013

Alloy, L.B., Abramson, L.Y., 1979. Judgment of contingency in depressed and nondepressed students: Sadder but wiser? J. Exp. Psychol. Gen. https://doi.org/10.1037/0096-3445.108.4.441

Blanke, O., Metzinger, T., 2009. Full-body illusions and minimal phenomenal selfhood. Trends Cogn. Sci. 13, 7–13. https://doi.org/10.1016/j.tics.2008.10.003

Bouton, C.E., Shaikhouni, A., Annetta, N. V., Bockbrader, M.A., Friedenberg, D.A., Nielson, D.M., Sharma, G., Sederberg, P.B., Glenn, B.C., Mysiw, W.J., Morgan, A.G., Deogaonkar, M., Rezai, A.R., 2016. Restoring cortical control of functional movement in a human with quadriplegia. Nature 533, 247–250. https://doi.org/10.1038/nature17435

Buchholz, V.N., David, N., Sengelmann, M., Engel, A.K., 2019. Belief of agency changes dynamics in sensorimotor networks. Sci. Rep. 9, 1–12. https://doi.org/10.1038/s41598-018-37912-w

Busch, N.A., Dubois, J., VanRullen, R., 2009. The phase of ongoing EEG oscillations predicts visual perception. J. Neurosci. 29, 7869–7876. https://doi.org/10.1523/JNEUROSCI.0113-09.2009

Chambon, V., Moore, J.W., Haggard, P., 2015. TMS stimulation over the inferior parietal cortex disrupts prospective sense of agency. Brain Struct. Funct. 220, 3627–3639. https://doi.org/10.1007/s00429-014-0878-6

Chambon, V., Wenke, D., Fleming, S.M., Prinz, W., Haggard, P., 2013. An online neural substrate for sense of agency. Cereb. Cortex 23, 1031–1037. https://doi.org/10.1093/cercor/bhs059

Cortes, C., Vapnik, V., 1995. Support-Vector Networks. Mach. Learn. 20, 273–297. https://doi.org/10.1023/A:1022627411411

Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., Jeannerod, M., 1997. Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. Cognition 65, 71–86. https://doi.org/10.1016/S0010-0277(97)00039-5

Desideri, D., Zrenner, C., Ziemann, U., Belardinelli, P., 2019. Phase of sensorimotor μ-oscillation modulates cortical responses to transcranial magnetic stimulation of the human motor cortex. J. Physiol. 597, 5671–5686. https://doi.org/10.1113/JP278638

Fletcher, P.C., Frith, C.D., 2009. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat. Rev. Neurosci. 10, 48–58. https://doi.org/10.1038/nrn2536

Fries, P., 2015. Rhythms for Cognition: Communication through Coherence. Neuron 88, 220–35. https://doi.org/10.1016/j.neuron.2015.09.034

Fries, P., 2005. A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. Trends Cogn. Sci. 9, 474–480. https://doi.org/10.1016/j.tics.2005.08.011

Frith, C.D., Blakemore, S.J., Wolpert, D.M., 2000. Abnormalities in the awareness and control of

action. Philos. Trans. R. Soc. B Biol. Sci. 355, 1771–1788. https://doi.org/10.1098/rstb.2000.0734

Gallagher, S., 2000. Philosophical conceptions of the self: implications for cognitive science. Trends Cogn. Sci. 4, 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Haggard, P., Clark, S., Kalogeras, J., 2002. Voluntary action and conscious awareness. Nat. Neurosci. 5, 382–385. https://doi.org/10.1038/nn827

Hanslmayr, S., Gross, J., Klimesch, W., Shapiro, K.L., 2011. The role of alpha oscillations in temporal attention. Brain Res. Rev. 67, 331–343. https://doi.org/10.1016/j.brainresrev.2011.04.002

Hanslmayr, S., Volberg, G., Wimber, M., Dalal, S.S., Greenlee, M.W., 2013. Prestimulus oscillatory phase at 7 Hz gates cortical information flow and visual perception. Curr. Biol. 23, 2273–2278. https://doi.org/10.1016/j.cub.2013.09.020

Hur, J.-W., Kwon, J.S., Lee, T.Y., Park, S., 2014. The crisis of minimal self-awareness in schizophrenia: A meta-analytic review. Schizophr. Res. 152, 58–64. https://doi.org/10.1016/j.schres.2013.08.042

