A GENERATIVE MODEL OF BODY OWNERSHIP

AND MINIMAL SELFHOOD

Dissertation

zur Erlangung des akademischen Grades

Doctor of Philosophy (Ph.D.)

am Fachbereich Erziehungswissenschaften und Psychologie

der Freien Universität Berlin

vorgelegt von

Dipl.-Psych. Jakub Limanowski

Berlin 2014

Erstgutachter: Prof. Dr. Felix Blankenburg

Zweitgutachter: Prof. Dr. Arno Villringer

Tag der Disputation: 11.02.2015

Acknowledgments

Contents

Supplements

Abbreviations

AI   Active inference

BOI   Body ownership illusion

BOLD Blood oxygenation level dependent

DCM   dynamic causal modeling

EBA   extrastriate body area

FEP   free energy principle

fMRI   functional magnetic resonance imaging

IPS   intraparietal sulcus

LOC   lateral occipito-temporal cortex

MPS   minimal phenomenal selfhood

PMv   ventral premotor cortex

PPC   posterior parietal cortex

PPI   psycho-physiological interaction

RHI   rubber hand illusion

SI   primary somatosensory cortex

SII   secondary somatosensory cortex

TMS   transcranial magnetic stimulation

Abstract

The experience of a body as one's own enables one to perceive and interact with the world. Body ownership has long been recognized as one of the enabling conditions for *minimal selfhood*. The *rubber hand illusion* (RHI) demonstrates that simultaneous touch to a dummy hand and one's hidden real hand induces an illusion of feeling the touches on the dummy hand, and a corresponding illusion of "ownership" of the dummy hand. The mechanisms underlying the RHI have been speculatively explained in terms of Bayesian inference on ambiguous sensory input, and the implementation of such mechanisms has been ascribed to fronto-parietal brain areas. However, research has not established a link between neuroanatomical models of the RHI and models of probabilistic information processing in the brain, nor a connection of these models to philosophical accounts of minimal selfhood.

The papers presented here were aimed at addressing these issues, via an application of the *free energy principle* (FEP), an account of cortical information processing focused on predictive coding. In two fMRI studies, we demonstrated an involvement of visual body-selective brain regions in the RHI. During the RHI these areas interacted with the fronto-parietal circuit identified previously; the resulting network showed enhanced connectivity from somatosensory and visual areas to multimodal parietal areas. A TMS experiment showed that disrupting activity of the body-selective visual cortex increased behavioral effects of the RHI. Finally, in two theoretical papers, the compatibility of the FEP with philosophical accounts of minimal selfhood is demonstrated. Taken together, these papers show that the information flow during manipulations of body ownership may well be formalized within a predictive coding scheme, and thus suggest the FEP as the missing link between models of the RHI and cortical information flow, and as an explanation of the mechanisms underlying minimal selfhood in terms of prediction error suppression across a hierarchical generative model of the world.

Zusammenfassung

Einen Körper als „meinen" zu erleben, ermöglicht es, die Welt wahrzunehmen und mit ihr zu interagieren. Der „Besitz" eines Körpers wurde schon früh als eine Grundvoraussetzung für ein *minimales Selbstbewusstsein* erkannt. Die *Gummihand-Illusion* (GHI) zeigt, dass die simultane Berührung einer künstlicher Hand und der eigenen Hand die Illusion erzeugt, die Berührung auf der künstlichen Hand zu spüren, was einen illusorischen „Besitz" der künstlichen Hand zur Folge hat. Es wurde spekuliert, dass die Grundlage der GHI durch Bayessche Inferenz erklärt werden kann. Die neuronale Implementation solcher Mechanismen vermutet man in fronto-parietalen Hirnarealen. Jedoch ist bisher noch keine Verknüpfung zwischen neuroanatomischen Modellen der GHI und Modellen probabilistischer Informationsverarbeitung im Gehirn, oder philosophischen Erklärungen des minimalen Selbst aufgezeigt worden.

Die hier vorgestellten Arbeiten sollten mit Hilfe des *Freie-Energie-Prinzips* (FEP), einer Erklärung kortikaler Informationsverarbeitung fokussiert auf Prädiktionskodierung, diesen offenen Fragen nachgehen. In zwei fMRT Untersuchungen weisen wir die Beteiligung visuell körperselektiver Hirnregionen an der GHI nach. Während der GHI interagierten diese Regionen mit den schon beschriebenen fronto-parietalen Hirnarealen. Darüber hinaus war im Gesamtnetzwerk die Konnektivität vom somatosensorischen und visuellen Kortex zum multimodalen parietalen Kortex verstärkt. In einem weiteren Experiment rief ein Stören der Aktivität im körperselektiven visuellen Kortex mittels TMS eine Verstärkung behavioraler Effekte der GHI hervor. Schließlich wird in zwei theoretischen Artikeln die Kompatibilität des FEP mit philosophischen Zugängen zum minimalen Selbst herausgestellt. Insgesamt zeigen diese Arbeiten, dass der Informationsfluss während illusionärem „Körperbesitz" plausibel mittels Prädiktionskodierung formalisiert werden kann. Damit ermöglicht das FEP den fehlenden Brückenschlag zwischen Modellen der GHI und solchen kortikaler Informationsverarbeitung. Darüber hinaus bietet das FEP eine Erklärung der Mechanismen des minimalen Selbst durch Unterdrückung eines Vorhersagefehlers in einem hierarchischen generativen Modell der Welt.

List of original articles

This dissertation is based on the following articles:

Limanowski, J., Lutti, A., Blankenburg, F. (2014). The extrastriate body area is involved in illusory limb ownership. *NeuroImage, 86*, 514-524. doi: 10.1016/j.neuroimage.2013.10.035

Limanowski, J., Blankenburg, F. (2015). Network activity underlying the illusory self-attribution of a dummy arm. *Human Brain Mapping (Epub. ahead of print).* doi: 10.1002/hbm.22770

Wold, A., **Limanowski, J.**, Walter, H., & Blankenburg, F. (2014). Proprioceptive drift in the rubber hand illusion is intensified following 1 Hz TMS of the left EBA. *Frontiers in Human Neuroscience, 8*, 390. doi: 10.3389/fnhum.2014.00390

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

Limanowski, J. (2014). What can body ownership illusions tell us about minimal phenomenal selfhood? *Frontiers in Human Neuroscience, 8,* 946. doi: 10.3389/fnhum.2014.00946

1.   Introduction

"What is real? How do you define 'real'? If you're talking about what you can feel,

what you can smell, what you can taste and see, then 'real' is simply electrical signals

interpreted by your brain."

Morpheus (*The Matrix*, 1999)

Few things of such crucial importance for our experience of the world are as blindly taken for granted as being a self in possession of a particular body. The body—as aptly put by early phenomenologist Maurice Merleau-Ponty (1945)—is my "vehicle of being in the world": the body structures my perception of the world, and it allows me to interact with the world (James, 1890; Gibson, 1977). The phenomenal access to these two aspects of the body—the "sense of body ownership" (the self-identification with a particular body) and the "sense of agency" (the awareness that it is my body who acts)—has been proposed as the basis for our self-experience (Gallagher, 2000; see Bermúdez, 1998; Anderson and Perlis, 2005). But not only ours: analytic philosophy has identified them as the fundamental enabling conditions for *any* kind of self-experience of any organism or system—in short, for *minimal selfhood*[1] (Metzinger, 2005; see also Gallagher, 2000; Anderson and Perlis, 2005; Blanke and Metzinger, 2009; Tsakiris, 2010; Gallagher and Zahavi, 2010). Thereby body ownership has been proposed as a more "fundamental" condition than the sense of agency (Gallagher, 2000; Legrand, 2006; Tsakiris, 2010), because as William James (1890) recognized, the

---

[1] Whereas some authors (e.g. Gallagher, 2000) use "minimal selfhood", the term "minimal phenomenal selfhood" is used by Metzinger. In the parts of this thesis where I largely draw on Metzinger's conceptualizations, I will use his wording.

body is "always there", even without action. This grounding of self-experience in body ownership naturally determines the research questions for experimental approaches to the study of minimal selfhood.

In the recent years, new advances in cognitive science and the philosophy of mind have revealed that indeed the brain seems to have developed remarkable mechanisms for identifying with a particular body and distinguishing it from external things. But luckily for us scientists, as with almost any mechanism that works efficiently, there are some rare, deviant cases in which the mechanism's guiding rules lead to false inference. In exploiting exactly these cases, illusions are an excellent opportunity to reveal the mechanisms underlying the self-attribution of the body and the sense of agency. The papers presented in this thesis have made use of perhaps the most established such illusion to investigate the sense of body ownership and its enabling mechanisms in healthy individuals: the so-named *rubber hand illusion* (RHI, Botvinick and Cohen, 1998).

1.1. Manipulating the sense of body ownership: The rubber hand illusion

Matthew Botvinick and Jonathan Cohen (1998) discovered that when they touched a participant's hidden hand together with an ipsilateral dummy hand placed next to the real hidden hand (the dummy was made of rubber, thus the name of the illusion), the participant often would report that she felt the touches occurring on the dummy hand instead of on her own hand. More strikingly, most participants also reported feeling as if the dummy hand was a part of their own body. When asked to report the perceived location of their real hidden hand, participants' exhibited a bias towards the location of the dummy hand. Many replications and extensions of the RHI paradigm have been published, which have further shown that a threat to the self-attributed dummy hand (e.g., approaching or stabbing the dummy hand with a knife) evokes an anxiety response measurable by increased skin conductance (Armel and Ramachandran, 2003; Ehrsson et al., 2007; Gentile et al.,

2013). Others have reported an accompanying drop in skin temperature on the real hand (Moseley et al., 2008), or an increased histamine reactivity on the real hand, indicative of a down-regulation of the immune system (Barnsley et al., 2011). Such findings have been taken as evidence for the hypothesis that the brain also decreases the self-attribution of the real hand during the illusion, i.e. that the dummy hand replaces the real hand in the "body representation" (Longo et al., 2008; Moseley et al., 2012).

Importantly, as already noted by Botvinick and Cohen (1998), one fundamental requirement for the RHI to occur is the temporal congruence of the seen and felt touches: All of the effects described above vanish when the touches to the dummy hand and the real hand are delivered asynchronously. Furthermore, the touches need to occur at corresponding locations on the dummy hand and the real hand (Costantini and Haggard, 2007). Botvinick and Cohen therefore speculated that the RHI emerges from the "perceptual matching" of the touch seen on the dummy hand and the touch felt on one's real hand. This interpretation of the RHI has been further developed into the "multisensory hypothesis of body ownership" (Ehrsson, 2012), according to which body ownership is based on the integration of concurrent body-related information from multiple sensory modalities (Ehrsson, 2012; Maravita et al., 2003; Makin et al., 2008; Tsakiris, 2010; Hohwy, 2012; Blanke, 2012). In general, multisensory integration in the brain seems to follow two simple rules introduced by Meredith and Stein (1986; see also Driver and Spence, 2000; Ernst and Bülthoff, 2004; Ma and Pouget, 2008): if two sensory signals come from approximately the same location, or emerge relatively close to one another in time, they are likely to have been caused by the same external event. With these rules, the brain solves the classical "binding problem" (Revensuo and Newman, 1999; Driver and Spence, 2000; Ernst and Bülthoff, 2004), namely, how do I know which of the signals I receive about the external world via different sensory modalities I should bind together—and hence assume that they provide information about the same external thing—and which signals I should treat as pertaining to different things? However, the case of the RHI is special because it involves

information from somatoception, which provides information about *only my own* bodily properties (Bermúdez, 1998; Anderson and Perlis, 2005). Thus the registration of correlations between visual and tactile events is unique to the body—no other object in the world except your body will provide the same correspondence of seen and felt touches—and this process has been called "self-specifying" (Botvinick, 2004). The fact that the RHI so easily exploits these mechanisms renders it such an intriguing paradigm to investigate body ownership.

Figure 1. The rubber hand illusion. In most individuals, the simultaneous tactile stimulation of corresponding anatomical locations on one's real hidden hand and a visible ipsilateral dummy hand (**a**; Botvinick and Cohen, 1998) induces an illusion of feeling the touch at the location where the dummy hand is seen being touched, and often leads to an illusory self-attribution of the dummy hand as part of one's body (**b**). The limits of the RHI seem to be defined by a representation of the body and the space surrounding it, since the RHI cannot be induced if the dummy hand is placed in an anatomically implausible position (**c**; Ehrsson et al., 2004; Tsakiris and Haggard, 2005), and its effect decreases with increasing distance between the real and dummy hand, with approximately 30 cm being a limit within which the illusion is strongly experienced (**d**; Lloyd et al., 2007).

There is a second class of restrictions to the RHI, which suggest that more than mere multisensory correspondence needs to be given for the illusion to occur. The dummy hand needs to be of the same laterality as the real hand (Tsakiris and Haggard, 2005), and it must be placed in an anatomically plausible posture (Ehrsson et al., 2004; Lloyd et al., 2007). Moreover, the effects of the RHI decrease with increasing distance of the dummy hand from the real hand, with 30 cm being a "spatial limit" outside of which the effects of the illusion vanish quickly (Lloyd et al., 2007). These constraints (which have partly also been demonstrated in monkeys by Graziano et al., 2000) have led

to proposals that body ownership results from an interaction between congruent multisensory input and an internal body representation (Graziano and Botvinick, 2002; Maravita et al., 2003; Tsakiris, 2010).

In the search for the neural correlates of the RHI, neuroimaging has emphasized the contribution of frontal and parietal areas. In the first fMRI study using the RHI paradigm, Ehrsson and colleagues (2004) found increased BOLD responses to the RHI in the intraparietal sulcus (IPS) and the ventral premotor cortex (PMv), which responded more to synchronous versus asynchronous stimulation of the hands, and when the dummy hand was positioned in an anatomically plausible posture. The involvement of the IPS and PMv in the RHI has been supported by several publications from Ehrsson's lab (Ehrsson et al., 2007; Petkova et al., 2011; Gentile et al., 2013; but some other studies have failed to replicate these results, Tsakiris et al., 2007). Further brain areas that have been less consistently reported in studies investigating body ownership are the insula (Tsakiris et al., 2007), the somatosensory cortex (Tsakiris et al., 2007; Makin et al., 2008), and the temporo-parietal junction (Ionta et al., 2011; Tsakiris, 2010).

In their pioneering work using electrophysiological recordings in monkeys, Rizzolatti and colleagues (1981, 1997) demonstrated that some neurons in the ventral premotor cortex (PMv) discharged when a visual stimulus was presented within reaching distance of the monkey. Crucially, the visual receptive fields of these neurons were overlapping with their tactile receptive fields. The emerging idea that these neurons code the location of the limb and the space around it was supported by Graziano and colleagues (1994, 1997), who demonstrated that the bimodal receptive fields of such a typical PMv neuron shifted accordingly when the monkey's hand was moved—irrespective of gaze direction. This suggests that the receptive fields of these neurons are body-centered, rather than, for example, retinotopic (Avillac et al., 2007; Graziano and Cooke, 2006), and moreover, that the receptive fields of the somatosensory and visual modalities may be "re-mapped"

onto one another if necessary (e.g. to keep them aligned when the arm moves but vision remains fixed at another location, Macaluso and Maravita, 2010; Holmes and Spence, 2004). Since then, neurons with similar properties have been reported in the intraparietal and posterior parietal cortex (PPC) of the monkey (Graziano, 1999; Graziano et al., 2000; Bremmer et al., 2001a; Avillac et al., 2007, see Graziano and Cooke, 2006, for a review). There is compelling evidence for a similar encoding of the space around the body in premotor and intraparietal areas of the human brain (for reviews see Maravita et al., 2003; Makin et al., 2008; Macaluso and Maravita, 2010): fMRI studies in humans have revealed similar polymodal motion processing in premotor and posterior parietal areas as in monkeys (Bremmer et al., 2001b), the existence of body-centered visuo-tactile maps in the human posterior parietal cortex (Sereno and Huang, 2006, 2014, Lloyd et al., 2002), and preferential responses in human premotor and parietal areas for moving stimuli near one's hand (Makin et al., 2007; Brozzoli, Gentile et al., 2012). Thus, the space around the body (the peripersonal space, PPS; Rizzolatti et al., 1997; Graziano et al., 2000; Làdavas, 2002) indeed seems to be represented in the brain, presumably in the premotor and intraparietal cortex. The basic function of the PPS representation may be, on the one hand, to protect the body from nearby potentially harmful stimuli (Graziano and Botvinick, 2002; Graziano and Cooke, 2006), and on the other hand, to represent objects that can be acted upon (e.g. within reaching distance, Rizzolatti et al., 1981, 1997; Maravita et al., 2003). In both cases, the PPS qualifies as a special attentional space (Graziano and Cooke, 2006), and sensory signals within the PPS should have a special "self-relevance" (Làdavas, 2002; Maravita et al., 2003; Makin et al., 2008). Thus not only our skin defines a boundary between our self and others—the space immediately surrounding our bodies may define another one. The empirical evidence from these studies may well be reconciled with the fact that the illusory self-attribution of body parts engages fronto-parietal areas of the brain (see Makin et al., 2008; Blanke, 2012; Ehrsson, 2012, for reviews).

The multisensory mechanisms that cause the RHI might be well explained in theory as a probabilistic (Bayesian) updating of the body representation (Armel and Ramachandran, 2003; Hohwy and Paton, 2010; Tsakiris, 2010; Apps and Tsakiris, 2014). Still, as recently been pointed out by Ehrsson (2012), there is no full functional model of the dynamic, multisensory information flow during the RHI that could be linked to theories of probabilistic information processing in the brain. The recent years have witnessed the emergence of a large-scale account of cortical information processing focused on the principles of predictive coding (Friston, 2005; Kilner et al., 2007; Friston and Stephan, 2007; Friston and Kiebel, 2009), which could finally help to bridge these gaps—the so-named free energy principle (FEP, Friston, 2005, 2010).

1.2. The free energy principle

Perception has often been compared to hypothesis testing (e.g., Gregory, 1980; Feldman and Friston, 2010; an idea that can be traced back to Helmholtz), whereby "perceptual content *is* the predictions of the currently best hypothesis about the world" (Hohwy, 2013, p.47). Thus perception can be formalized as classical Bayesian inference (Friston, 2010; Hohwy, 2013) based on a prior hypothesis or model of the world and the sensory input, with the ultimate adaptive goal of inferring the causes of the sensory input. The prior probability of the model (not all hypotheses or models are equally likely given our experience with the world), and the fit between the model and the observed data (the likelihood that the data would occur given this model) determine whether we accept this model as the most plausible explanation of the world (formally, the model's posterior or conditional probability).