Ikumi, N., Torralba, M., Ruzzoli, M., Soto-Faraco, S., 2019. The phase of pre-stimulus brain oscillations correlates with cross-modal synchrony perception. Eur. J. Neurosci. 49, 150–164. https://doi.org/10.1111/ejn.14186

Keil, J., Senkowski, D., 2018. Neural Oscillations Orchestrate Multisensory Processing. Neuroscientist 24, 609–626. https://doi.org/10.1177/1073858418755352

Landau, A.N., Fries, P., 2012. Attention samples stimuli rhythmically. Curr. Biol. 22, 1000–1004. https://doi.org/10.1016/j.cub.2012.03.054

Libet, B., Gleason, C.A., Wright, E.W., Pearl, D.K., 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. Brain. https://doi.org/10.1093/brain/106.3.623

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Mellor, C.S., 1970. First rank symptoms of schizophrenia. I. The frequnncy in schizophrenics on admission to hospital. II. Differences between individual first rank symptoms. Br. J. Psychiatry. https://doi.org/10.1192/s0007125000192116

Moore, J.W., Obhi, S.S., 2012. Intentional binding and the sense of agency: A review. Conscious. Cogn. 21, 546–561. https://doi.org/10.1016/j.concog.2011.12.002

Moore, J.W., Ruge, D., Wenke, D., Rothwell, J., Haggard, P., 2010. Disrupting the experience of control in the human brain: Pre-supplementary motor area contributes to the sense of agency. Proc. R. Soc. B Biol. Sci. 277, 2503–2509. https://doi.org/10.1098/rspb.2010.0404

Popovych, S., Rosjat, N., Toth, T.I., Wang, B.A., Liu, L., Abdollahi, R.O., Viswanathan, S., Grefkes, C., Fink, G.R., Daun, S., 2016. Movement-related phase locking in the delta–theta frequency band. Neuroimage 139, 439–449. https://doi.org/10.1016/j.neuroimage.2016.06.052

Rice, D.M., Hagstrom, E.C., 1989. Some evidence in support of a relationship between human auditory signal-detection performance and the phase of the alpha cycle. Percept. Mot. Skills. https://doi.org/10.2466/pms.1989.69.2.451

Schurger, A., Sitt, J.D., Dehaene, S., 2012. An accumulator model for spontaneous neural activity prior to self-initiated movement. Proc. Natl. Acad. Sci. 109, E2904–E2913. https://doi.org/10.1073/pnas.1210467109

Sinha, P., Kjelgaard, M.M., Gandhi, T.K., Tsourides, K., Cardinaux, A.L., Pantazis, D., Diamond, S.P., Held, R.M., 2014. Autism as a disorder of prediction. Proc. Natl. Acad. Sci. 111, 15220–15225. https://doi.org/10.1073/pnas.1416797111

Sperduti, M., Delaveau, P., Fossati, P., Nadel, J., 2011. Different brain structures related to self- and external-agency attribution: A brief review and meta-analysis. Brain Struct. Funct. 216, 151–157. https://doi.org/10.1007/s00429-010-0298-1

Sperduti, M., Pieron, M., Leboyer, M., Zalla, T., 2014. Altered Pre-reflective Sense of Agency in Autism Spectrum Disorders as Revealed by Reduced Intentional Binding. J. Autism Dev. Disord. 44, 343–352. https://doi.org/10.1007/s10803-013-1891-y

Synofzik, M., Vosgerau, G., Newen, A., 2008. I move, therefore I am: A new theoretical framework to investigate agency and ownership. Conscious. Cogn. 17, 411–424. https://doi.org/10.1016/j.concog.2008.03.008

Thézé, R., Giraud, A., Mégevand, P., 2020. The phase of cortical oscillations determines the perceptual fate of visual cues in naturalistic audiovisual speech. Sci. Adv. 6, eabc6348. https://doi.org/10.1126/sciadv.abc6348

Tomassini, A., Ambrogioni, L., Medendorp, W.P., Maris, E., 2017. Theta oscillations locked to intended actions rhythmically modulate perception. Elife. https://doi.org/10.7554/eLife.25618.001

VanRullen, R., 2016. How to evaluate phase differences between trial groups in ongoing electrophysiological signals. Front. Neurosci. 10, 1–22. https://doi.org/10.3389/fnins.2016.00426

VanRullen, R., Carlson, T., Cavanagh, P., 2007. The blinking spotlight of attention. Proc. Natl. Acad. Sci. 104, 19204–19209. https://doi.org/10.1073/pnas.0707316104

Wegner, D.M., 2002. The illusion of conscious will. MIT Press.