The FEP entails the Bayesian brain hypothesis, but at the same time provides explanations for how such inference, and the formation of priors necessary for inference, may be implemented in the brain (Friston and Stephan, 2007; Bastos et al., 2012). The FEP is built around the claim that self-

organizing systems operate based on a *hierarchical generative model* of the world (Friston, 2005). In brief, a generative model is a "working hypothesis" about the causal structure of the world, describing a causal process that generates some observable data; hierarchical generative models (HGMs) thereby describe such processes across multiple levels of a hierarchy, where higher levels generate causes at lower levels (Goodman and Tenenbaum, 2014; Friston, 2010). In a HGM, the causal structure of the world is thus implied in the hierarchy of top-down connections, thus every level of this hierarchy reflects "deeper", more abstract inference about the causal regularities in the world (Hohwy, 2013; Friston, 2012; Goodman and Tenenbaum, 2014). Of course, the plausibility of this model needs to be evaluated; the FEP describes a "winner takes all" strategy (Hohwy, 2007), where there is always just one model that is chosen as the currently best explanation of the state of affairs. In statistics, if a model is fitted to the data, the difference between this model and the data is the *error*. The smaller we can make this error, the better the model will fit the data. Similarly, the FEP is built around the claim that any "self-organizing system" must minimize the surprise originating from the world, where "surprise" basically refers to the improbability of some sampled data given a particular HGM of the world. Free energy is a measure derived from information theory that poses an upper bound on surprise, therefore minimizing free energy implies minimizing surprise (Friston, 2010). One neurobiologically plausible form of free energy minimization is prediction error minimization in a *predictive coding* scheme (Srinivasan et al., 1981; Friston, 2005; Friston and Kiebel, 2009; Bastos et al., 2012).

Predictive coding[2] is "simply" an inversion of the generative model with the aim of suppressing the prediction error throughout the model's hierarchy (Kilner et al., 2007). Thus predictive coding maps from sensory data, which are now generated at the lowest level, to

---

2   I am aware that this brief introduction does injustice to the much more complex mathematical background and neurobiological implementation of predictive coding (for this see Friston, 2005, 2010; Friston and Stephan, 2007; Friston and Kiebel, 2009; Bastos et al., 2012).

increasingly complex causes (Bastos et al., 2012). Crucially, the informative quantity that is communicated upwards[3] the model's hierarchy is not the sensory data *per se*, but the data-evoked prediction error, since this unpredicted data is the only information that still needs explanation (Feldman and Friston, 2010). The organism now has two options to suppress such prediction error ("surprise"): It may change the predictions of the next-highest level in the model's hierarchy, or it may change the data itself—this selective sampling of data is known as the *active inference* (AI) principle (Friston, 2010, 2011; Vershure et al., 2003), and is one of the features that make the FEP most appealing to cognitive science and philosophy. First, AI also explains the organism's actions (movements) as resulting from a suppression of prediction errors by the motor system, and hence implies that the brain must also have a model of what caused these actions (Friston, 2010, 2012). Thus the FEP applies the same inferential principles to perception and the causes of action (Kilner et al., 2007; Friston, 2012). Second, because the organism's options to sample data are limited by its phenotype, the FEP (similar to the concept of "affordances", Gibson, 1977) acknowledges the bidirectional role of embodiment: "not only does the agent embody the environment but the environment embodies the agent" (Friston, 2011; see also Hohwy, 2007, 2013; Frith, 2009; Clark, 2013). Finally, by such hierarchical suppression of prediction error in a predictive coding scheme, the HGM's priors can effectively be estimated from the data (in statistical terms, these are thus *empirical priors,* Friston, 2012). One essential implication of this is that ultimately, the organism's generative model will entail, as model evidence, evidence for the existence of the agent itself (Friston, 2011; Limanowski and Blankenburg, 2013). These implications of the FEP for embodied "self-modeling" (Friston, 2011, 2012; Hohwy, 2007; Clark, 2013) will be discussed further in Paper 4, but note how

---

3 The terminology with which feedback and feedforward information flow is assigned to anatomical connections may be initially confusing—it helps to remember the inverse relation between predictive coding and the generative model: In a generative model, the top-down connections from higher to lower levels convey the predictions and the respective bottom-up connections convey the feedback (the error). In a predictive coding scheme, the error is the driving signal that is passed along feedforward connections from lower to higher levels, and the (adjusted) predictions of the model are conveyed along the top-down connections as feedback.

they directly suggest the FEP as a candidate framework to formalize a causal model of minimal selfhood including body ownership and agency.

1.3. Overall aim of this thesis

The aim of this thesis was to explore, in a first step, the possibility that the information flow within the brain network that implements the RHI follows the principles of predictive coding as formalized in the FEP framework. Based on this potential application of the FEP to explain the dynamic information flow within the hierarchical brain network involved in the illusory self-attribution of body parts, this thesis aimed, in a second step, to explore whether the FEP is in principle even compatible with philosophical accounts of self-modeling, and may therefore provide the missing link between models of body ownership and philosophical accounts of minimal selfhood *per se.*

2.   Summary and discussion of papers

In this chapter I shall summarize the three empirical and two theoretical papers that form the core of this dissertation. In the first study (Limanowski, Lutti, and Blankenburg, 2014), we investigated the neuronal correlates of illusory limb ownership using fMRI in combination with a modified version of the RHI paradigm. We used a fully automatic setup featuring two stimulation locations (on the palm and the forearm, see Figure 2), which allowed us to implement a control condition that matched the RHI in terms of temporal synchrony, and to eliminate human interaction from the experimental procedure. Most importantly, in combination with separate functional localizers, we were able to show that BOLD responses within a body-selective region in the visual lateral occipito-temporal cortex (LOC), the so-named extrastriate body area (EBA), increased during the illusion. Moreover, the response in the EBA correlated with the experienced intensity of the RHI, and the somatosensory system increased its functional coupling with the EBA during the illusion.

In the second study (Limanowski and Blankenburg, 2015), we extended on these findings and sought to understand the dynamic interactions within the brain network that implement the mechanisms necessary for the self-attribution of body parts. We used the same fMRI-compatible setup as in Study 1, but employed the classical comparison of temporally congruent versus incongruent stimulation of the dummy arm and the real arm (Botvinick and Cohen, 1998). We were able to replicate the findings of Study 1, and to prove the consistency of the data obtained from both studies using conjunction analyses on both datasets. We observed that a network comprising PMv, IPS, and areas in the LOC, overlapping with the EBA, increased its activity and its functional coupling during the RHI. Using dynamic causal modeling (DCM), we were able to show that the RHI context enhanced the bottom-up connectivity from somatosensory (SII) and visual areas (LOC/EBA) to the multimodal IPS. We interpreted these results, according to predictive coding, as reflecting multisensory prediction errors about the location of one's real hand in space, which are passed on to

hierarchically higher multimodal convergence zones in the IPS. The results of both studies fit well with the spatio-temporal principles of multisensory integration (Meredith and Stein, 1986) and with the assumption of a probabilistic body representation that is continually updated via the integration of body-related information from multiple sensory modalities (Graziano and Botvinick, 2002; Makin et al., 2008; Tsakiris, 2010; Ehrsson, 2012). Crucially, our studies advance on previous publications with the demonstration of mechanisms integrating multisensory information within the PPS that do not only involve the fronto-parietal circuit, but already involve body-selective regions of the LOC. Moreover, they suggest that information flow during the RHI may be thought of as predictive coding in an inverted hierarchical generative model.

In the third study (Wold, Limanowski, Walter, and Blankenburg, 2014), we used repetitive transcranial magnetic stimulation (rTMS) to test whether the left EBA (the body-selective area in LOC implied by Study 1 and Study 2 as contributing to the RHI) was indeed causally involved in the illusory self-attribution of a dummy hand during the RHI. Employing a prominent behavioral measure (the "proprioceptive drift", Botvinick and Cohen, 1998; Tsakiris and Haggard, 2005) of the RHI, we were able to show that, indeed, interfering with the left EBA's activity substantially enhanced the relative misperception of one's real hand location as closer to the location of the dummy hand. This result supported the conclusions of our fMRI studies, namely, that the LOC/EBA is of essential importance for comparing multimodal information, and communicating to the self-attribution network, during the RHI.

In the fourth paper (Limanowski and Blankenburg, 2013), we aimed at situating the existing literature on (illusory) body ownership within the theoretical framework of cortical information processing provided by the FEP (Friston, 2005, 2010), to offer possibilities for the missing link between theoretical explanations of the illusion and general theories of probabilistic information processing in the brain (Ehrsson, 2012; see above). We argued that the assumptions made by the FEP

about probabilistic self-modeling are compatible with those of philosophical accounts of minimal phenomenal selfhood (MPS; Metzinger, 2005), and moreover, that this compatibility may be empirically testable. This combination might thus inform not only models of body ownership, but furnish richer, more complex models of body representation and perhaps even minimal selfhood. The fifth paper (Limanowski, 2014) is a brief follow-up on Paper 4, in which I discuss how working with body ownership illusions may provide insights into MPS *per se*, given the special perceptual role of the body implied by classical phenomenology.

Study 1: The extrastriate body area is involved in illusory limb ownership

As described in the introduction, most models of body ownership rest on the assumption that the self-attribution of body parts is implemented in fronto-parietal areas that integrate convergent body-related information from multiple sensory modalities (Makin et al., 2008; Tsakiris, 2010; Ehrsson, 2012), but they seem to largely ignore the role of the visual modality. Two further potential caveats that such models need to take into account may arise from previous implementations of the RHI paradigm: First, one typically compares synchronous versus asynchronous tactile stimulation of the dummy and the real hand to identify the behavioral and neuronal correlates of the illusion. This serial presentation of seen touch on the dummy hand and felt touch on one's real hand may introduce neuronal effects due to the anticipation of touch (see e.g. Carlsson et al., 2000; Keysers et al., 2010; Kuehn et al., 2012) or due to serial sensory predictions in general (Schubotz & von Cramon, 2002, 2003, 2004). Second, the stimulation of the two hands is usually delivered manually by the experimenter (Ehrsson et al., 2004). Many of the brain areas supposedly involved in the illusory experience during the RHI contain neurons with mirror-like properties, i.e., they also respond to other people's actions (PMv; Ebisch et al., 2008) and sensations (SI/SII; Keysers et al., 2010), or to the mere vision of others' bodies (EBA; Downing et al., 2001). It is conceivable that the visible interaction of the

experimenter's hands with "one's hand" (i.e., the dummy hand during the RHI but not during the control condition) might bias the neuronal responses in such areas.

In this study (Limanowski, Lutti, and Blankenburg, 2014), we sought to address all of these potential confounds. We designed a fully automatic RHI setup featuring two anatomically distinct stimulation locations on a full-size right dummy arm (Figure 2). The computer-controlled experimental setup effectively eliminated human interaction (and thus all potential influences of the experimenter) from the illusion induction. Moreover, the multi stimulation-site design allowed us to implement a new control condition in which stimulation was delivered in synchrony to anatomically incongruent locations of the dummy arm and the real arm, thus matching the RHI condition in terms of temporal congruence. Participants lay inside the scanner facing the dummy arm in full view, and were required to maintain fixation on a small black dot in the middle of the dummy arm; their real arm was completely hidden from view in a corresponding posture about 13 cm behind the dummy arm. Stimulation was delivered in blocks of 17.9 s length (11.2 s rest) by back-and-forth brush strokes at about 1.3 Hz, whereby each brush was driven by a separate computer-controlled stepping motor that transmitted its torque into the scanner room via nonmagnetic cables (Figure 1). FMRI data (3T Siemens Tim Trio, 32 channel head coil, custom 3D-EPI sequence with 2 mm³ isotropic voxels at 2240 ms volume acquisition time) were recorded from 20 participants, each completing five RHI runs and one localizer run, which was used to verify the somatotopical representation of the stimulation sites (palm versus forearm) in SI. Each condition was modeled as a regressor in the first-level GLMs, and the resulting contrast images were entered into a flexible factorial GLM for the second-level analysis. To quantify individual differences in the susceptibility to the RHI, we obtained verbal ratings of the strength and the prevalence (onset and duration) of the ownership illusion in the RHI and control conditions from our participants after scanning.
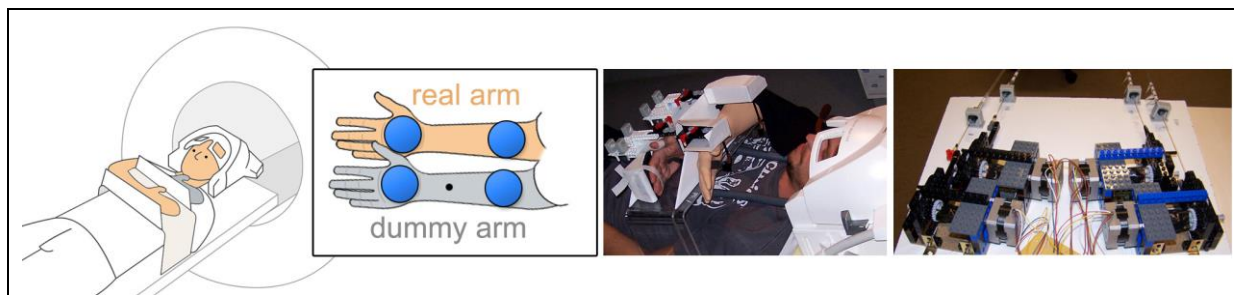
Figure 2. Automated experimental setup used in the fMRI studies. A custom console was used to deliver tactile stimulation to the real arm and the dummy arm. Participants were lying inside the fMRI scanner with their own arm hidden from view behind a realistic ipsilateral dummy arm, which was fully visible. While participants maintained fixation onto the middle of the dummy arm (marked with a black dot), stimulation could occur in various combinations at two anatomically distinct locations on each arm (palm and forearm, schematically depicted with blue circles). Stimulation consisted of back-and-forth brush strokes, whereby the brushes were driven by four identical computer-controlled stepping motors placed outside the scanner room (right panel).

The behavioral ratings affirmed that the automated induction of the RHI produced a strong ownership illusion in our participants, while the control condition did not induce such an illusion. Thus we demonstrated that synchronous visuo-tactile stimulation only elicits the RHI when it occurs on the same body part. In the fMRI data, the main effect of the RHI versus control conditions was observable as significant BOLD signal increase in voxels within bilateral anterior insula and the left LOC, centered on the middle occipital gyrus (mOCG). The cluster of RHI-related activity we found in left mOCG matched previously reported coordinates of the EBA. Since the EBA is defined functionally by its preferred response to human bodies, we had our participants (four drop-outs) take part in an additional functional localizer of the EBA consisting of visual presentation of human body parts (hands and feet) versus control objects (motorcycle parts), following the procedure described by Urgesi et al. (2007). The contrast of body parts versus objects revealed strong BOLD signal increases in bilateral LOC, as expected. Crucially, the RHI-responsive cluster in left LOC identified in the main GLM analysis was almost entirely contained within the body-selective EBA as independently localized. Moreover, a whole-brain regression revealed that participants' left EBA activity during the RHI versus control condition correlated significantly positively with a self-report measure reflecting the illusory

self-attribution of the dummy arm. Finally, we were also able to demonstrate an increased functional coupling between left primary somatosensory cortex (SI) and bilateral EBA during the RHI versus control condition, and found evidence for a similarly increased coupling of the left EBA with the supramarginal gyrus and the posterior parietal cortex, multimodal areas that had been implied by previous RHI work (e.g. Brozzoli, Gentile et al., 2012).

In sum, these results demonstrate an involvement of the body-selective left EBA in the illusory self-attribution of a right dummy arm following synchronous stimulation of anatomically congruent versus incongruent locations of the two arms. It seems reasonable that a visual area contributing to the network underlying the self-identification with a body should be body-selective (see Jeannerod, 2004, 2007; Astafiev et al., 2004; Peelen and Downing, 2007). We were able to draw such inferences because (i) we implemented sample-specific functional localizations of the EBA in each of our fMRI studies, and (ii), because our automatic setup effectively eliminated the actions and body parts of the human experimenter that might have biased previous studies. Activity differences within the left EBA reflected not only the perceptual difference between the RHI and control condition on the group level, but also between-participant differences in the strength and prevalence of the experienced ownership illusion. Moreover, we observed an increased functional coupling between visual EBA and somatosensory and multimodal areas in the parietal cortex during the illusion. These effects were observable despite the fact that the arm was fully visible throughout the whole experiment. These results imply a more complex function of the EBA that goes beyond the mere processing of visual features of human bodies, and add support for the involvement of visual areas in the self-attribution of body parts.

Study 2: Network activity underlying the illusory self-attribution of a dummy arm

The results of Study 1 (Limanowski et al., 2014) had suggested that the illusory self-attribution of a dummy arm engages the body-selective EBA, and had moreover provided evidence for an interaction of the somatosensory and multimodal parietal cortex with the EBA. These findings added visual body-selective LOC/EBA to the candidate brain regions of the brain network underlying the RHI. An involvement of not only frontal and parietal cortex (Ehrsson et al., 2004; Petkova et al., 2011), but also body-selective extrastriate visual areas in the RHI is very plausible, given for example the strong cross-modal effects of vision of the body on tactile information processing (Kennett et al., 2001; Taylor-Clarke et al., 2002; Haggard et al., 2007) and PPS processing (Graziano et al., 2000; Makin et al., 2007; see Introduction). The reported BOLD signal increases in the PMv and IPS (Ehrsson et al., 2004; Petkova et al., 2011) and the LOC/EBA during the RHI (Limanowski et al., 2014), as well as the illusion-related increases in the functional coupling among these areas (Gentile et al., 2013; Limanowski et al., 2014) strongly suggest that the illusory self-attribution of the dummy limb depends on interactions among the pathways of this network. However, neither simple BOLD contrasts nor analyses based on BOLD signal correlations between voxels of the seed region and the rest of the brain (PPIs) allow any clear inferences about the directionality of the interactions within a given brain network. In the present study (Study 2, Limanowski and Blankenburg, 2015), we aimed at contributing to a deeper understanding of the mechanisms within the suggested brain network by which the brain infers whether to self-attribute the dummy limb or not.

We used the same setup (automated induction of the RHI at two possible stimulation locations of the right arm) and fMRI pulse sequence (3D-EPI, 2mm³ functional resolution, now at a volume acquisition time of 2520 ms) as in Study 1, whereby this time, we induced the RHI via synchronous versus asynchronous (control condition) stimulation of the corresponding locations on the dummy arm and the real arm (Botvinick and Cohen, 1998; Ehrsson et al., 2004; Tsakiris and

Haggard, 2005; Petkova et al., 2011). This gave us three possible congruent versus incongruent comparisons: at the palm location, at the forearm location, and at both locations. 20 participants received stimulation via back-and-forth brush strokes, block-wise for 20.16 s (12.6 s rest) at a frequency of about 1.3 Hz, while in the asynchronous conditions we introduced a delay of 50 % (about 400 ms) between the seen and felt brush strokes. Again, based on our hypothesis that body-selective regions would contribute to the RHI, we implemented a functional EBA localizer (see Study 1). One aim of this study was to try to replicate the behavioral effects, and more importantly, the involvement of the EBA in the RHI observed in Study 1, this time via synchronous versus asynchronous tactile stimulation of corresponding locations of the dummy arm and the real arm. Specifically, we wanted to demonstrate the comparability of the results of the present experiment (spatial congruence while manipulating the temporal congruence of visuo-tactile stimulation) with those from Study 1 (Limanowski et al., 2014: temporal congruence while manipulating the spatial congruence of visuo-tactile stimulation). Therefore, we jointly preprocessed both datasets, and sought to identify brain areas whose BOLD signal would increase during spatially congruent versus incongruent stimulation (Study 1) *and* during temporally congruent versus incongruent stimulation of the dummy arm and the real arm (present study). To test for significantly consistent responses to congruent versus incongruent visuo-tactile stimulation in brain regions across stimulation locations in the present experiment (palm, forearm, or both) and for the comparison with the Study 1 dataset (spatial or temporal incongruence at a given location), we used global conjunction analyses, which reveals voxels whose activity or activity difference is consistent across several contrasts.