Yomogida, Y., Sugiura, M., Sassa, Y., Wakusawa, K., Sekiguchi, A., Fukushima, A., Takeuchi, H., Horie, K., Sato, S., Kawashima, R., 2010. The neural basis of agency: An fMRI study. Neuroimage 50, 198–207. https://doi.org/10.1016/j.neuroimage.2009.12.054

Zaadnoordijk, L., Hunnius, S., Meyer, M., Kwisthout, J., van Rooij, I., 2015. The developing sense of agency: Implications from cognitive phenomenology, in: 2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob). IEEE, pp. 114–115. https://doi.org/10.1109/DEVLRN.2015.7346126

Zito, G.A., Wiest, R., Aybek, S., 2020. Neural correlates of sense of agency in motor control: A

neuroimaging meta-analysis. PLoS One 15, e0234321.
https://doi.org/10.1371/journal.pone.0234321

Supplementary information

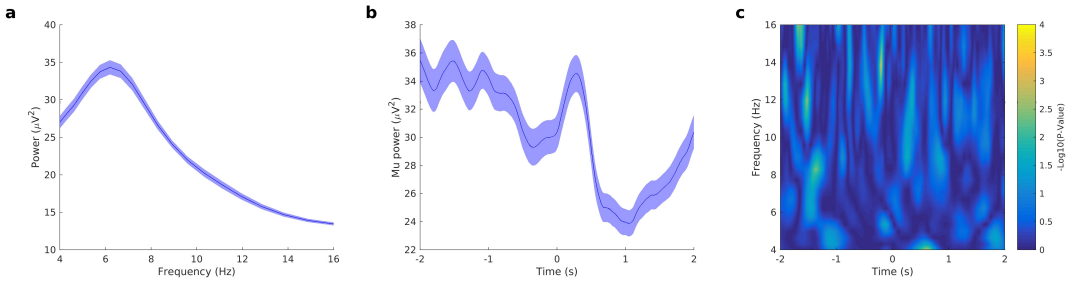


Figure S1: *analyses on oscillatory power. (a) Power by frequency band in the baseline period ranging from -2 to -1 seconds from movement. In our participant, the mu peak is relatively low in frequency, with the maximal spectral density found at 6.2 Hz. (b) Time course of mu power (6.2 Hz), averaged over all trials, showing a decrease immediately before and after the movement (ERD). The rebound after movement onset is likely due to the transient effect of the ERP elicited by the NMES induced movement. The shaded area denotes the standard error. (c) T-test on 4-16 Hz power between high and low agency trials, as in the main text the analysis was done on the subset of V-/S+ and V+/S- trials. Cluster based correction found no significant differences between the two conditions.*