These analyses revealed a predominantly left-lateralized network comprising the LOC (as expected, contained within the body-selective EBA), IPS, and PMv, which responded consistently more strongly to congruent versus incongruent stimulation of the dummy arm and the real arm across stimulation locations. The comparison with the Study 1 dataset confirmed that the left LOC/EBA and the left IPS moreover showed such consistent effects across stimulation locations *and*

type of (in)congruence (spatial or temporal). We also replicated the positive correlation of the left EBA's response to congruent versus incongruent stimulation with the reported intensity of illusory self-attribution, for the present dataset, and again, also in a conjunction with the data from Study 1. Corresponding conjunctions of PPI analyses (each PPI was performed separately for each congruent versus incongruent contrast at each stimulation location) revealed that during congruent versus incongruent visuo-tactile stimulation, the left PMv and the left IPS consistently increased their functional coupling with the LOC/EBA. These analyses strongly supported our assumptions, and therefore we were confident to analyze the directionality of the illusion-related interactions within the suggested network using dynamic causal modeling (DCM). Random effects Bayesian model selection (RFX BMS) identified as the most plausible intrinsic architecture of the SII-LOC-IPS-PMv network (we included SII as the input area for somatosensory information) a model with bidirectional connections between SII and IPS, LOC and IPS, and IPS and PMv, whereby sensory input entered SII (assumed somatosensory input) and LOC (assumed visual input). Based on this architecture, which is plausible given the knowledge about the RHI and PPS processing, we created a model space comprising models with illusion-related modulations of bottom-up or top-down connectivity, or both (The DCM analysis was performed separately for each stimulation location). RFX BMS suggested the same model as the most probable explanation of the data, featuring a modulation of the bottom-up connections from SII to IPS and from LOC to IPS under the illusion. The analysis of the parameter estimates of these connections showed that these were significantly enhanced in each case. Again, as a proof of concept, we performed the same DCM analysis on the dataset from Study 1, which yielded the same winning model in both cases (although the parameter estimates from LOC to IPS were not consistent and did not reach significance). However, in sum the DCM analyses strongly support the univariate analyses and consistently point to a modulation of bottom-up connectivity from somatosensory and visual areas to the IPS during the RHI.

We interpreted these results within a predictive coding scheme of the RHI, in which a prediction error stemming from the ambiguous position of one's arm in space is generated in the LOC and SII, and is passed on to the IPS via feedforward connections. Thereby the first prediction error would be generated within the LOC, which first associates or binds the seen and felt touch (the somatosensory information being presumably communicated already via the IPS) due to their coincidence on a body part within peripersonal space, and then registers the mismatch between the visual (dummy arm-centered) arm representation and that predicted by the IPS. This prediction error may then be communicated to the IPS, which may try to suppress it by adjusting its predictions about one's arm's position—biased by the dominant visual modality. These predictions may then be sent down to LOC (where they may "explain away" the prediction error) and also to SII—there, they do *not* match the somatosensory information about one's arm's position, and may therefore evoke another prediction error. Such an interplay of prediction errors could be ongoing throughout the illusion; indeed, we found some supporting evidence for this interpretation in the fact that activity within the left SII increased more over the time of stimulation during the congruent than during incongruent stimulation. The intrinsic connectivity of the DCMs featured significantly enhanced connectivity from the IPS to the PMv during visuo-tactile stimulation of the two arms *per se*, i.e., in the congruent and incongruent conditions versus baseline. This supports the presumed directionality of information flow from parietal to frontal areas during PPS processing (e.g. Graziano and Cooke, 2006; Makin et al., 2008), but does not directly suggest significantly different parieto-frontal communication during congruent versus incongruent visuo-tactile information. Perhaps, our experimental design captured the first enabling step of the RHI, namely, the visual capture of touch (Pavani et al., 2000), while other processes involving the PMv might either take more time (e.g. Petkova et al., 2011, had used longer stimulation blocks) or a different setup (in e.g. Ehrsson et al., 2004, the hands were placed next to each other, not in the line of sight). In models of PPS processing, the PMv receives information from the IPS to initiate defensive behavior (Graziano and Cooke, 2006); models of the RHI assign functions

25

such as "representation of the dummy hand for action" to the PMv (Makin et al., 2008). Whereas our data suggest that the IPS may signal a higher-level prediction error to the PMv (according to the assumptions of predictive coding), the contribution of PMv and its information flow remains to be clarified by future research.

In sum, Study 2 replicated and extended the results of Study 1 (Limanowski et al., 2014). Our results contribute to the understanding of the EBA's role in body perception. These results support the assumed role of body-selective visual LOC/EBA and its interaction with the fronto-parietal circuit implied by previous studies during the illusory self-attribution of body parts; moreover, we have provided first evidence that interactions within the network activated during illusory body ownership may follow a predictive coding scheme.

Study 3: Proprioceptive drift in the rubber hand illusion is intensified following 1 Hz TMS of the left EBA

This behavioral experiment was designed to follow up on the novel observation of BOLD signal increases in visually body-selective left extrastriate body area (EBA) during the illusory self-attribution of a dummy arm (Study 1: Limanowski et al., 2014; see also Study 2: Limanowski and Blankenburg, 2015). Our hypothesis was that if the involvement of the EBA in the network underlying the self-attribution of body parts was indeed of practical significance, then a lesion to this area should substantially alter the effects of the RHI. To test for such a causal role of the EBA in the RHI, we applied repetitive transcranial magnetic stimulation (rTMS) over this region. TMS is a non-invasive means of temporally interfering with neuronal activity in a pre-defined brain region. In this way, the researcher may test for the causal involvement of this region in a given mechanism, which makes TMS a welcome method to examine any assumptions derived from neuronal models based on functional

neuroimaging. Previous studies had applied TMS over the EBA to disentangle its role from that of the premotor cortex in action processing (Urgesi et al., 2007), but we found no publication reporting TMS over the EBA (or occipito-temporal cortex in general) during a RHI paradigm.

Before the TMS experiment, we acquired fMRI data using the functional EBA localizer described in Study 1 (Limanowski et al., 2014), and used them to identify each participant's peak body-selective region in left extrastriate cortex. We applied rTMS over these left EBAs (1200 pulses at 1 Hz over the course of 20 minutes per stimulation block), whereby we used a stimulation intensity of 80 % of the motor threshold to interfere with EBA activity, and 40 % of the motor threshold as a "sham" stimulation intended as a control condition against which the effects of rTMS were compared. After rTMS or sham stimulation, participants completed four 3.5 minute-stimulation blocks, two featuring synchronous touch to the real and the dummy hand (RHI condition), and two featuring asynchronous touches (control condition). We quantified the intensity of the illusory self-attribution of the dummy hand using a prominent behavioral measure of the RHI, the so-named "proprioceptive drift" (Tsakiris and Haggard, 2005; Botvinick and Cohen, 1998). This measure assesses the relative mislocalization of one's real hand towards the location of the dummy hand following synchronous, as compared to asynchronous stimulation of the dummy hand and the real hand. Typically, a proprioceptive drift of about 30 % of the distance between the real and the dummy hand is reported, which is interpreted as evidence for a recalibration of the body-centered proprioceptive reference frame onto, or at least towards, the dummy hand. In our case, the distance between the dummy hand and the real hand was approximately 20 cm. Proprioceptive drift was assessed after each run as the difference between the reported perceived location of the participant's real hand after stimulation, as compared with a pre-stimulation baseline. Thus we accounted for default between-participant differences in the ability to localize one's hand without direct vision. This design also nicely complemented our fMRI analyses (Study 1 and 2), during which no assessment of proprioceptive drift was possible due to restricted space. In addition, we also assessed the intensity

of the illusion via the standard verbal self-attribution ratings (Botvinick and Cohen, 1998), and the participant's on-line indication of the time point during stimulation, at which this illusion was first perceived (Ehrsson et al., 2004).

Our main finding was that on average, rTMS (versus sham stimulation) over the left EBA significantly increased the relative mislocalization of one's real hand towards the location of the dummy hand (the proprioceptive drift) in the RHI versus control condition. This result demonstrates that interfering with neuronal activity in the left EBA enhances the illusory self-attribution of a dummy hand during the RHI, as revealed by a prominent behavioral measure. We did not observe similar interactions for the other two measures, the ownership ratings and the reported illusion onsets. These measures were significantly affected by synchronous versus asynchronous touch, but showed no differential effect of rTMS versus sham stimulation.

The results of this study should be further pursued. While the increased behavioral effect of the RHI following a disruption of EBA activity is plausible (see the discussion of that study), a similar study by Kammers et al. (2009) found a *decreased* RHI after rTMS over the left inferior parietal lobule. Notably, the authors also found that rTMS only affected the mislocalization of one's real hand but not the self-attribution ratings. One potential explanation for the dissociation of the measures observed in our data and by Kammers et al. (2009) is that the targeted areas do not underlie the illusory experience of "ownership" of the dummy hand, but only the resolution of the visuo-somatosensory conflict. This is an interesting option, which, however, needs to be clarified by future research. However, the results of Kammers et al. and our data are compatible with the hierarchical inferential structure of the RHI network as discussed above. Thus, while interfering with a lower-level area may hinder the generation of prediction errors evoked by the ambiguous input arising from the illusion (and therefore enhance the illusion), interfering with higher-level inferior parietal areas could well prevent the "resolution" of the intersensory conflict by recalibration of one's hand's position (and

therefore reduce the proprioceptive drift). However, it could also be that with the rTMS intervention we had targeted a whole different process in the EBA, such as for example a purely visual distinction between one's own and others' body parts. There have been some reports that during the RHI, the dummy hand is perceived as more similar to one's real hand ("perceptual assimilation", Longo et al., 2008). Interfering with the EBA's body part perception could have enhanced this assimilation and therefore increased the illusion. These options need to be clarified using more specific study designs, using for example non-hand objects in addition to dummy hands, and possibly by targeting not only the EBA but also the IPS in the same experiment.

In sum, the enhanced behavioral effect of the RHI following rTMS over the left EBA fits well with the results of our two fMRI studies (Limanowski et al., 2014; Limanowski and Blankenburg, 2015), which had both demonstrated the involvement of the left EBA in a brain network underlying the self-attribution of body parts. In Study 1, we had discussed a possible role of EBA being the detection of multimodal prediction errors during the RHI, in accordance with previous speculations about the EBA's role (Jeannerod, 2004; Apps and Tsakiris, 2014). Study 2 provided evidence supporting this speculation. The enhanced behavioral effects of the RHI following rTMS over the EBA observed in this study may thus be interpreted as the result of an interference with the EBA's ability to detect such prediction error, and to pass it on to hierarchically higher regions like the IPS. In other words, if the EBA is not able to detect and communicate mismatches (prediction errors) between the self-model's predictions and the sensory input, then the ambiguous sensory input could be easier reconciled with the self-representation, perhaps leading to a facilitated self-attribution of the dummy hand.

Paper 4: Minimal self-models and the free energy principle

This theoretical article was inspired by the—in our opinion apparent—compatibility of the assumptions made by the free energy framework about probabilistic self-modeling (Friston, 2010, 2011), and philosophical accounts of MPS (Metzinger, 2004, 2005; Blanke and Metzinger, 2009). Our first aim was to show that these two approaches share key assumptions about the nature and the mechanisms underlying MPS. Second, we wanted to provide an overview of recent applications of predictive coding schemes (or such proposals) to key aspects of MPS, including first and foremost, multisensory integration and the sense of body ownership, but also the first-person perspective of subjective experience, the "mineness" of actions and sensations, the special status of interoceptive information, and finally, also the modeling of other selves. The FEP has proven valuable in explaining mechanisms of low-level perception (Mumford, 1992; Rao and Ballard, 1999; Murray et al., 2002) and action (Kilner et al., 2007; Friston, 2012; see also Vershure et al., 2003), but it remains to be seen whether it can provide similar insights into more complex mechanisms underlying self-modeling (Friston, 2011, 2012).

Thomas Metzinger's conceptualization of MPS (Metzinger, 2004, 2005, 2013; Gallese and Metzinger, 2003; Blanke and Metzinger, 2009) is one of the philosophical accounts of minimal selfhood mentioned in the introduction to this thesis. These accounts try to identify the necessary and sufficient properties of the most basic, pre-reflective kind of self-awareness possible. The MPS account is a relatively formal one (which makes it attractive for the natural sciences), built upon the idea of the brain as a representational system that simulates a model of the world, in order to explain or reduce ambiguity of the sensory input coming from the world. This idea is also one of the FEP's key tenets (Friston, 2005, 2010, 2011, 2012, see Introduction). Moreover, the MPS account emphasizes the special role of the body as the self-representation of the system, upon which its world-model is spatially, temporally, and phenomenally centered. In this way the body defines the "perspectivalness"

of experience (Metzinger, 2004). Crucially, the system-model does not recognize itself *as* a model (Metzinger, 2004, 2005), and thus the model is phenomenally "transparent" to the subject of experience. In philosophical terms this means that only the contents, not the mechanisms of the system-model are accessible to introspection; "You *are* the content of your PSM [phenomenal self-model]", as Metzinger (2005) has put it. All these assumptions are again found in the FEP, which by its nature admits of the central role of the body in the world-model, as the sensory apparatus that enables being and interacting with the world (Friston, 2012, see Introduction): In a hierarchical generative model that, by minimizing prediction error across its hierarchy, tries to infer the causes of its sensory input originating in the world, this sensory apparatus necessarily has to be included. It has to be included as part of the world, but at the same time this sensory apparatus enables being in and interacting with the world (and thus a whole new way of minimizing prediction error as described by the active inference principle, Vershure et al., 2003; Friston, 2010, 2011, see Introduction). Thus the FEP acknowledges the "paradox" role of the body as pointed out by phenomenology (Merleau-Ponty, 1945; Legrand, 2006; see Paper 5 below). Even the phenomenal transparency of the self-model can be found in the FEP: Because with ascending hierarchy, causes of the sensory data are modeled more and more abstractly, they also become more and more cognitively inaccessible (Hohwy, 2007, 2010; Kiebel et al., 2008). In sum, the FEP and MPS accounts both share the view that the coherent subject of experience is the result of a dynamic self-modeling process that tries to explain—to predict—the world to reduce surprise.

Building upon this, the second part of our paper was therefore concerned with trying to outline existing or potential applications of predictive coding within the FEP to the individual key properties of MPS (Blanke and Metzinger, 2009) and their proposed implementing mechanisms, including multisensory integration, the first-person perspective of subjective experience, the "mineness" of actions and sensations, the special role of interoceptive information, and the modeling of other selves. A full summary of these would mean rehashing large parts of the paper. Hence I will

only briefly highlight the explanations that such accounts offer for the mechanisms underlying the self-attribution of the body: The self-representation (or body representation) is thus a result of hierarchical prediction error suppression along an inverted generative model, where higher levels try to suppress prediction error in lower levels. If an error cannot be explained at a given level, it will be passed upwards—until in the worst case, the whole model may have to be abandoned in favor of a better explanation of the world and the organism itself (Friston, 2011; Hohwy, 2010). However, if the error is successfully explained away at a given level, there is no need to adjust higher-level representations. The RHI may be thus explained by a successful suppression of prediction error about one's hand's position in space at lower and intermediate levels: the sensory data suggest that touch happens at two different arms, which is explained away by re-mapping the sensory reference frames so that they correspond to one another. This then makes it unnecessary to adjust any higher-level representation about the characteristics of one's body, for example, that it has only one right arm (Hohwy, 2010; Friston, 2011, 2012). This strong (unchanged) prior assumption about a one-right-armed body thus renders an alternative model, which would infer that touch information is in fact coming from two different right arms (the dummy arm and the real arm) at once, unlikely (Hohwy, 2010, 2012).

To conclude, in this paper we presented an argument for the compatibility of predictive coding accounts with conceptualizations of MPS; we moreover proposed that predictive coding schemes can also formally explain key mechanisms underlying MPS, such as the (illusory) self-attribution of body parts, as the result of a dynamic process across an (inverted) hierarchical generative self-model. The suggestions offered for the mechanisms underlying the RHI are very well supported by the results of Studies 1-3 (Limanowski et al., 2014; Limanowski and Blankenburg, 2015; Wold, Limanowski, Walter, and Blankenburg, 2014), most notably, by the enhanced forward-flowing connections from lower sensory to higher multimodal areas during the RHI reported in Study 2, which

may be interpreted as a prediction error informing the self-model that some of its predictions need adjustment for an unsurprising world-perception.

Paper 5: What can body ownership illusions tell us about minimal phenomenal selfhood?

This paper was a brief follow-up on Paper 4 (Limanowski and Blankenburg, 2013) that still might raise some interesting points. The insights into the mechanisms underlying the self-attribution or "ownership" of body parts derived from working with the RHI and its extensions are meant to inform us about the nature of MPS. However, MPS is conceptualized as a global and unitary experience of a coherent self (Gallagher, 2000; Metzinger, 2005; Blanke and Metzinger, 2009), and a similarly coherent, yet probabilistic self-model is also implied by the FEP (Friston, 2010, 2011; Hohwy, 2013, see Paper 4). Body ownership illusions (BOIs) such as the RHI are designed to target individual properties of the self-model, such as the location of one's limbs in space by introducing some intersensory conflict. The combination of philosophical accounts and the FEP (Paper 4) very elegantly accounts for why such changes in low-level properties of the system, which can be accounted for by adjusting correspondingly low-level representations, may not affect the model (or the system's self-identification with it) *per se*. Thus when using BOIs to draw inferences about MPS, one should keep in mind that these paradigms first and foremost target individual properties of the self-model. There may be a fruitful link to classical phenomenology (Husserl, Merleau-Ponty, and others) with its dual conceptualization of the body, distinguishing between the "objective body" and the "lived body" (an "experientially absent" body, which is not perceived as a thing in the world while we live and act through it, Gallagher, 1986; Gallagher and Zahavi, 2010). The FEP recognizes this dual role of the body (see Paper 4 above). Hence following this definition, it may be helpful to investigate whether BOIs, with their focus on the physical properties of the body, may in fact not target the "lived body" but the

33

body as perceived. I conclude that BOIs may provide deep insights into the construction and representation processes of some features of MPS, but it should be formally clarified about which levels of a generative self-model as proposed in Paper 4 they really can inform us.

3.   General discussion

The main aim of the present thesis was the evaluation of the FEP as a potential framework in which a generative model of body ownership, exemplified by the case of the RHI, may be formalized—with the overarching goal of contributing to a model of minimal selfhood. The fMRI studies presented in this thesis (Study 1 and 2) used the automatic, spatio-temporally congruent stimulation of a right dummy arm together with one's real counterpart to induce an illusory self-attribution of the dummy arm (the RHI). In conjunction, the results of these studies suggest that the neural mechanisms underlying this illusion are implemented in a predominantly left-lateralized network comprising LOC, IPS, and PMv. Thereby, using independent functional localizers, we showed that the regions of LOC that responded to the illusion showed strong preferential responses to vision of the body, thus corresponding to the EBA. Using PPI analyses, we demonstrated that the PMv and IPS increasingly interacted with the LOC/EBA during the illusion (Study 1 and 2), and DCM suggested that the most likely explanation for the illusion-related modulations within the RHI network were enhanced bottom-up connections from SII and LOC/EBA to the IPS (Study 2). In a TMS experiment (Study 3), we proved a causal involvement of the body-selective EBA in the RHI by demonstrating that temporarily interfering with its activity enhanced the illusion-related mislocalization of one's real hand towards the dummy hand.

Taken together, the results of the empirical research articles presented here (summarized below in Figure 3) provide evidence that the brain's interpretation of the ambiguous sensory input provided by the RHI follows a predictive coding scheme in a hierarchical network. In the theoretical Papers 4 and 5, I argued that the self-attribution of body parts may be conceived of as hierarchical inference operating along an inverted generative model according to the principles of predictive coding as formalized in the FEP, and that such an interpretation is compatible with philosophical

accounts of MPS. I concluded that the FEP may help explain body ownership illusions, and perhaps, in the long run, may thus also furnish a larger-scale generative model of minimal selfhood.

3.1. The contribution of visual body-selective areas to the rubber hand illusion

The RHI has become such a popular paradigm because it is a surprisingly simple, yet effective way of interfering with the brain's mechanisms underlying the sense of body ownership, one of the fundamental properties of MPS (Gallagher, 2000; Blanke and Metzinger, 2009). Explanations of the mechanisms underlying the RHI emphasize the resolution of a conflict between vision and somatosensation (Makin et al., 2008; Tsakiris, 2010; Ehrsson, 2012), in line with Bovinick and Cohen's (1998) first interpretation of the illusion. Such a conflict arises because the touch seen on the dummy hand and the touch felt on one's real hand seemingly correspond, but visual and proprioceptive information about the location of the touched hand is contradicting.

One potential shortcoming of existing RHI accounts is that they largely focus on the fronto-parietal multimodal areas. In monkeys, the PPC constructs a body- or body-part-centered (in contrast to for example retinotopic or allocentric) spatial representation of the world based on integrated multisensory information (Graziano and Cooke, 2006; Avillac et al., 2007). Thereby areas in and around the IPS seem to be of particular importance, as they receive inputs from extrastriate visual areas (but not from early visual areas), and from primary and secondary somatosensory cortices (Graziano and Botvinick, 2002; Beauchamp et al., 2007, 2010; Ehrsson, 2012), and communicates to premotor areas (Graziano and Cooke, 2006). Based on these findings it has been argued that during the RHI, the function of the IPS may be a re-mapping of the PPS onto the dummy hand (Brozzoli, Gentile et al., 2012) and a communication of such information to the PMv (Makin et al., 2008). In a model of information flow during the RHI proposed by Makin and colleagues (2008), the IPS receives information from the visual and somatosensory cortex, and is suggested to integrate this information

into a common hand-centered reference frame, while the PMv is associated with "representation of the dummy hand for action". Similarly, in Tsakiris' (2010) model, the posterior parietal cortex (PPC, supposedly comprising the IPS) compares somatosensory and visual information about touch, while "touch referral" to the dummy hand is assigned to the PMv. Although the models differ in other neuroanatomical and functional aspects (e.g. Tsakiris proposes an involvement of the insula and temporo-parietal junction), they share the assumption that visual and somatosensory information interact in the parietal cortex.

The proposed importance of multimodal areas such as the PMv and the IPS for body ownership (e.g. by localizing one's body parts in space) is compatible with the notion of a hierarchical generative model following the assumptions of the FEP. However, the FEP emphasizes the suppression of prediction error throughout the whole predictive hierarchy, reaching down to the individual senses. A full model of the RHI in particular, or of body ownership or MPS in general, would hence also have to include the mechanisms at the individual sensory modalities.

The papers I have presented here support the assumption of such a hierarchical model. They suggest that the integration of visuo-tactile information that underlies the self-attribution of the dummy hand already begins in the LOC. Thereby according to our interpretation, the LOC would already associate the seen and felt touches as one coherent event occurring on a body part within PPS (a role that was previously assigned to the IPS, Makin et al., 2008). This binding based on the learned associations between somatosensory and visual stimuli on the body (Botvinick, 2004; Anderson and Perlis, 2005) is thus an example of low-level predictive coding of bodily features. Crucially, in our model the LOC then generates a prediction error that informs the hierarchically higher IPS to adjust its predictions about the location of one's arm in space.

The contribution of visual LOC to the generative body ownership model implied by our results is supported by evidence from other lines of research: the mere vision of a body part can substantially

enhance the processing of tactile stimuli on this body part ("visual enhancement of touch", Kennett et al., 2001; Taylor-Clarke et al., 2002; Whiteley et al., 2004; Haggard et al., 2007), and the processing of visual stimuli in the PPS around it (Graziano et al., 2000; Làdavas, 2002; Makin et al., 2007). Moreover, the fact that during the RHI, the touch is felt where I see it (on the dummy hand, not where my hand is actually located) suggests a dominance, or stronger weighting, of vision over proprioception ("visual capture of touch", Pavani et al., 2000; Holmes and Spence, 2004; Lloyd et al., 2007; Makin et al., 2008). There is strong evidence for information transfer between the somatosensory cortex and extrastriate visual cortex via indirect pathways involving the IPS (Grefkes et al., 2002, Grefkes and Fink, 2005; Avillac et al., 2007; Macaluso and Maravita, 2010) and corresponding cross-modal effects of touch and vision in LOC (Macaluso and Driver, 2001; Beauchamp et al., 2007, 2010). It has been proposed that within the LOC, the EBA (Downing et al., 2001) is perhaps the most plausible candidate where such modulatory effects of vision of the body on somatosensation could be implemented (Haggard et al., 2007; Costantini et al., 2011; Blanke, 2012). In fact, Blanke (2012) had speculated about comparable multimodal neurons in the LOC/EBA as have been demonstrated in the IPS.
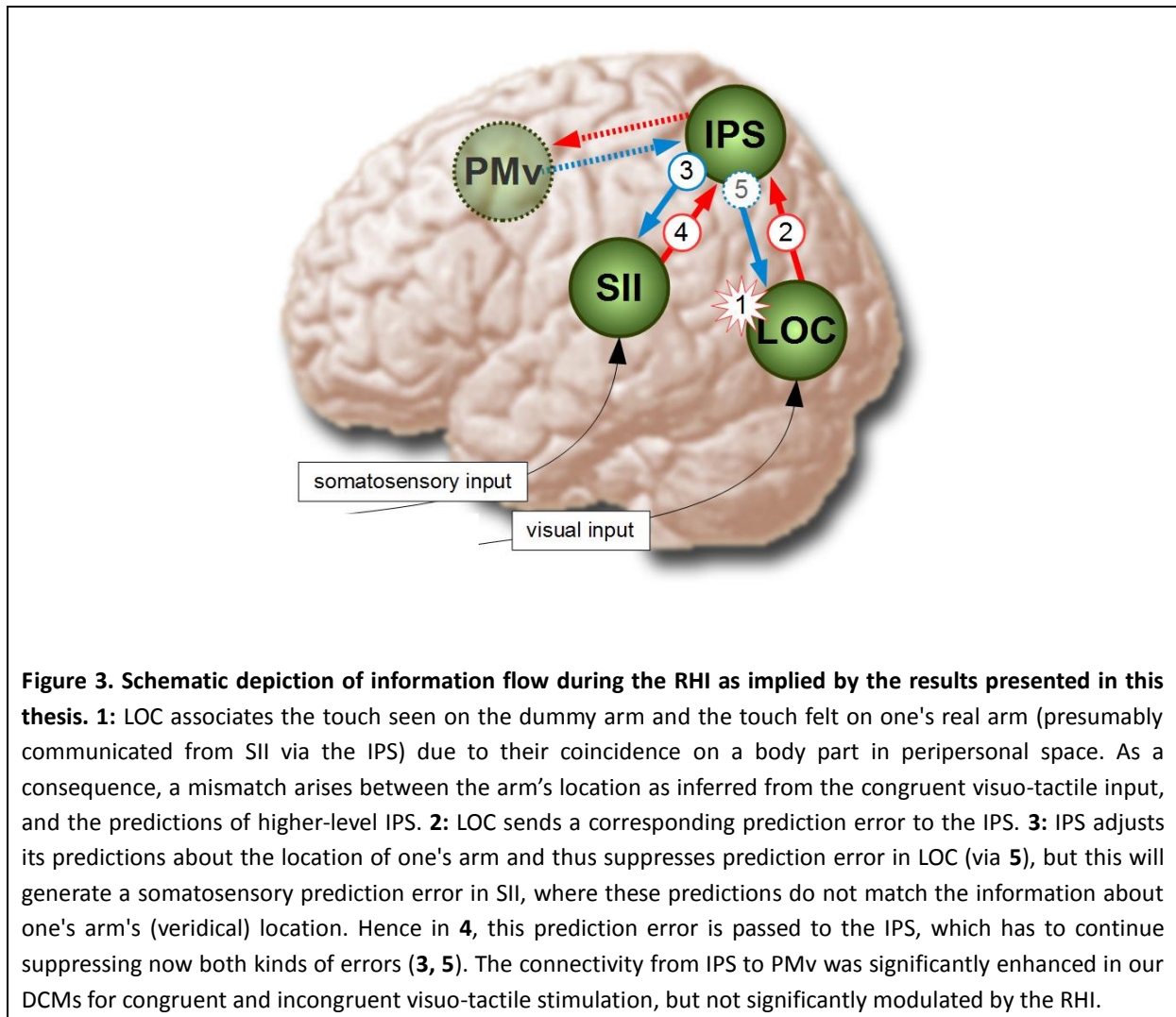
The EBA was initially identified and reported (Downing et al., 2001) because it showed a remarkable preferential response to the visual presentation of human body parts and full bodies versus all other sorts of visual stimuli (including faces, interestingly). However, by now, converging evidence suggests that the EBA is in fact a "supramodal" region (Kitada et al., 2014): For example, even in blind people the EBA responds to the haptic exploration of body shapes (Kitada et al., 2009, 2014), and to "seeing" body shapes with an auditory sensory substitution device (Striem-Amit and Amedi, 2014). Still the most intriguing evidence for a function of EBA that goes beyond the purely unimodal processing of visual body shape comes from a study by Astafiev et al. (2004), who showed that the EBA responded to seen and unseen movements of one's arm, strongly suggesting that the EBA also processes proprioceptive information. Building upon this, other studies have repeatedly

implied a role of the EBA in action processing (Jackson et al., 2006; David et al., 2007; Bracci et al., 2010; Saygin et al., 2012; van Elk, 2014). Notably, the detection of multimodal mismatches and the generation of corresponding prediction errors in the EBA has been implied speculatively by authors using different paradigms: Saygin et al. (2012) found that the EBA responded to violations of predictions about biological motion (when non-biological motion was performed by a human-like agent). David et al. (2007, 2009) showed that the EBA was involved in the detection of visuo-motor mismatches, and that TMS over the EBA interfered with the ability to detect such mismatches (note that this also supports the interpretation of the results of Study 3).

Taken together, the papers presented in this thesis suggest a model underlying body ownership featuring essential inferential mechanisms at hierarchically lower levels than the hitherto implied frontal or parietal areas, and thereby comply with the assumptions of the FEP that prediction error suppression is ongoing throughout the complete hierarchy of a generative model. In the following, I shall describe this proposal in a little more detail.

3.2. A generative model of body ownership underlying the rubber hand illusion

Some authors had speculated that the functional interplay of unimodal and multimodal areas during the RHI may reflect an updating of the probabilistic body model according to the principles of predictive coding (Hohwy, 2010; Apps and Tsakiris, 2014), but to date there is neither direct evidence for this, nor possible links between neuroanatomical models of the RHI and models of probabilistic information processing in the brain. A guiding idea behind this thesis was that this shortcoming—and just as well, the limitation of this network to relatively high-level multimodal areas described in the previous section—may be addressed by the formulation of the inferential network underlying the RHI as a predictive coding scheme. Figure 3 summarizes the information flow during the RHI in such a hierarchical scheme as implied by the results of the empirical papers presented in this thesis.

Figure 3. Schematic depiction of information flow during the RHI as implied by the results presented in this thesis. 1: LOC associates the touch seen on the dummy arm and the touch felt on one's real arm (presumably communicated from SII via the IPS) due to their coincidence on a body part in peripersonal space. As a consequence, a mismatch arises between the arm's location as inferred from the congruent visuo-tactile input, and the predictions of higher-level IPS. **2:** LOC sends a corresponding prediction error to the IPS. **3:** IPS adjusts its predictions about the location of one's arm and thus suppresses prediction error in LOC (via **5**), but this will generate a somatosensory prediction error in SII, where these predictions do not match the information about one's arm's (veridical) location. Hence in **4**, this prediction error is passed to the IPS, which has to continue suppressing now both kinds of errors (**3, 5**). The connectivity from IPS to PMv was significantly enhanced in our DCMs for congruent and incongruent visuo-tactile stimulation, but not significantly modulated by the RHI.

One crucial point in which this proposal differs from previous ones about the information flow underlying the RHI is the involvement of the visual LOC (see previous section). Second, the information exchange here follows a predictive coding scheme, in which the brain uses hierarchical inference on prediction errors to decide whether or not to self-attribute the seen dummy arm or not. Accordingly, the enhanced bottom-up connectivity from lower-level somatosensory and visual cortices to the IPS can be interpreted as propagating prediction errors arising from a mismatch of the ambiguous sensory data and the IPS' predictions about one's arm's location (see discussion of Study 2). One important interpretation of brain activity derived from predictive coding is that stimulus-evoked increases in neuronal activity reflect the generation of a prediction error (Friston, 2005;

Summerfield and Koechlin, 2008; Kok et al., 2012). Thus, the BOLD activations within the LOC and IPS (and over time, also in SII) we observed in Study 2 may reflect such prediction errors; It is also tempting to speculate that the fronto-parietal activations reported by other RHI studies also reflect the generation of prediction errors that potentially inform higher levels of the body model.

To conclude, the results of the presented papers, summarized in Figure 3, comply with previous conceptualizations of the mechanisms underlying the RHI (Makin et al., 2008; Tsakiris, 2010; Ehrsson, 2012), but they extend on them by offering a possible formal implementation of the inferential process in an inverted hierarchical generative model, following a predictive coding scheme (Friston and Stephan, 2007; Friston and Kiebel, 2009). Thereby the results suggest a generative model that extends beyond the traditionally implied areas in the PMv and IPS to include somatosensory and visual areas. The presented data hint toward an important "first step" of illusory body part self-attribution during the RHI in the "low-level" visual EBA: the interaction of body-related information and consequential generation of a low-level prediction error. In sum, these papers support speculations about the contribution of the EBA to the self-identification with a body (Jeannerod, 2004; Astafiev et al., 2004), and furthermore add strong support to speculations, including our own proposal (Limanowski and Blankenburg, 2013, Paper 4), about predictive coding as an important organizing principle of the information flow within the RHI network.

3.3. Towards a generative model of minimal selfhood: Challenges and future perspectives

The application of the FEP to body ownership as exemplified by the RHI was one aim of the present thesis. Second, I also wanted to explore the possibility that a generative model of body ownership (built upon insights gained from work on the RHI) may inform a larger-scale model of

minimal selfhood. This was an explorative endeavor, since it was clear that the RHI is a paradigm that isolates one condition of MPS (i.e., body ownership). The RHI is obviously only a temporary misinterpretation of sensory data by the brain (its effects vanish, for example, as soon as the participant moves her hand); while this is very helpful to learn about the mechanisms underlying body ownership, it cannot tell us everything about MPS (as I argued in Paper 5).

Our aim with this proposal was to highlight the hierarchical inferential structure potentially used by the brain to implement the key features of MPS as assumed by the FEP. Thomas Metzinger (2004, 2005, 2013; Gallese and Metzinger, 2003; Blanke and Metzinger, 2009) proposed that MPS emerges because the brain entails a model of the world, which includes the organism itself, and which for the sake of simplicity self-identifies with its own content. Metzinger's (2005) conclusions such as "no such things as selves exist in the world" were perhaps put forth a bit provocatively (and sparked an interesting debate with Shaun Gallagher). But in essence, his theory is very compatible with our knowledge about (self-)perception, and most importantly, as we (Limanowski and Blankenburg, 2013; corroborated by Metzinger, 2013) argued, with the FEP (Friston, 2005, 2010).

Even the schematic proposal we offer for a generative model of MPS (Figure 1 of Paper 4) already shows how much more complex such a model—if specified to include all key constituents of MPS—will have to be. Within such or a similar generative model of MPS, the mechanisms suggested by our DCMs (see Figure 4) will be but a tiny branch of the inferential tree. Nevertheless, situating the inferential hierarchy underlying the RHI within such a model may help derive testable hypotheses about *which level* of the generative self-model is actually targeted by the illusion. Thus could potentially clarify long open questions such as whether the RHI is indeed a full-blown experience of ownership of the seen dummy hand, or primarily just a visual capture of touch illusion (Holmes and Spence, 2004; Botvinick, 2004; Ehrsson, 2012). However, perhaps the most promising aspect of such a speculative formalization of philosophical MPS accounts within the FEP is that the resulting

generative minimal self-model may extend beyond models of the RHI, and thus it may do justice to an investigation of the self as an embodied agent (Merleau-Ponty, 1945; James, 1890; Gibson, 1977; Friston, 2011).

Certainly, future work is needed to validate the interpretation of these results, and to evaluate the proposed possibility of formalizing minimal selfhood in a generative model offered here. Some questions have remained unanswered, others have been newly opened up. One important question relates to the assumption that the brain uses the same network(s) for processing self- and non-self-relevant information. The processing the self and others in such "shared circuits" (Hurley, 2008; see also Jeannerod, 2004, 2007; Kilner et al., 2007; Gallese and Sinigaglia, 2011; Limanowski and Blankenburg, 2013; Apps and Tsakiris, 2014) is implied by overlapping neuronal activity for the processing of touch (Keysers et al., 2004, 2010), pain (Singer et al., 2004), and as suggested by the discovery of neurons with mirror properties, also of actions (Di Pellegrino et al., 1992; Gallese et al., 1996) of oneself and others. Interestingly, there have been recent arguments for a shared representation of the PPS around one's own and others' body parts in the PMv (Brozzoli et al., 2013). Following the predictive coding account, one can also argue that the brain uses hierarchical inference along the same pathways for the self and others. For example, Kilner and colleagues (2007) have presented a predictive coding account of hierarchical inference of action understanding by suppression of prediction error along the structures of the human mirror neuron system. But if the brain uses the same structures for self- and other-processing, how does it then distinguish between the self and others (Jeannerod, 2007)? This point has been put forth by Jeannerod (2004, Jeannerod and Pacherie, 2004) to argue that in fact, the mirror neuron system does not implement the self-other distinction, since it responds to the action itself (the "naked content" of the representation, Jeannerod and Pacherie, 2004), but not to the actor[4].