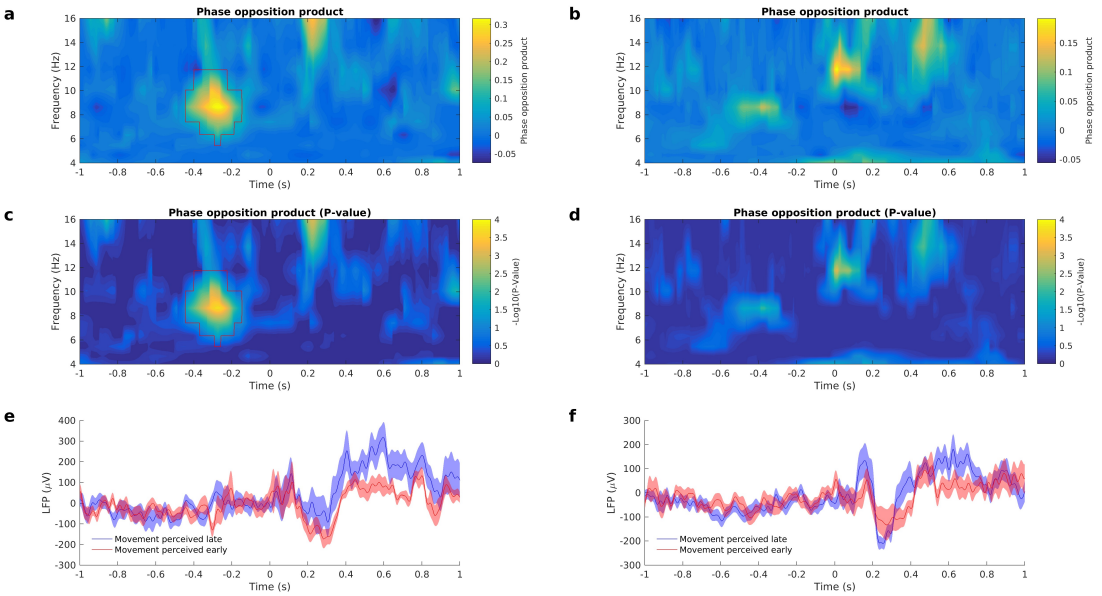


Figure S2: *Comparison of operant and non-operant movement. Left panels refer to the operant movement, right panels to the non-operant movement. (a-b) indicate the phase opposition product, (c-d) the relative P-value. The largest cluster for the operant movement, indicated by the red contour, was significant with p = 0.04, while it was well below significance for the non-operant movement (p = 0.38). (e-f) show the averaged LFP time course, with shades indicating its standard error.*