---

4   Interestingly, Jeannerod (ibd.) offers the EBA as a potential candidate region for self-other distinction (following up on the findings by Astafiev et al., 2004, who showed that the EBA responds preferentially to self-generated movements).

What could predictive coding accounts contribute to this debate? Following the FEP, the ongoing suppression of prediction error throughout the complete hierarchy of the generative model is, ultimately, encoding evidence for the existence of the agent itself in the form of model evidence (Friston, 2011). Therefore, there will be always less prediction error for the self than for others (concerning both actions and sensations). In other words, in generative models where the "mineness" of actions or sensations is encoded in the correct predictions of the multimodal input (Hohwy, 2007, 2013), and evidence for the self is potentially encoded as the current world-model's evidence (Friston, 2011; Metzinger, 2005; Limanowski and Blankenburg, 2013), the prediction error is what potentially tells me that something is "not me". Interestingly, in the RHI, the error that should be informative in this way is explained away and self-other distinction fails. We (Limanowski and Blankenburg) believe that one fruitful line of investigation, guided by the assumptions of the FEP, is the comparison of illusory to "veridical" body experience. As a possible first step in this direction, we have just acquired fMRI data to compare touch that is experienced on one's real arm versus an arm that is "owned" as a result of the RHI. These data will hopefully constitute a contribution to an extension of FEP explanations beyond the RHI paradigm to include the "real" body.

In conclusion, illusions like the RHI suggest that the experienced "reality" of oneself in the world rests—just as the character Morpheus says in the introductory quote from the 1999 science fiction film *The Matrix*—upon the current interpretation of sensory data by the brain. Thereby "current" and "interpretation" are two keywords that are equally emphasized by the FEP. The results presented in this thesis imply that the information flow in the hierarchical network underlying the self-attribution of body parts may follow a predictive coding scheme according to the FEP. This supports previous speculations about the RHI as arising from hierarchical probabilistic inference, and it also supports the conceptualization of body ownership and minimal selfhood as following similar principles: the suppression of prediction error across an inverted hierarchical generative model

entailed by the brain. Currently, the FEP with its emphasis on embodiment and active inference seems particularly well-suited to define the underlying functional architecture of such models.

References

Anderson, M. L., & Perlis, D. R. (2005). The roots of self-awareness. *Phenomenology and the Cognitive Sciences, 4,* 297-333.

Apps, M. A., & Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, *41*, 85-97.

Armel, K. C., & Ramachandran V. S. (2003). Projecting sensations to external objects: Evidence from skin conductance response. *Proceedings of the Royal Society of London B, 270*, 1499-1506.

Astafiev, S. V., Stanley, C. M., Shulman, G. L., & Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. *Nature Neuroscience 7*, 542–8.

Avillac, M., Hamed, S. B., & Duhamel, J. R. (2007). Multisensory integration in the ventral intraparietal area of the macaque monkey. *The Journal of Neuroscience*, *27*, 1922-1932.

Barnsley, N., McAuley, J. H., Mohan, R., Dey, A., Thomas, P., & Moseley, G. L. (2011). The rubber hand illusion increases histamine reactivity in the real arm. *Current Biology, 21*, R945-R946.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76*, 695-711.

Beauchamp, M. S., Yasar, N. E., Kishan, N., & Ro, T. (2007). Human MST but not MT responds to tactile stimulation. *The Journal of Neuroscience*, *27*, 8261-8267.

Beauchamp, M. S., Pasalar, S., & Ro, T. (2010). Neural substrates of reliability-weighted visual-tactile multisensory integration. *Frontiers in Systems Neuroscience*, *4*.

Bermúdez, J. L. (1998). *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press.

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

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences, 13*, 7-13.

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

Bracci, S., Ietswaart, M., Peelen, M. V., & Cavina-Pratesi, C. (2010). Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. *Journal of Neurophysiology*, *103*, 3389-3397.

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann K, et al. (2001a). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron, 29,* 287–96.

Bremmer, F., Schlack, A., Duhamel, J. R., Graf, W., & Fink, G. R. (2001b). Space coding in primate posterior parietal cortex. *NeuroImage*, *14*, S46-S51.

Brozzoli, C., Gentile, G., Bergouignan, L., & Ehrsson, H. H. (2013). A shared representation of the space near oneself and others in the human premotor cortex. *Current Biology, 23*, 1764-1768.

Brozzoli, C., Gentile, G., & Ehrsson, H. H. (2012). That's near my hand! Parietal and premotor coding of hand-centered space contributes to localization and self-attribution of the hand. *The Journal of Neuroscience*, *32*, 14573-14582.

Carlsson, K., Petrovic, P., Skare, S., Peterson, K. M., & Ingvar, M. (2000). Tickling expectations: Neural processing in anticipation of a sensory stimulus. *Journal of Cognitive Neuroscience, 12,* 691-703.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*, 181-204.

Costantini, M., & Haggard, P. (2007). The rubber hand illusion: Sensitivity and reference frame for body ownership. *Consciousness and Cognition, 16,* 229–240.

Costantini, M., Urgesi, C., Galati, G., Romani, G. L., & Aglioti, S. M. (2011). Haptic perception and body representation in lateral and medial occipito-temporal cortices. *Neuropsychologia, 49,* 821–9.

David, N., Cohen, M. X., Newen, A., Bewernick, B. H., Shah, N. J., Fink, G. R., & Vogeley, K. (2007). The extrastriate cortex distinguishes between the consequences of one's own and others' behavior. *NeuroImage*, *36,* 1004-1014.

David, N., Jansen, M., Cohen, M. X., Osswald, K., Molnar-Szakacs, I., Newen, A., ... & Paus, T. (2009). Disturbances of self–other distinction after stimulation of the extrastriate body area in the human brain. *Social neuroscience*, *4*, 40-48.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research, 91*, 176-180.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293,* 2470–2473.

Downing, P. E, & Peelen, M. V. (2011). The role of occipitotemporal body-selective regions in person perception. *Cognitive Neuroscience, 2,* 37–41.

Driver, J., & Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Current Biology*, *10*, R731-R735.

Ebisch, S. J. H., Perrucci, M. G., Ferretti, A., Del Gratta, C., Romani, G. L., & Gallese, V. (2008). The sense of touch: embodied simulation in a visuotactile mirroring mechanism for observed animate or inanimate touch. *Journal of Cognitive Neuroscience, 20,* 1611–1623.

Ehrsson, H. H. (2012). *The concept of body ownership and its relation to multisensory integration.* In: The New Handbook of Multisensory Processes, B.E. Stein (Ed.), MA: MIT Press (Cambridge).

Ehrsson, H. H., Spence, C., & Passingham, R. E. (2004). That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science, 305,* 875–877.

Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J., & Passingham, R. E. (2007). Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. *Proceedings of the National Academy of Sciences, 104*, 9828-9833.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162-169.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*, 815-836.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11,* 127-138.

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.

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

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*, 417-458.

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

Frith, C. (2009). *Making up the mind: How the brain creates our mental world*. Wiley. Com.

Gallagher, S. (1986). Lived body and environment. *Research in phenomenology, 16*, 139-170.

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Science, 4,* 14–21.

Gallagher, S., & Zahavi, D. (2010). *Phenomenological approaches to self-consciousness.* In: The Stanford Encyclopedia of Philosophy (Winter 2010 Edition), Edward N. Zalta (ed.), http://plato.stanford.edu/archives/win2010/entries/self-consciousness-phenomenological/.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593-609.

Gallese, V., & Metzinger, T. (2003). Motor ontology: the representational reality of goals, actions and selves. *Philosophical Psychology, 16*, 365-388.

Gentile, G., Guterstam, A., Brozzoli, C., & Ehrsson, H. H. (2013). Disintegration of multisensory signals from the real hand reduces default limb self-attribution: an fMRI study. *The Journal of Neuroscience, 33*, 13350-13366.

Gibson, J. J. (1977). *The theory of affordances.* Hilldale, USA.

N. D. Goodman and J. B. Tenenbaum (electronic). Probabilistic Models of Cognition. Retrieved 01.08.2014 from http://probmods.org.

Graziano, M. S., Yap, G. S., & Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science, 266*, 1054-1057.

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

Graziano, M. S. (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*, 10418-10421.

Graziano, M. S., Cooke, D. F., & Taylor, C. S. (2000). Coding the location of the arm by sight. *Science*, *290*, 1782-1786.

Graziano, M. S., & Botvinick, M. M. (2002). How the brain represents the body: insights from neurophysiology and psychology. *Common mechanisms in perception and action: Attention and performance XIX*, 136-157.

Graziano, M. S., & Cooke, D. F. (2006). Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia*, *44*, 845-859.

Grefkes, C., Weiss, P. H., Zilles, K., & Fink, G. R. (2002). Crossmodal processing of object features in human anterior intraparietal cortex: an fMRI study implies equivalencies between humans and monkeys. *Neuron, 35*, 173-184.

Grefkes, C., & Fink, G. R. (2005). REVIEW: The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, *207*, 3-17.

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 290*, 181-197.

Haggard, P., Christakou, A., & Serino, A. (2007). Viewing the body modulates tactile receptive fields. *Experimental Brain Research*, *180*, 187-193.

Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche, 13*, 1-20.

Hohwy, J. (2010). *The hypothesis testing brain: Some philosophical applications.* In ASCS09: Proceedings of the 9th Conference of the Australasian Society for Cognitive Science (pp. 135-144).

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in psychology, 3,* 96.

Hohwy, J. (2013). *The predictive mind*. Oxford: Oxford University Press.

Hohwy, J., & Paton, B. (2010). Explaining away the body: Experiences of supernaturally caused touch and touch on non-hand objects within the rubber hand illusion. *PLoS ONE, 5*, 1-10.

Holmes, N. P., & Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cognitive Processing, 5*, 94-105.

Ionta, S., Heydrich, L., Lenggenhager, B., Mouthon, M., Fornari, E., Chapuis, D., Gassert, R., et al (2011). Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective. *Neuron, 70,* 363–374.

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *Neuroimage*, *31*, 429-439.

James, W. (1890). *The principles of psychology.* New York: Dover.

Jeannerod, M. (2004). Visual and action cues contribute to the self–other distinction. *Nature Neuroscience*, *7*, 422-423.

Jeannerod, M. (2007). Being oneself. *Journal of Physiology-Paris, 101*, 161-168.

Kammers, M. P., Verhagen, L., Dijkerman, H. C., Hogendoorn, H., De Vignemont, F., & Schutter, D. J. (2009). Is this hand for real? Attenuation of the rubber hand illusion by transcranial magnetic stimulation over the inferior parietal lobule. *Journal of Cognitive Neuroscience, 21*, 1311-1320.

Kennett, S., Taylor-Clarke, M., & Haggard, P. (2001). Noninformative vision improves the spatial resolution of touch in humans. *Current Biology*, *11*, 1188-1191.

Keysers, C., Wicker, B., Gazzola, V., Anton, J. L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron, 42*, 335-346.

Keysers, C., Kaas, J. H., & Gazzola, V. (2010). Somatosensation in social perception. *Nature Reviews Neuroscience, 11*, 417-428.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology, 4*, e1000209-?

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive Processing, 8*, 159-166.

Kok, P., Jehee, J. F., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron*, *75*, 265-270.

Kuehn, E., Trampel, R., Mueller, K., Turner, R., & Schütz-Bosbach, S. (2012). Judging roughness by sight-A 7-tesla fMRI study on responsivity of the primary somatosensory cortex during observed touch of self and others. *Human Brain Mapping, 34,* 1882-1895.

Làdavas, E. (2002). Functional and dynamic properties of visual peripersonal space. *Trends in Cognitive Sciences*, *6*, 17-22.

Legrand, D. (2006). The bodily self: The sensori-motor roots of pre-reflective self-consciousness. *Phenomenology and the Cognitive Sciences, 5*, 89-118.

Lloyd, D. M., Shore, D. I., Spence, C., & Calvert, G. A. (2002). Multisensory representation of limb position in human premotor cortex. *Nature Neuroscience*, *6,* 17-18.

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.

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

Ma, W. J., & Pouget, A. (2008). Linking neurons to behavior in multisensory perception: A computational review. *Brain research*, *1242*, 4-12.

Macaluso, E., & Maravita, A. (2010). The representation of space near the body through touch and vision. *Neuropsychologia, 48,* 782-795.

Makin, T. R., Holmes, N. P., Zohary, E. (2007). Is that near my hand? Multisensory representation of peripersonal space in human intraparietal sulcus. *The Journal of Neuroscience, 27,* 731–40.

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

Maravita, A., Spence, C., & Driver, J. (2003). Multisensory integration and the body schema: close to hand and within reach. *Current Biology*, *13*, R531-R539.

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*, 640-662.

Merleau-Ponty, M. (1945 / 2013). *Phenomenology of perception.* Hoboken: Taylor and Francis.

Metzinger, T. (2004). The Subjectivity of Subjective Experience: A Representationalist Analysis of the First-Person Perspective. *Networks 3-4,* 33-64.

Metzinger, T. (2005). *Being no one: The self-model theory of subjectivity.* The MIT Press.

Metzinger, T. (2013). Why are dreams interesting for philosophers? The example of minimal phenomenal selfhood, plus an agenda for future research. *Frontiers in Psychology, 4.*

Moseley, G. L., Olthof, N., Venema, A., Don, S., Wijers, M., Gallace, A., & Spence, C. (2008). Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. *Proceedings of the National Academy of Sciences, 105,* 13169–13173.

Moseley, G. L., Gallace, A., & Spence, C. (2012). Bodily illusions in health and disease: physiological and clinical perspectives and the concept of a cortical 'body matrix'. *Neuroscience & Biobehavioral Reviews, 36*, 34-46.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological Cybernetics, 66*, 241-251.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, *99*, 15164-15169.

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

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience, 8,* 636–648.

Petkova, V. I., Björnsdotter, M., Gentile, G., Jonsson, T., Li, T. Q., & Ehrsson, H. H. (2011). From part-to whole-body ownership in the multisensory brain. *Current Biology, 21,* 1118–1122.

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

Revonsuo, A., & Newman, J. (1999). Binding and consciousness. *Consciousness and Cognition, 8,* 123-127.

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.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1997). The space around us. *Science*, *277*, 190-191.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience, 4,* 413-422.

Sereno, M. I., & Huang, R. S. (2006). A human parietal face area contains aligned headcentered visual and tactile maps. *Nature Neuroscience, 9,* 1337–43.

Sereno, M. I., & Huang, R. S. (2014). Multisensory maps in parietal cortex. *Current Opinion in Neurobiology, 24,* 39-46.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*, 1157-1162.

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences, 216,* 427-459.

Striem-Amit, E., & Amedi, A. (2014). Visual Cortex Extrastriate Body-Selective Area Activation in Congenitally Blind People "Seeing" by Using Sounds. *Current Biology*, *24*, 687-692.

Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron*, *59*, 336-347.

Taylor-Clarke, M., Kennett, S., & Haggard, P. (2002). Vision modulates somatosensory cortical processing. *Current Biology*, *12*, 233-236.

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

Tsakiris, M., & Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *J. Exp. Psychol. Hum. Percept. Perform., 31,* 80–91.

Tsakiris, M., Hesse, M. D., Boy, C., Haggard, P., & Fink, G. R. (2007). Neural signatures of body ownership: a sensory network for bodily self-consciousness. *Cerebral Cortex, 17,* 2235–2244.

Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. M. (2007). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature Neuroscience, 10,* 30–31.

van Elk, M. (2014). The left inferior parietal lobe represents stored hand-postures for object use and action prediction. *Frontiers in Psychology*, *5*.

Verschure, P. F., Voegtlin, T., & Douglas, R. J. (2003). Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature, 425*, 620-624.

Whiteley, L., Kennett, S., Taylor-Clarke, M., & Haggard, P. (2004). Facilitated processing of visual stimuli associated with the body. *Perception*, *33*, 307-314.

Original publications

Study 1

Limanowski, J., Lutti, A., Blankenburg, F. (2014). The extrastriate body area is involved in illusory limb

ownership. *NeuroImage, 86*, 514-524. doi: 10.1016/j.neuroimage.2013.10.035
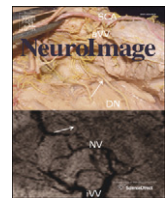
The extrastriate body area is involved in illusory limb ownership

Jakub Limanowski [a,*], Antoine Lutti [b], Felix Blankenburg [a,c,d]

[a] Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, 10117 Berlin, Germany
[b] LREN, Departement des neurosciences cliniques, CHUV, University Lausanne, Switzerland
[c] Dahlem Institute for Neuroimaging of Emotion, Freie Universität Berlin, 14195 Berlin, Germany
[d] Center for Adaptive Rationality (ARC), Max Planck Institute for Human Development, 14195 Berlin, Germany

ARTICLE INFO

ABSTRACT

The Rubber Hand Illusion (RHI) is an established paradigm for studying body ownership, and several studies have implicated premotor and temporo-parietal brain regions in its neuronal foundation. Here we used an automated setup to induce a novel multi-site version of the RHI in healthy human participants inside an MR-scanner, with a RHI and control condition that were matched in terms of synchrony of visual and tactile stimulation. Importantly, as previous research has shown that most of the ownership-related brain areas also respond to observed human actions and touch, or body parts of others, here such potential effects of the experimenter were eliminated by the automated procedure. The RHI condition induced a strong ownership illusion; we found correspondingly stronger brain activity during the RHI versus control condition in contralateral middle occipital gyrus (mOCG) and bilateral anterior insula, which have previously been related to illusory body ownership. Using independent functional localizers, we confirmed that the activity in mOCG was located within the body-part selective extrastriate body area (EBA). Crucially, activity differences in participants' peak voxels within the left EBA correlated strongly positively with their behavioral illusion scores. Thus EBA activity also reflected interindividual differences in the experienced intensity of illusory limb ownership. Moreover, psychophysiological interaction analyses (PPI) revealed that contralateral primary somatosensory cortex had stronger brain connectivity with EBA during the RHI versus control condition, while EBA was more strongly interacting with temporo-parietal multisensory regions. In sum, our findings demonstrate a direct involvement of EBA in limb ownership.

© 2013 Elsevier Inc. All rights reserved.

Introduction

To be oneself among others, one needs to identify with a particular body (Blanke and Metzinger, 2009; Gallagher, 2000; Jeannerod, 2007). Most accounts of body ownership have emphasized multimodal information integration in hierarchical cortical networks as a fundamental mechanism underlying a coherent self-representation (Apps and Tsakiris, 2013; Blanke, 2012; Hohwy, 2007, 2010; Petkova et al., 2011; Seth et al., 2011; Tsakiris, 2010). These theories are supported by recent neuroimaging experiments that have provided novel insights into how the brain self-attributes body parts based on such integration of visual, tactile, and proprioceptive information. In the Rubber Hand Illusion (RHI; Botvinick and Cohen, 1998), synchronous stroking of a dummy body part together with one's own corresponding body part typically misleads the brain to self-attribute the dummy limb (Botvinick and Cohen, 1998; Ehrsson et al., 2004; Tsakiris and Haggard, 2005) or even a whole body (Ehrsson, 2007; Lenggenhager et al., 2007). The experience of (illusory) body ownership has been linked to activity in frontal brain regions, predominantly the ventral premotor cortex (PMv;

Ehrsson et al., 2004, 2005; Petkova et al., 2011), but also posterior regions like the right temporo-parietal junction (rTPJ; Blanke et al., 2002, 2005; Ionta et al., 2011; Tsakiris et al., 2008), posterior parietal cortex and intraparietal sulcus (PPC/IPS; Brozzoli et al., 2012; Ehrsson et al., 2004; Gentile et al., 2011; Petkova et al., 2011; Shimada et al., 2005; Tsakiris, 2010), and occipito-temporal regions like the body part-selective extrastriate body area (EBA; Arzy et al., 2006; Blanke and Mohr, 2005; Downing et al., 2001; Ionta et al., 2011). Primary somatosensory cortex (SI; Kanayama et al., 2007, 2009; Lenggenhager et al., 2011; Tsakiris et al., 2007) and the anterior insula (AI; Ehrsson et al., 2007) have also been associated with body ownership. Activity in these regions has been interpreted as reflecting the degree of illusory self-attribution or "incorporation" of the fake limb or body (Blanke, 2012; Ehrsson et al., 2004; Holmes and Spence, 2004; Petkova et al., 2011; Tsakiris, 2010).

Here, we used a fully automated setup to induce a novel, multi-site version of the RHI inside an fMRI scanner with high spatial resolution, addressing two potential caveats of the procedures typically used to evoke the illusion. First, we matched visual and tactile stimuli of both RHI and control condition in temporal synchrony, in contrast to the typically used asynchronous stroking control condition where observed touch on the dummy hand and felt touch on the own hand are presented serially. In our control condition, observed and felt touch were presented

---

* Corresponding author at: Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Luisenstraße 56, Haus 1, 10117 Berlin, Germany.
E-mail address: jakub.limanowski@hu-berlin.de (J. Limanowski).

synchronously at spatially incongruent locations (palm and forearm). This synchronous stimulation countered potential problems associated with a serial, isolated presentation of observed and felt touch: For example, premotor cortex has been shown to be engaged in (serial) sensory predictions even in tasks using abstract, nonbiological stimuli (Schubotz and von Cramon, 2002, 2003, 2004), and the presentation of observed touch before felt touch at the same location could potentially be influenced by effects of anticipation of touch (see e.g. Carlsson et al., 2000; Keysers et al., 2010; Kuehn et al., 2012). Moreover, the resulting design enabled us to calculate a joint contrast comparing two RHI and control conditions, in which spatiotemporal differences between stimuli in the conditions were averaged out, and thus the resulting effects were attributable to the experienced illusion only. Second, by fully automating our experimental setup, we eliminated the human experimenter from the procedure. The induction of the RHI by touch from another person may interfere with self-related information processing, as many brain regions associated with body ownership (e.g., EBA, insula, PMv, and SI) also respond to observed human actions and touch, or mere vision of bodies of others (Bernhardt and Singer, 2012; Blanke, 2012; Ebisch et al., 2008; Keysers et al., 2010; Peelen and Downing, 2007; Zaki and Ochsner, 2012). Therefore, we aimed to isolate body ownership mechanisms from effects introduced by social interaction. The RHI has been induced automatically in one PET study (Tsakiris et al., 2007), but to our knowledge no automated MR-compatible RHI setup has been reported to date. We tested for BOLD signal differences between the RHI versus control condition within the ownership-related regions identified in previously published studies, expecting effects in regions whose response to the illusion is *not* influenced by receiving human touch. Moreover, we tested whether activity in those regions would reflect individual differences in the experienced intensity of the ownership illusion (Ehrsson et al., 2004; Petkova et al., 2011; Tsakiris et al., 2007).

Materials and methods

Participants

20 healthy volunteers (22–36 years old; 13 females; 19 right-handed, one classified as "mixed left-handed", measured with the Edinburgh Handedness Inventory, Oldfield, 1971; normal or corrected-to-normal vision) participated in the experiment; 16 of these participants took part in an additional scanning session for the functional EBA localizer. All participants gave written informed consent before the experiment and the study was approved by the local Ethical Committee of the Charité University Hospital (Berlin) and corresponded to the Human Subjects Guidelines of the Declaration of Helsinki.

Apparatus and procedure

A realistic life-size right dummy arm was mounted on a custom console made of transparent acrylic glass, which was set up atop the participant's chest (Fig. 1A). The participant's right arm was placed horizontally behind the dummy arm in a corresponding posture (distance between arms ~13 cm). To ensure that the location of visual stimuli in eye-centered coordinates remained the same, the participant was instructed to fixate a small dot in the middle of the dummy arm throughout the whole experiment, while her or his own arm was completely occluded from view (Fig. 1B). In contrast to previous studies (Ehrsson et al., 2004), our participants were not subjected to any prior information about the RHI and we collected the illusion intensity ratings after, not during the functional scanning sessions. For full, direct vision of the dummy arm, the participant's head was slightly tilted within the head coil (approx. 20–30°), her or his head and shoulders were foam-padded, the right arm was attached to the console with Velcro strips to eliminate motion during the experiment, and the gap between dummy arm and the participant's shoulder was covered with a black piece of cloth. Two pairs of sponge brushes were installed at anatomically corresponding locations at the palm and forearm of the own and dummy arms (Fig. 1B). Each of the brushes was separately moveable in back-and-forth 180° rotations, thereby applying touch at a specific location. To eliminate the influence of being touched by a human (seeing touch delivered with a hand may have specific effects on somatosensation; Ebisch et al., 2008; Keysers et al., 2010), and to ensure continuous temporal synchrony of strokes and corresponding stroking patterns, the brushes were driven by four separate electrical stepping motors placed outside the scanner room. The stepping motors (1.8° stepping angle; RS Components GmbH, Mörfelden-Walldorf, Germany) were controlled by a custom MATLAB (The MathWorks, Inc., Natick, USA) script via a computer parallel port, which also received the scanner-triggers to synchronize stimulation onsets with the fMRI acquisition. The motors' movements were mechanically transmitted to the brushes via a custom construction of nonmagnetic Plexiglas cables and plastic gears. During stimulation, the respective brushes performed strokes at 1.3 Hz, with random inter-stroke intervals (0, 50, or 150 ms), as an irregular stroking pattern has been shown to increase the RHI (Armel and Ramachandran, 2003). Before the start of the experiment, the two brushes touching the participant's own arm were adjusted and tested each, to assure reliable touch sensation. The participant then completed a brief practice run to get acquainted with the setup and the different stimulation types, and proceeded with the five experimental runs (see below). Subsequently, the strength of experienced ownership of the dummy arm in each condition was quantified (the
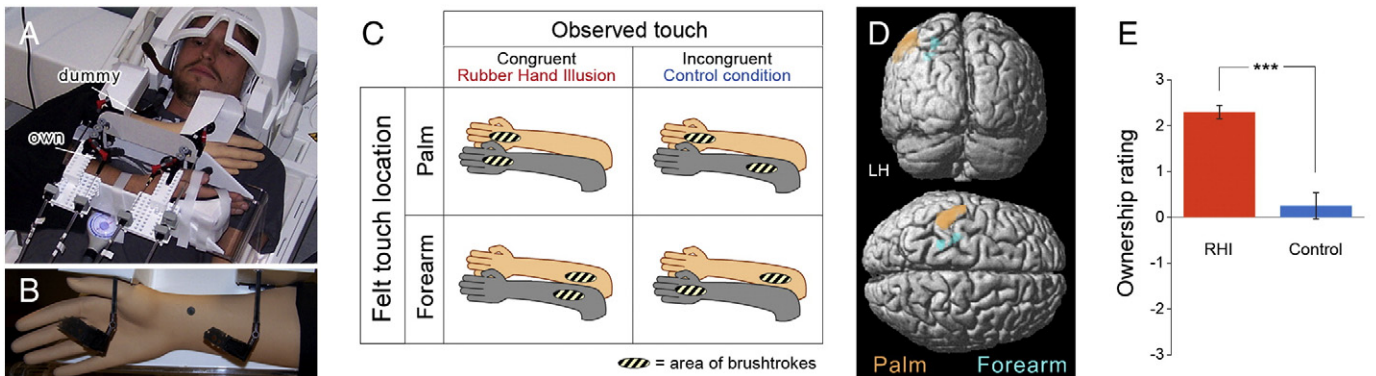
Fig. 1. (A) Experimental apparatus with the own arm occluded from view behind the dummy arm. (B) Participants' view of the dummy arm. (C) Locations of synchronous stroking on the dummy (gray) and own arm for the RHI and control condition. (D) Tactile stimulation produced significant ($p < 0.05$ FWE, small volume corrected with the left SI) activations in contralateral SI. The surface render shows the significant main effects ($p < 0.001$ uncorrected to visualize somatotopic arrangement) of stroking at the palm ($x = -48$, $y = -38$, $z = 60$, $t = 5.44$) and forearm ($x = -24$, $y = -38$, $z = 56$, $t = 3.78$) location during the visuo-tactile localizer runs, masked with anatomical left SI. (E) Participants' mean ratings of experienced ownership of the dummy arm during the RHI and control condition; error bars are standard errors of the mean, significance level obtained from Wilcoxon's signed-rank test ($z = 3.99$, $n = 20$, $p = 0.00007$).

respective two stimulation types of each condition were presented sequentially) by asking the participant to indicate her agreement with the following statement on a 7-point Likert-scale ranging from -3 ("completely disagree") to $+3$ ("completely agree"): "During the stimulation, it felt as if the dummy arm was my own arm." (Botvinick and Cohen, 1998). For the RHI condition the individual onset of the ownership illusion was assessed as well (Ehrsson et al., 2004): The participant was instructed to give a brief verbal response as soon as she would feel (and only if she would feel) that the dummy arm felt as if it was her own arm. The elapsed time between the beginning of stimulation and the participant's first verbal statement of experienced ownership of the dummy arm was measured with a stopwatch to represent the individual onset of the ownership illusion. After the scanning session, the participant completed a German version of the Interpersonal Reactivity Index (IRI, Davis, 1983), which has been used to measure trait empathy in other fMRI studies (e.g., Schaefer et al., 2012).

Experimental design

The scanning comprised two sessions: one for the RHI experiment and one for the functional localization of EBA and hMT + (see below). In the first scanning session, the RHI experiment was conducted as a repeated-measures block design comprising four conditions: the RHI condition, a control condition, and a visual and tactile stroking-only condition (localizers). Each participant completed five runs, with each condition presented four times per run (presentation order of conditions was randomized for each run). Note that, due to the multi-site setup, two spatially different types of stroking could occur in each condition: The RHI condition was operationalized as synchronous stroking of anatomically corresponding locations of own and dummy arms (Botvinick and Cohen, 1998), i.e., the own and dummy arms were simultaneously either both touched at the palm, or the forearm location. In the control condition, synchronous strokes were applied to anatomically incongruent locations of own and dummy arms (i.e., simultaneous touch at the own palm and dummy forearm, or vice versa), in contrast to the typically used asynchronous stroking control condition (Armel and Ramachandran, 2003; Botvinick and Cohen, 1998; Ehrsson et al., 2004). This novel synchronous control condition was enabled by the multi-site setup; pilot experiments confirmed that, despite temporal synchrony of observed and felt touch, this condition did not induce the RHI. As functional localizers for the different visuo-tactile stimulations, we implemented a visual-only (dummy arm touch) and a tactile-only (own arm touch) condition into the design. Each stimulation was presented in a block of 17.9 s duration (18 brushstrokes), followed by a rest period of 11.2 s.

Functional localization of extrastriate body area and hMT +

In an additional scanning session, we employed a standard functional EBA localizer. This was done to functionally verify that the activation in the left middle occipital gyrus during the RHI versus control condition indeed corresponded to the location of the extrastriate body area (Downing et al., 2001). The EBA has been shown to respond more strongly to pictures of body parts versus objects (Downing and Peelen, 2011; Downing et al., 2001; Urgesi et al., 2007). Therefore, to localize EBA, participants were shown color photographs of human body parts (hands and feet), and object parts (motorcycles, following Urgesi et al., 2007) on a white background (presented on a screen viewed via a mirror at $18.7° \times 13.7°$ visual angle). Stimulus categories were presented block-wise in random order, the order of stimuli within each block was also randomized. Fig. 3A shows sample stimuli. Each picture was presented for 700 ms followed by a 150 ms blank screen within blocks of 20 s; a black fixation cross was shown between the blocks for 20 s and served as a baseline. We calculated the contrast BODY–OBJECT to identify the effects of vision of body parts versus objects. In addition, we also localized the adjacent motion-sensitive area hMT +,

because body part- (EBA) and motion-sensitive (hMT+) responses in extrastriate cortex may overlap (Spiridon et al., 2006). We used a standard motion localizer (Tootell et al., 1995): participants fixated the center of an annulus of 12° diameter, consisting of 300 randomly arranged stationary, or radially moving white dots against a black background. During motion, dots were periodically moving towards, or away from the center of the annulus (alternating every 2.25 s). Moving and stationary dots were presented with the Psychtoolbox (Brainard, 1997) alternating in blocks of 18 s length; each condition was presented 8 times. To reveal the effect of moving versus stationary stimuli, we calculated the contrast MOTION–STATIC. Sixteen participants of our RHI experiment (four were not able to participate) were scanned in this additional session. The fMRI parameters, data preprocessing, and analyses used for the functional data obtained in this scanning session were identical as described in the following for the RHI experiment.

fMRI data acquisition

The experiment was conducted on a whole-body 3 T scanner (Tim Trio, Siemens, Germany), equipped with a 32-channel head coil. T2*-weighted functional images were acquired using a customized 3D-EPI sequence (Lutti et al., 2012). Parallel imaging (GRAPPA image reconstruction) was used along the phase and partition directions (acceleration factor 2), yielding an acquisition time of 2240 ms per image volume (image resolution: $2.0 \times 2.0 \times 2.0$ mm³, TR = 70 ms, matrix size [96, 96, 64], TE = 33 ms, flip angle = 20°, BW = 1408 Hz). A total of 1055 functional volumes were recorded for each participant (five runs with 211 volumes each). After the functional runs and ownership ratings (see below), a high-resolution T1-weighted structural image was acquired for each participant (3D MPRAGE, voxel size = 1 mm × 1 mm × 1 mm, FOV = 256 mm × 256 mm, 176 slices, TR = 1900 ms, TE = 2.52 ms, flip angle = 9°).

Data preprocessing and analysis

Data were preprocessed and analyzed using SPM8 (Wellcome Department of Cognitive Neurology, London, UK: www.fil.ion.ucl.ac.uk/spm/). One dummy volume (38.255 s) was routinely recorded at the beginning of each run, and excluded from the analysis. Furthermore, five volumes (acquired during two rest periods) from one participant had to be discarded due to extensive movement artifacts. Individual slices of all volumes were scanned for physically-based artifacts, and, if necessary, repaired by interpolation using the SPM ArtRepair toolbox (Mazaika et al., 2009) with default settings (art_slice program; 0.08% of slices corrected). Images were then realigned to the first image of each run to correct for head motion, using a least squares approach and a 6 parameter rigid body transformation. Each participant's structural image was co-registered with the realigned functional images, and segmented into white matter, gray matter, and cerebrospinal fluid (CSF). A mask image was created from the structural CSF-segment using the SPM Volumes toolbox (thresholded to ensure 90% tissue probability), and applied to the timeseries of each run. To minimize the effect of physiological noise, the averaged timeseries of all voxels within the CSF-mask was later included into the first level design matrices as a nuisance regressor (Weissenbacher et al., 2009). Functional images were spatially normalized to the MNI standard brain (SPM8 EPI template), and spatially smoothed by an isotropic Gaussian kernel of 5 mm FWHM. Data were detrended using a linear mean global signal removal script (Macey et al., 2004). Outlier volumes showing excessive movement were identified and repaired with the SPM ArtRepair toolbox by interpolation (art_global program; default movement threshold = 0.5 mm/TR after motion correction; 2.99% of volumes repaired).

Statistical parametric maps were calculated using a standard two-level mixed-effects model. In the first-level analysis, a general linear regression model was fit to each participant's dataset. Microtime onset was set to the middle slice of each volume, and low-frequency signal

drifts in the images were removed by a high-pass filter (cut-off frequency 300 s). The six movement parameters and the extracted CSF-timeseries (see Materials and methods) were added as nuisance regressors to each run. Each stimulation type was modeled as regressors with a boxcar function and convoluted with the standard hemodynamic response function of SPM. Because of the two spatially distinct stroking types, this resulted in two regressors per condition: for the RHI ($RHI_{palm/palm}$ & $RHI_{arm/arm}$), control ($CONTROL_{palm/arm}$ & $CONTROL_{arm/palm}$), visual only ($VISUAL_{palm}$ & $VISUAL_{arm}$), and tactile only ($TACTILE_{palm}$ & $TACTILE_{arm}$) conditions. For each regressor, T-contrasts versus baseline were calculated in the GLM. The resulting contrast images of all participants were entered into flexible factorial within-subject GLMs at the second-level (random effect, group analysis), including a between-subjects factor modeling the subject constants. We tested for BOLD signal differences between the RHI and control condition (see Fig. 1C) with the factors Felt touch location (palm, forearm), and Observed touch (congruent, incongruent). For the RHI versus control comparison, we were interested in the effects of synchronous congruent (RHI condition) versus synchronous incongruent touch (control condition). We therefore combined the two stimulation types for each location in the contrast $RHI–CONTROL = (RHI_{palm/palm} + RHI_{arm/arm}) - (CONTROL_{palm/arm} + CONTROL_{arm/palm})$. Moreover, following the procedure described by Ehrsson et al. (2004), we tested for activity that was specifically related to the period before, or after illusion onset. To this end, we used each participant's individually assessed illusion onset (see below) to divide each of the first-level RHI and CONTROL regressors into a regressor modeling the pre-illusion-onset phase, and one modeling the post-illusion-onset phase. For each regressor, we then calculated a contrast comparing the period before illusion onset with the period after illusion onset (e.g., $RHI_{palm/palm[pre]} - RHI_{palm/palm[post]}$). On the second level, the resulting contrast images were entered into the same flexible factorial design as used for the RHI versus control comparison. To compare brain activity during the post-illusion versus pre-illusion phase, we calculated the inverse contrast on the second level. The effects of the functional localizers were investigated in a design with the factors Modality (tactile, visual), and Stroking location (palm, forearm).