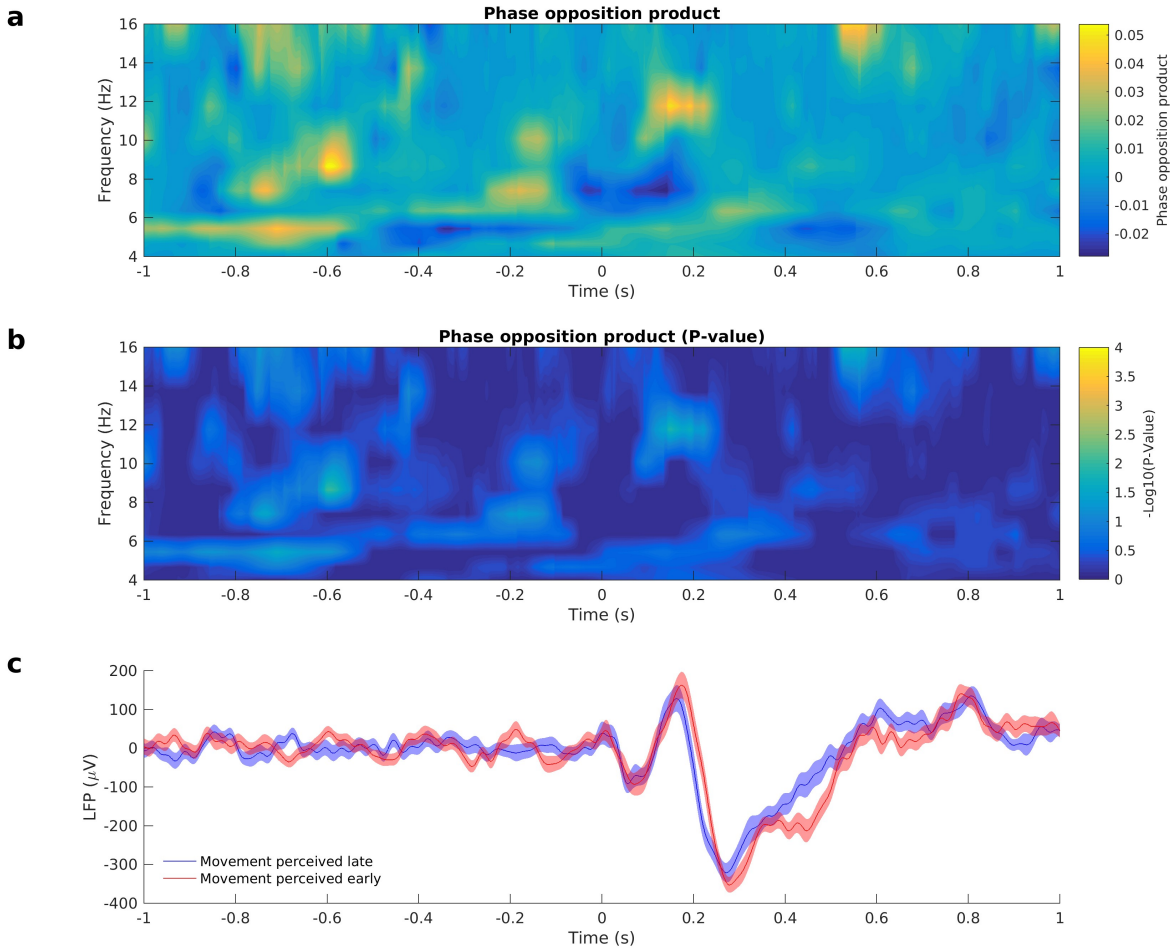


Figure S3: *Control analysis on passive movement condition. (a) Phase opposition product contrasting trials with late-early perception of passively induced movements. (b) P-values for the phase opposition product. No significant cluster emerged after correction. (c) Averaged time course of the LFP in the two conditions. Shaded areas indicate standard errors.*

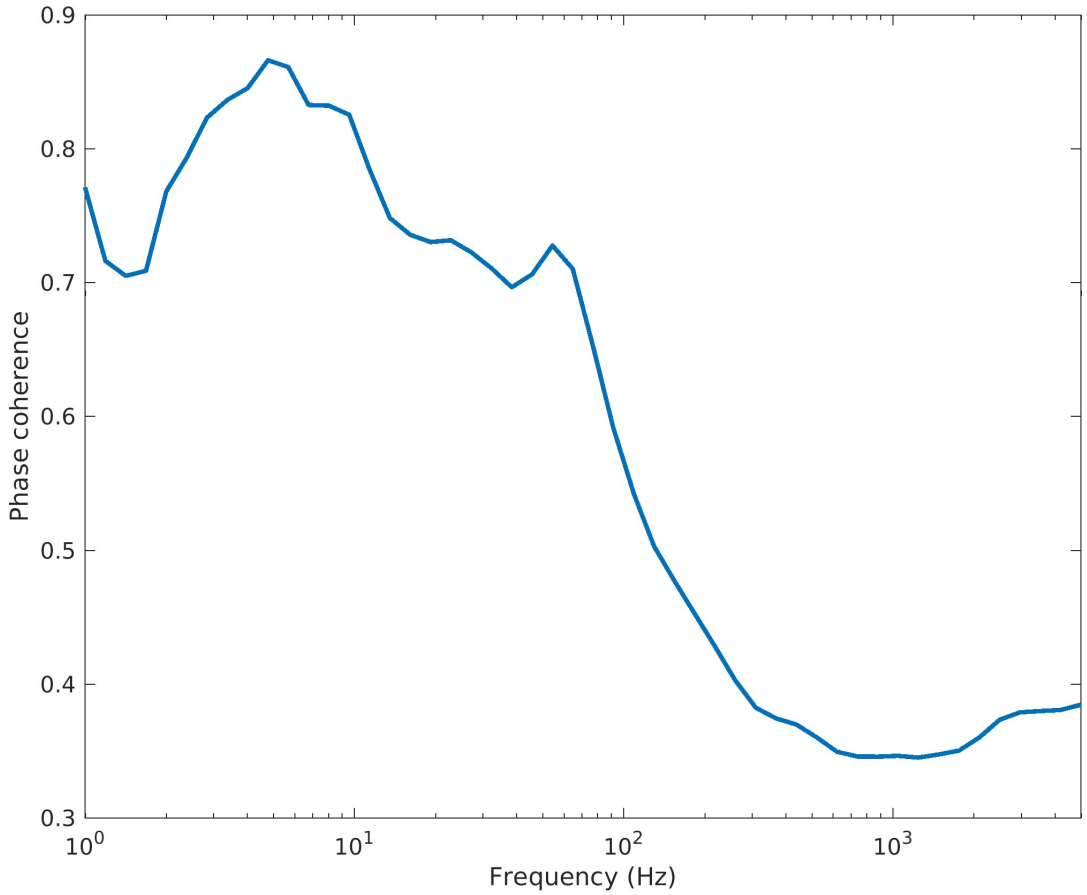


Figure S4: *phase coherence across channels. The phase coherence was computed as the average ITC (formula (3) of the main text) across channels over a 15 minutes sample recording (raw data after artefact blanking), for frequencies between 1 Hz and 5 KHz. In the mu range, the ITC is consistently above 0.8, indicating small loss of information in this frequency band after averaging across channels.*