We analyzed changes in brain connectivity of SI during the RHI versus control condition by means of psychophysiological interactions (Friston et al., 1997). To account for the fact that stroking occurred at two distinct locations, we calculated two separate PPIs, one for touch at the palm ($RHI_{palm/palm} - CONTROL_{arm/palm}$), and one for touch at the forearm ($RHI_{arm/arm} - CONTROL_{palm/arm}$). Spheres of 2 mm radius were constructed around each participant's individually thresholded peak voxel within contralateral SI (mean coordinates of seed regions for touch at the palm: $x = -44.1 \pm 6.5$, $y = -31.9 \pm 5.2$, $z = 52.4 \pm 8.6$; and touch at the forearm: $x = -32.1 \pm 5.7$, $y = -37.0 \pm 4.3$, $z = 60.1 \pm 10.3$; MEAN $\pm$ SD), and the first eigenvariate of the BOLD signal was extracted. The psychophysiological interaction terms for the RHI versus control condition were created, and included into GLMs. For each participant, the contrast images of the two PPIs were averaged using the imCalc function of SPM8 to obtain a single contrast image, which was entered into a one-sample t-test at the group-level. We also calculated a PPI with seed regions in the left EBA (mean coordinates of seed regions: $x = -49.8 \pm 4.4$, $y = -69.2 \pm 4.8$, $z = 4.2 \pm 3.2$; MEAN $\pm$ SD) to examine connectivity of EBA during the RHI versus control. The procedure differed from the described PPI only in that we calculated a single PPI on the joint first-level contrast ($RHI_{palm/palm} + RHI_{arm/arm}) - (CONTROL_{palm/arm} + CONTROL_{arm/palm}$). We also calculated a regression analysis on the first-level contrast images comparing the RHI versus the control condition ($RHI_{palm/palm} + RHI_{arm/arm}) -(CONTROL_{palm/arm} + CONTROL_{arm/palm}$) using the illusion scores (see below) as a covariate.

Effect sizes within clusters obtained from the second-level contrasts were calculated as percent signal change as follows: each cluster of interest, thresholded at $p < 0.001$, uncorrected, was saved as a binary image. We used the SPM rfxplot toolbox (Gläscher, 2009) to extract the parameter estimates for each participant's peak voxel within the cluster mask image. These values were averaged over participants to calculate group-level effect sizes; for correlation analyses, we used the individual parameter estimates of each subject. The anatomical mask for the left SI (BAs 3a, 3b, 1, and 2) was created with the Anatomy toolbox (Eickhoff et al., 2005). All reported coordinates correspond to the MNI space as used by SPM8. Neuroanatomical labels were derived from the SPM Anatomy toolbox where possible. For visualization of the results, the statistical maps were projected onto the SPM8 canonical single-subject T1 template, or rendered on a surface template. Based on our strong a priori assumptions (following published findings, see Introduction) for the anatomical location of regions involved in the RHI, we corrected for multiple comparisons using small volume correction based on pre-defined regions of interest (8 mm ROIs), applying a statistical threshold of $p < 0.05$ familywise error corrected (FWE). We used the following ROIs (RH = right hemisphere, LH = left hemisphere; odd MNI coordinates were approximated due to our voxel size of 2 versus typically 3 mm, e.g., for $x = 51$ we used $x = 52$): PMv (from Ehrsson et al., 2004; RH: $x = 48$, $y = 18$, $z = 40$; LH: $x = -58$, $y = 16$, $z = 10$), AI (from Ehrsson et al., 2007; RH: $x = 40$, $y = 28$, $z = 6$; LH: $x = -42$, $y = 20$, $z = 10$), PPC/IPS (from Ehrsson et al., 2004; RH: $x = 34$, $y = -46$, $z = 52$; LH: $x = -36$, $y = -42$, $z = 52$). For the EBA, we initially used the coordinates provided by Downing et al. (2001; RH: $x = 50$, $y = -70$, $z = 0$; LH: $x = -50$, $y = -72$, $z = 8$). However, as we also functionally defined the EBA in a separate localizer in our own sample, we were able to create sample-specific ROIs for the left and right EBA by binary saving the respective clusters in the left and right lateral occipital cortex that were activated ($p < 0.05$ FWE, whole-brain) by the EBA localizer. For the right TPJ, we constructed a ROI based on a transformation of the Talairach coordinates reported by Blanke et al. (2005; RH: $x = 66$, $y = -38$, $z = 18$). For all other brain regions, we report only those activations that survived a threshold of $p < 0.05$, whole-brain FWE corrected.

Behavioral data

Participants' verbal ownership ratings did not pass the Kolmogorov–Smirnov test for normality, and were therefore compared using the nonparametric Wilcoxon's signed-rank test. For the correlation analyses with brain activity differences, we calculated individual illusion scores as a compound measure reflecting both strength and prevalence of the experienced ownership illusion during the RHI condition, based on the procedure described by Ehrsson et al. (2004): each participant's difference between ownership ratings for the RHI condition and the control condition was multiplied by the duration of experienced ownership during the timeframe of stimulation in the scanning session (i.e., by subtracting the time of reported onset of the RHI from total stimulation duration). To account for the small sample size and possible effects of behavioral outliers, we used the nonparametric Spearman's rho test for the correlation analyses. All significance levels were assessed using two-tailed tests; we report only those results that survived Bonferroni correction for the number of tests performed.

Results

Behavioral results

Participants' mean reported ownership ratings for the RHI condition were significantly higher than those for the control condition (Fig. 1E; Wilcoxon's signed-rank test, $n = 20$, $Z = 3.99$, $p = 0.00007$). Moreover, the RHI condition was the only condition in which all participants affirmed experiencing ownership of the dummy arm (i.e., all ratings were positive; mean ownership rating = 2.30, SD = 0.66). On average, participants reported experiencing the illusion after 5.66 s (SD = 5.87 s),

which means that the duration of stimulation in each block (17.9 s) was long enough to evoke the RHI. Scores on the Interpersonal Reactivity Index, including the individual subscales, did neither correlate with illusion scores of the RHI, nor with brain activity differences in any of the specified regions (Pearson's correlation coefficients, all $ps > 0.05$).

The Rubber Hand Illusion versus control condition produces brain activity in extrastriate body area and anterior insula

We were interested in the specific effects of limb ownership on brain activity as induced via our RHI setup, which differed from those used in previous fMRI studies in that (i) it was fully automated and (ii) the control condition was synchronous (i.e., temporal synchrony of visual and tactile information was given). Therefore, we first tested for BOLD signal differences between the RHI and control condition, computing the contrast $(RHI_{palm/palm} + RHI_{arm/arm}) - (CONTROL_{palm/arm} + CONTROL_{arm/palm})$. Importantly, this contrast was fully matched in terms of physical stimulus properties, i.e., location and timing for the RHI and control condition. Effects of the RHI as obtained from this contrast should reveal which of the regions of interest (see Materials and methods) were responsive to the ownership illusion, even when it was induced without social interaction. Results of this random effects group analysis are shown in Fig. 2.

We observed significant ($p < 0.05$, FWE corrected, see Table 1) increases in the BOLD signal during the RHI versus control condition in the left middle occipital gyrus (mOCG), spanning into the left anterior occipital sulcus ($x = -42$, $y = -68$, $z = 8$, $t = 3.54$), and in the anterior insula ($x = -40$, $y = 16$, $z = 10$, $t = 4.16$); there also was activity in the right anterior insula ($x = 48$, $y = 8$, $z = 10$, $t = 3.43$, $p < 0.001$ uncorrected).

Localization of extrastriate body area

Notably, the activity in the left mOCG for the RHI versus control condition (Fig. 2) comprised coordinates of the extrastriate body area as reported by several studies (Astafiev et al., 2004; Downing et al., 2001). However, the EBA is adjacent to, and may be overlapping with the motion-sensitive region hMT + (Peelen and Downing, 2007; Spiridon

et al., 2006). We therefore ran standard functional localizers for EBA and hMT + in 16 of the same participants in a separate scanning session (see Materials and methods), to attribute the activation in the left mOCG either to the EBA (a body part-selective region), or hMT + (a motion-sensitive region) in our own sample.

For the group-level contrast BODY–OBJECT, i.e., vision of body parts versus motorcycle parts (Fig. 3A, following Urgesi et al., 2007), we found significant ($p < 0.05$, FWE) activations in bilateral middle occipital gyrus (LH: $x = -52$, $y = -66$, $z = 6$, $t = 7.62$; RH: $x = 56$, $y = -60$, $z = 14$, $t = 6.49$), left superior parietal lobule ($x = -32$, $y = -44$, $z = 56$, $t = 5.52$), and left IPC/supramarginal gyrus ($x = -62$, $y = -28$, $z = 28$, $t = 5.46$), see Fig. 3B. Thus the EBA localizer produced strongest activity in bilateral mOCG, in agreement with published findings (Astafiev et al., 2004; Costantini et al., 2011). Next, we calculated the MOTION–STATIC contrast to locate hMT +, and found significant ($p < 0.05$, FWE) activity in the bilateral middle and inferior occipital cortex. The resulting activation in the left mOCG ($x = -42$, $y = -70$, $z = 0$, $t = 10.62$) was located more posterior and more inferior, and only marginally overlapped with the left EBA as defined by the BODY–OBJECT contrast (only 3 common voxels, see Fig. 3B), and not at all with the activation found for the RHI versus control condition (no common voxels, see also Fig. 2 and Table 1). Most importantly, the activation within the left mOCG as obtained from the RHI–CONTROL contrast was largely (72.7% of voxels, or 86.4% of voxels, for a mask threshold of $p < 0.05$, FWE, or $p < 0.001$, uncorrected) located within the area defined by the BODY–OBJECT contrast, and correspondingly significant at $p < 0.05$ FWE using small volume correction

Table 1
Group results: BOLD signal differences between the RHI versus control condition.

Anatomical location	Peak MNI ($p < 0.001$)			Peak t value
	x	y	z	
L. middle occipital gyrus	−42	−68	8	3.54[a]
L. anterior insula	−40	16	10	4.16[a]
R. anterior insula	48	8	10	3.49

Significant BOLD activations for the contrasts RHI–CONTROL.
[a] $p < 0.05$ FWE corrected based on pre-defined ROIs.
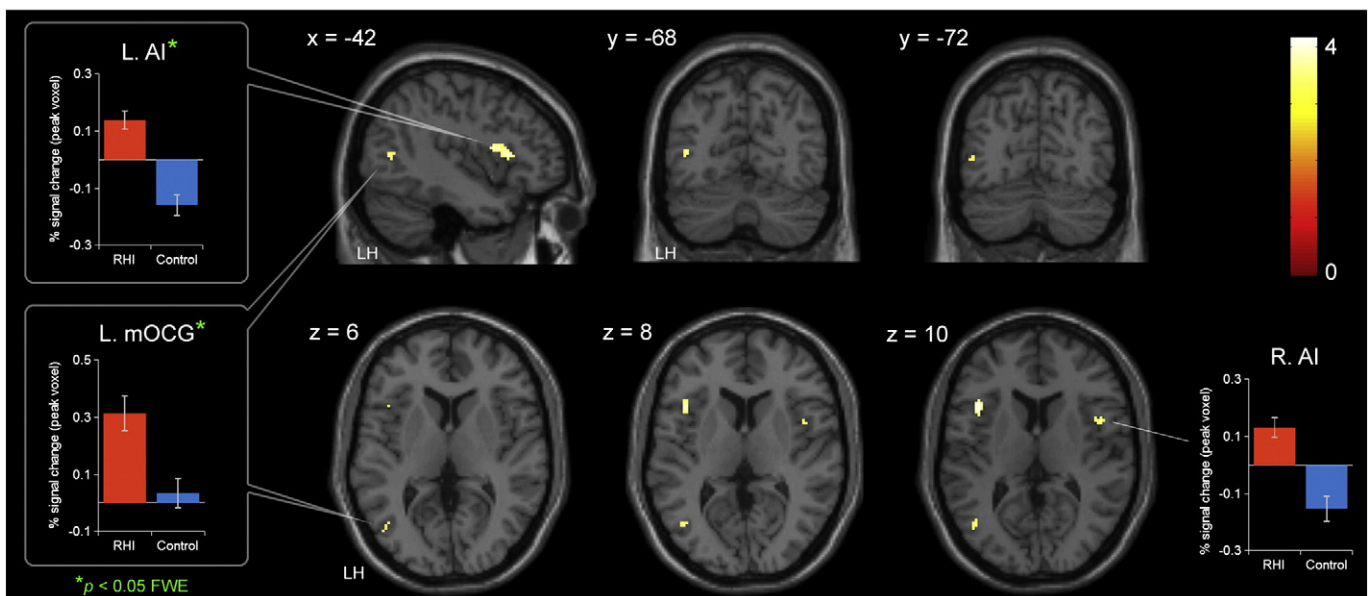
Fig. 2. Significant ($p < 0.05$ FWE correction based on pre-defined ROIs marked by a green asterisk) group-level BOLD signal differences between the Rubber Hand Illusion versus control condition ($RHI_{palm/palm} + RHI_{arm/arm}) - (CONTROL_{palm/arm} + CONTROL_{arm/palm})$. The statistical parametric maps of the T-contrast (superimposed onto the single-subject T1 template of SPM8, displayed at $p < 0.001$) show stronger activity in the left middle occipital gyrus (L. mOCG; $x = -42$, $y = -68$, $z = 8$, $t = 3.53$), left anterior insula (L. AI; $x = -40$, $y = 16$, $z = 10$, $t = 4.16$), and right anterior insula (R. AI; $x = 48$, $y = 8$, $z = 10$, $t = 3.49$, n.s.). The bar graphs show the mean BOLD signal changes (in percent) at peak voxels within each of these regions during the RHI and control conditions, error bars are standard errors of the mean. LH, left hemisphere.
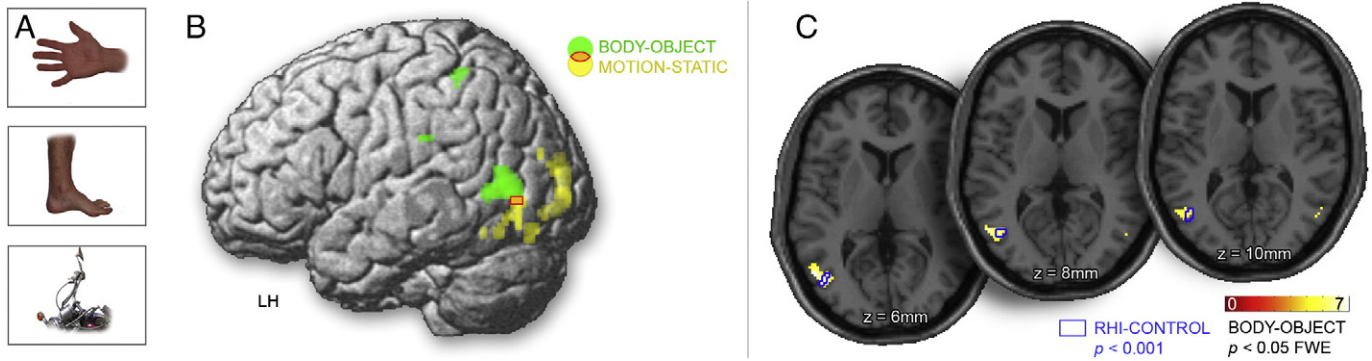
Fig. 3. (A) Sample stimuli used for locating specific responses to vision of body parts versus objects (EBA localizer). (B) Surface render of the significant ($p < 0.05$, FWE) voxels obtained from the EBA localizer (BODY–OBJECT, in green) and the MT+ localizer (MOTION–STATIC, in yellow). Overlap of these activations is indicated by a red outline (3 shared voxels). (C) Comparison of the location of activity in the left mOCG: The significant cluster in the left mOCG obtained from the BODY–OBJECT contrast ($p < 0.05$, FWE) contained 72.7% ($p < 0.05$ FWE, small volume corrected) of the significant voxels in the left mOCG as obtained from the RHI–CONTROL contrast ($p < 0.001$, uncorrected, cluster volume marked by blue outline).

with the left EBA mask (see Figs. 3B and C). Thus the functional localizer confirmed that the activity in the left mOCG for the RHI–CONTROL contrast can indeed be attributed to the left EBA, and does not overlap with the motion-sensitive area hMT+.

Activity differences in extrastriate body area correlate strongly positively with respective individual illusion scores
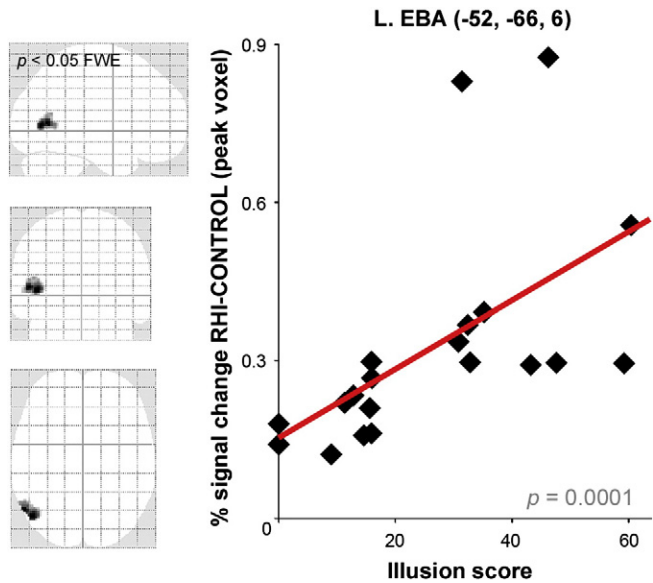
Our group-level comparison revealed a stronger *overall* activity of the left EBA and AI during the RHI versus control condition. However, the susceptibility to the RHI typically varies between individuals, and it is thus desirable to relate brain activity differences produced by the illusion to the observed between-subject differences in the experienced illusion. Therefore, we extracted the parameter estimates of participants' left EBA peak voxels of the RHI–CONTROL group-level contrast, selecting from within an independent ROI defined by the functional EBA localizer (Fig. 4, left) and correlated them with the respective behavioral illusion scores (a quantification of the intensity, i.e., strength

Fig. 4. Plot showing the strong positive correlation (Spearman's $\rho = 0.756$, $n = 20$, $p = 0.0001$) between BOLD signal differences at EBA peak voxels between RHI and control condition, and participants' respective illusion scores (see Materials and methods), and a least squares regression line. Peak voxels were selected from within the left EBA as localized by the BODY–OBJECT contrast in the second scanning session (see maximum intensity projection of the ROI mask, $p < 0.05$, FWE).

and prevalence of the ownership illusion during the RHI relative to the control condition, see Materials and methods). Participants' BOLD signal differences between RHI and control condition within the left EBA correlated significantly positively (Spearman's $\rho = 0.756$, $n = 20$, $p = 0.0001$) with their respective illusion scores: the higher participants scored on the illusion measure, the higher was the activity in their left EBA during the RHI versus control condition (Fig. 4, right). Activity in the left EBA thus directly reflected the subjectively experienced strength and prevalence of the ownership illusion of the dummy arm. Peak voxels within the left and right AI (regions of interest from the RHI–CONTROL group-level contrast) did not correlate significantly with illusion scores ($ps > 0.2$; the correlation of the left EBA voxels and illusion scores remained significant at $p < 0.001$ after Bonferroni correction accounting for the three tests performed). To test how specific the correlation of illusion scores in the left EBA was, we also calculated a whole-brain regression analysis on the RHI versus control contrast images, using the calculated illusion scores of our participants as a covariate. This analysis revealed that the intensity of the illusion was indeed significantly reflected only within the left EBA ($x = -52$, $y = -64$, $z = 2$, $t = 3.79$, $p < 0.05$ FWE, small volume corrected within the left EBA localizer mask) and the right PPC/IPS ($x = 18$, $y = -60$, $z = 66$, $t = 8.29$, $p < 0.05$ FWE). Importantly, only 3 of the voxels activated by this analysis were contained within the activation obtained from the hMT+ localizer (these did not survive small volume correction). This analysis thus confirmed the specificity of the relationship between EBA activity and the reported intensity of the illusion.