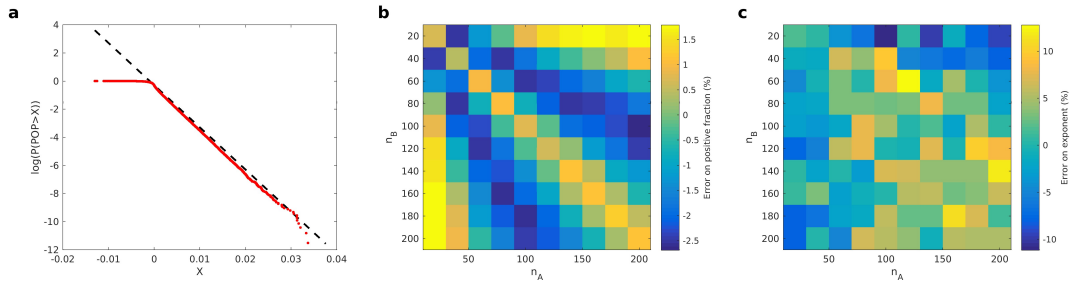


Figure S5: *estimation of the null distribution of POP values. The null distribution is expected to depend on the number of trials in each condition, n_A and n_B. Therefore, we sampled a 10x10 grid of n_A and n_B values, and for each pair we generated n_A and n_B random angles 100000, and computed the relative POP values. First of all, we noticed that, when POP values are positive, their cumulative distribution is approximately exponential (a). Then, to obtain a good approximation of the distribution for positive POP values, it is sufficient to estimate how the exponent and the positive fraction of POP values depend on n_A and n_B (negative values of POP are not relevant as they cannot be associated with significant p-values). Good fitting functions were identified heuristically, and are shown in formula (5) of the main text. Here, we plot the relative error between simulated values and our approximation for the positive fraction (b) and the exponent (c) of the null cumulative distribution. Errors were in most cases between -5% and +5%, which we considered acceptable since the p-values obtained with this method are only a preliminary step in cluster based correction, which is not expected to bias the final statistical results (see main text, methods).*

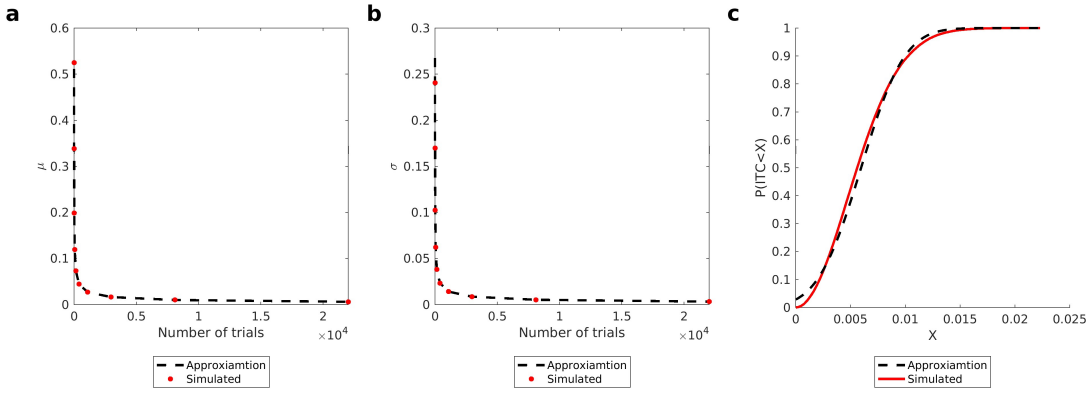


Figure S6: *estimation of the null distribution for ITC values, used in the multiunit analysis. To obtain samples from the null distribution of ITC values, for each tested number of trials N, we generated N random angles 100000 times, and computed the ITC value for each sample. Panel (a) shows the empirical mean value of ITC as a function of the number of trials (red dots), compared to the theoretical approximation from formula (7) of the main text. Panel (b) shows the same for the standard deviation, where the theoretical approximation is obtained as per formula (8). In panel (c) we show that, indeed, the distribution of ITC values is Gaussian with good approximation, by comparing the empirical cumulative distribution of ITC values with that of a Gaussian with mean and variance predicted by our approximation.*