Stronger functional coupling among somatosensory cortex and extrastriate body area during the Rubber Hand Illusion
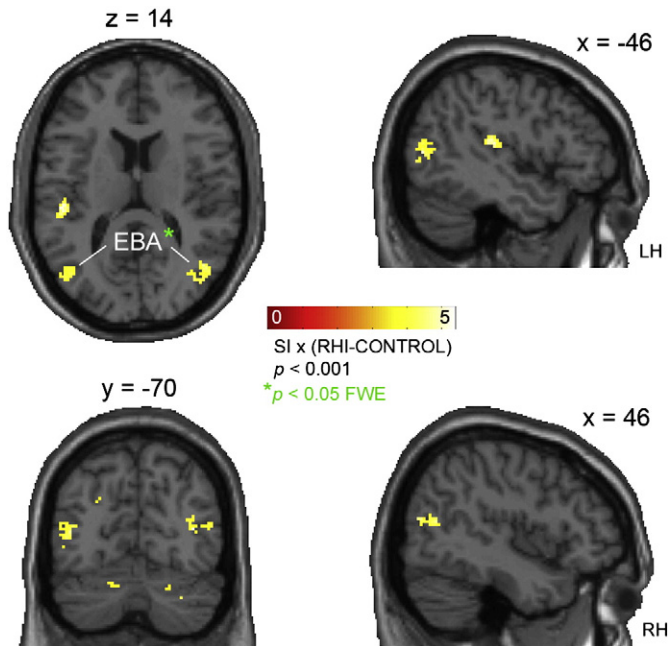
Activity differences between the RHI and control condition did not emerge in SI, which is not surprising due to the well-matched stimuli of the RHI and control condition (i.e., in the RHI–CONTROL contrast, tactile information was equal). However, we were interested in whether SI would still show a different connectivity pattern during the RHI versus control condition. Hence we calculated a PPI with seed regions located in the left SI (see Materials and methods). This analysis revealed that the left SI showed a significantly stronger coupling with the left ($x = -46$, $y = -72$, $z = 8$, $t = 4.98$) and right EBA ($x = 50$, $y = -66$, $z = 12$, $t = 3.99$) during the RHI versus control condition ($p < 0.05$, FWE, small volume corrected for bilateral EBA as defined by our functional localizer). Notably, the activity in the left mOCG revealed by this PPI also contained 45.5% of the voxels of the significant cluster in the left mOCG obtained from the RHI–CONTROL group-level contrast. Fig. 5 shows the location of significant voxels from this PPI analysis.

We also calculated a PPI analysis with seed regions located in the left EBA (see Materials and methods). This revealed that during the RHI

Fig. 5. Significant voxels obtained from the psychophysiological interaction analysis revealing stronger coupling of the left SI with the left (x = −46, y = −72, z = 8, t = 4.98, p < 0.05 FWE, small volume corrected) and right EBA (x = 50, y = −66, z = 12, t = 3.99, p < 0.05 FWE, small volume corrected), during the RHI versus control condition (displayed at p < 0.001, uncorrected).

versus control condition, the left EBA showed increased connectivity to a number of brain regions that were also activated by the EBA localizer (contrast BODY–OBJECT, thresholded at p < 0.001 uncorrected), namely the left supramarginal gyrus (x = −52, y = −26, z = 30, t = 5.60, p < 0.05 FWE, small volume corrected with the EBA localizer), right parietal operculum (x = 54, y = −26, z = 20, t = 4.67), and right anterior IPS (x = 40, y = −38, z = 50, t = 4.27). Although these activations partly did not survive correction for multiple comparisons, the coordinates correspond to those reported by previous related studies, which have demonstrated that these regions are involved in multisensory integration in hand-centered space (Brozzoli et al., 2011, 2012; Gentile et al., 2011; Makin et al., 2008).

Specific brain activity related to the periods before and after illusion onset

No premotor activity emerged in any of these analyses. However, in the fMRI study by Ehrsson et al. (2004), PMv activity during the RHI condition was specifically associated with the period after, relative to the period before illusion onset. We hence aimed at testing for similar interactions in our data. It should, however, be noted that in our experiment, illusion ratings were collected in the post-scanning phase, and thus these analyses have to be considered with some caution. Following the procedure described by Ehrsson et al. (2004), we first tested for activity differences between RHI and control condition during the period before participants reported the illusion, relative to the period after illusion onset. Coordinates and corrections for multiple comparisons of these activations are reported in Table 2. We replicated the findings by Ehrsson et al. (2004) during the pre-illusion period in the right dorsal premotor cortex and supplementary motor area, as well as in the left PPC/aIPS. Moreover, we found significant activations in the right TPJ (corresponding to published coordinates), left supramarginal gyrus (SMG), and bilateral EBA. Interestingly, the activations we found in EBA, PPC/aIPS, and SMG during the pre-illusion period corresponded to activations within brain regions produced by the EBA localizer (see Table 2). Next, we tested for brain activity specifically associated with the illusion period, relative to the period before onset, i.e., the inverse interaction. We found significant (p < 0.05 FWE) activity in the left

Table 2
Interaction: brain activity differences between RHI versus control condition before, relative to after illusion onset.

Anatomical location	Peak MNI (p < 0.001)			Peak t value
	x	y	z	
L. supramarginal gyrus	−46	−72	8	5.95[ac]
R. TPJ	50	−36	26	5.82[ab]
L. precuneus	−14	−66	34	5.59[a]
R. lateral occipital cortex (EBA)	52	−64	12	4.68[c]
R. dorsal premotor cortex	42	−2	58	4.46[b]
R. supplementary motor area	12	2	76	4.40[b]
L. PPC/aIPS	−28	−48	60	4.34[bc]
L. lateral occipital cortex (EBA)	−48	−72	12	4.21[c]

Stronger activity during the RHI versus control condition in the period before, relative to the period after reported illusion onset (p < 0.05 FWE corrected based on [a]whole-brain; [b]pre-defined ROIs; [c]EBA localizer).

paracentral lobule, spanning to the left precuneus (x = −6, y = −32, z = 80, t = 5.90). We also found activity in the left (x = −32, y = 20, z = 54, t = 5.19) and right (x = 40, y = 20, z = 56, t = 3.67) dorsolateral prefrontal cortex, left superior temporal gyrus (x = −44, y = −50, z = 12, t = 5.08) and right cerebellum (x = 32, y = −66, z = −38; t = 5.07), but none of these activations survived statistical correction for multiple comparisons. We did not find significant activity in PMv even when the statistical threshold was lowered to p < 0.01 uncorrected.

Discussion

Illusory ownership of a dummy arm was successfully induced by our novel, fully automated RHI setup, as indicated by the participants' ratings. Moreover, we found significant brain activity differences between the RHI and control condition in several of the expected brain regions. Our results, in particular the correlation of illusion scores and left EBA activity, further demonstrate a correspondence of behavioral and neural measures of illusory ownership. As we excluded the possibility that another person's presence or actions would bias participants' neural responses during stimulation, the resulting brain activity changes can be interpreted as directly underlying the illusory limb ownership, caused by congruent multisensory stimulation independent of social interaction. This is further supported by the fact that behavioral and BOLD effects of the RHI were independent of participants' trait empathy scores, which suggests that the induced ownership experience cannot be explained as a mere empathic reaction, but involves more basic mechanisms of body ownership. We will now discuss the individual findings in more detail.

Activity in extrastriate body area reflects illusory limb ownership

We found stronger brain activity in EBA during the RHI versus control condition, contralateral to the stimulated arm. ROI analyses based on published coordinates and an independent functional localizer session verified that this activation was indeed located in the body-part selective EBA. Importantly, BOLD signal responses in EBA were reflecting not only group-level differences between the RHI and the control condition, but also correlated strongly positively with interindividual differences in intensity of the RHI experience. Note also that these differences in EBA activity emerged even though a human-like arm was visible throughout the whole experiment, which alone should suffice to activate EBA.

The importance of vision for body-perception (Peelen and Downing, 2007) and specifically for the RHI (Armel and Ramachandran, 2003; Botvinick and Cohen, 1998; Pavani et al., 2000; Tsakiris and Haggard, 2005) is widely acknowledged, and has been supported by the demonstration of strong modulatory effects of vision on touch perception in peri-hand space (Làdavas et al., 2000; Lamm and Decety, 2008; Makin

et al., 2007, 2008). However, the role of visual cortex has been mainly defined as representing the dummy hand's position in space (e.g. Makin et al., 2008). Our results now provide novel evidence that EBA is directly involved in body ownership, and thus complement recent advances in understanding the functional role of the EBA: This region responds selectively to vision of bodies and body parts (Downing and Peelen, 2011; Downing et al., 2001; Pitcher et al., 2009), changes in limb position (Astafiev et al., 2004), actions with the same limb (Orlov et al., 2010), and mental imagery of embodied self-location (Arzy et al., 2006). It has moreover been suggested that the EBA integrates visual representations of the body with somatosensory information about body parts (Costantini et al., 2011; see also Apps and Tsakiris, 2013), and that it is involved in self-identification with a body (Ionta et al., 2011). However, there is an ongoing debate about the exact nature and specifically the dynamics of representations in EBA (see Downing and Peelen, 2011 for a discussion). Our results seem to suggest an at least somewhat sophisticated function of EBA, as discussed in the following.

Interactions of visuo-tactile systems during the RHI

Analyses of brain connectivity based on PPIs revealed a stronger coupling between the left SI and bilateral EBA during the RHI versus control condition, despite the well-matched stimulations in the RHI and control condition. This finding complements the results of BOLD signal differences between the RHI versus control condition, as it suggests that the EBA not only responds to the RHI, but that the somatosensory cortex also interacts more closely with this region during illusory limb ownership. It is also noteworthy that a second PPI analysis revealed that, during the RHI versus control, left EBA activity was more strongly coupled to a number of body-selective areas that have been shown to integrate multisensory information in hand-centered space (see e.g. Brozzoli et al., 2011, 2012; Gentile et al., 2011). These results support the claim that EBA is involved in integrating somatosensory with visual information about the body (Costantini et al., 2011). Together with the somatosensory system, EBA may thus be part of the often proposed "body representation" into which multisensory input must be integrated to be self-attributed (Costantini and Haggard, 2007; Kammers et al., 2006; Tsakiris, 2010; Tsakiris et al., 2007).

Recently, it has been proposed that one function of EBA could be to minimize prediction error within a hierarchical generative model of sensory input (Apps and Tsakiris, 2013; Saygin et al., 2012). These accounts follow the assumption that the brain contains hierarchical generative models that predict its sensory input (Friston, 2010; Friston and Kiebel, 2009; Hohwy, 2007). In an inversion of such a model, the driving signal is now the prediction error (the discrepancy between predicted and actual sensory input), which has to be explained away at some level of the hierarchy. This notably fits well with the classical assumption that illusory percepts emerge from Bayesian inference, i.e., an interpretation of ambiguous sensory input under a prior model (Apps and Tsakiris, 2013; Friston, 2005). During the RHI, for instance, observed and felt touch are "bound together" by these inference mechanisms (Hohwy, 2012), which explain away prediction error associated with discrepant visual, tactile, and proprioceptive input (Apps and Tsakiris, 2013; Hohwy, 2010). Although a detailed discussion of predictive coding accounts (Friston, 2010) is beyond the scope of this discussion, we would like to emphasize how well these map onto empirical data and theoretical accounts of multisensory self-processing (Blanke, 2012; Hohwy, 2007, 2010; Tsakiris, 2010; see Limanowski and Blankenburg, 2013, for a review). For our data, one potential explanation would be that indeed EBA and the interacting body-selective regions explain away prediction error that is associated with sensory input during the illusion (such as the discrepancy in visual appearance between the own and dummy arms). This would explain why activity in these regions was stronger in the period before illusion onset. In fact, a recent theoretical paper (Apps and Tsakiris, 2013) has proposed a cortical

network subserving prediction error minimization during the RHI, involving the EBA at intermediate levels, and AI, rTPJ, and PMv as multimodal areas at higher levels. Interestingly, a recent study (Apps et al., 2013) found that, along with activity in multimodal rTPJ and IPS, activity in unimodal visual cortex was related to illusory self-identification with another face induced by multisensory stimulation (a paradigm similar to the RHI). Specifically, illusory ownership of a face was related to activity in the face-selective occipital face area, which nicely complements our findings of an involvement of the body-selective EBA during illusory ownership of an arm. Our results are thus in line with the findings by Apps et al. (2013) and with the claim that representations of the self are dynamically updated during these experimentally induced illusions (Apps and Tsakiris, 2013; Apps et al., 2013; Hohwy, 2012). However, as our experiment was not designed to test these theories directly, future studies will have to address whether EBA is involved in perceptual assimilation of the dummy arm during the illusion (Longo et al., 2009), and if such effects can indeed be explained within a predictive coding framework (Apps and Tsakiris, 2013; Friston, 2010; Hohwy, 2007).

It is noteworthy that activity in posterior regions, including rTPJ, left PPC/aIPS and SMG, and bilateral EBA, in addition to PMd and SMA, was stronger during the period before illusion onset. These regions are well-known to receive multimodal input (Blanke, 2012; Brozzoli et al., 2011, 2012; Gentile et al., 2011; Petkova et al., 2011; Tsakiris, 2010). Thus we partly replicated the results of Ehrsson et al. (2004), who found several of these regions to be similarly involved in the "re-calibration phase" before illusion onset, presumably by resolving inter-sensory conflict (Makin et al., 2008). The strong activation of rTPJ during this phase should be mentioned in particular, as this region has been proposed to represent and integrate information in internal models of the body (Ionta et al., 2011; Tsakiris et al., 2008). Lesions to rTPJ are associated with out-of-the-body experiences (Blanke et al., 2002, 2005), and similarly, experimentally manipulated self-location activates this region (Ionta et al., 2011).

Finally, our results suggest that several of the regions showing an early activation during the RHI are also body part-selective, as demonstrated by their activation by the EBA localizer. In this light, the reflection of illusion scores by activity in the left EBA and right PPC we found in the regression analysis is particularly interesting, as both regions were also activated by the EBA localizer. Right PPC is known to integrate spatio-temporal information and represent external reference frames (Azañón et al., 2010; Dijkerman and de Haan, 2007; Tsakiris, 2010), and has often been shown to be involved in the RHI (Evans and Blanke, 2012; Makin et al., 2008) and in self-other differentiation in general (Decety and Sommerville, 2003; Shimada et al., 2005). In conclusion, it seems that the experience of illusory hand ownership during the RHI is enabled by an early activation of a network of multimodal body-selective areas.

Anterior insula is active during the Rubber Hand Illusion

We found significant activity differences in bilateral anterior insula during the RHI versus control condition. Previous experiments have revealed an involvement of the insula in the RHI: In a PET study (Tsakiris et al., 2007), activity in the right insula reflected the mislocalization of participants' own arm during the RHI. An fMRI study (Ehrsson et al., 2007) found increased activity in bilateral AI when a dummy hand was threatened during the illusion; this threat response also correlated positively with participants' ownership ratings. While one of the main functions of AI is interoception, a more general role of AI in self-related information processing has been suggested by recent proposals. In the human insular cortex, a posterior-to-anterior increase in the complexity of representations has been suggested (Craig, 2010, 2011; Lamm and Singer, 2010), with AI involved in representations of the self, interoception, and self-awareness (Craig, 2009, 2011; Critchley et al., 2004). AI also seems to be engaged in a sense of agency (Tsakiris, 2010). A role of AI in body ownership has also been implied by the fact that individuals with an obsessive desire to amputate their limb

have a smaller cortical volume of AI, predominantly contralateral to the affected limb (Hilti et al., 2012). Crucially, AI also seems to be involved in the prediction and integration of intero- and exteroceptive information (Lamm and Singer, 2010; Singer and Lamm, 2009). Seth et al. (2011) have proposed a role of AI in a model of integrated self-representation based on interoception, "alongside models of body ownership based on proprioception and multisensory integration". The link between interoceptive and exteroceptive self-processing receives support from studies using the RHI paradigm, in which interoceptive sensibility predicted the susceptibility to the RHI (Tsakiris et al., 2011), and illusory hand or full body ownership has been shown to influence homeostatic regulation (Moseley et al., 2008; Salomon et al., 2013). Interestingly, the AI has also been suggested to be involved in conscious error processing per se (Klein et al., 2013). Correspondingly, in a recent predictive coding account of self-recognition (Apps and Tsakiris, 2013, Fig. 1), the AI is emphasized as a multimodal brain area involved in explaining away prediction error associated with the Rubber Hand Illusion. Thus, although our procedure did not include threatening the dummy arm or locating the own arm, the stronger activity we found in bilateral AI during the RHI versus control condition supports the assumed importance of AI in self-perception and body ownership.

Differences to previous findings and potential limitations

We developed a fully automated setup, using computer-controlled stroking to eliminate the human agent from the RHI induction. The somatotopic arrangement enabled us to use a combined contrast to compare brain activity in the RHI versus control condition. Moreover, our control condition allowed us to present visual and tactile stimuli simultaneously, and thus to avoid the potentially problematic serial presentation of observed and felt touch (see Introduction). Contrasting with previous findings (cf. Ehrsson et al., 2004, 2005; Petkova et al., 2011), we failed to replicate effects of the illusion in ventral premotor cortex. It should be emphasized that these results have to be considered with caution, as our participants rated the strength and onset of the illusion off-line, i.e., after the scanning session (cf. Ehrsson et al., 2004). However, post-scanning ratings have been employed elsewhere (Petkova et al., 2011), and as our rating session followed immediately after the last scanning run, and participants lay in the scanner exactly as during the image acquisition, the ratings should still reflect the experienced illusion to a sufficient degree. Interestingly, a PET study (Tsakiris et al., 2007) that also used an automated setup to induce the RHI also found no PMv activity during the RHI. Similarly, a recent study using fully automated multisensory stimulation to induce illusory self-identification with another face also failed to replicate any involvement of PMv in the ownership illusion (Apps et al., 2013). This suggests that PMv activity could reflect another human agent touching one's own body—thus it would still be a measure of the illusion (and not as prominent in the control condition), albeit an indirect one. Another possible explanation for the different findings could be that premotor activity might be enabled by non-motor functions of the premotor cortex like attentional control or working memory (Schubotz et al., 2010; see Schubotz and von Cramon, 2003, for a detailed review). Premotor activity has been associated with "prospective attention to sensory events", i.e., sensory predictions (Schubotz and von Cramon, 2002, 2003) even for abstract, non-biological stimuli (Schubotz and von Cramon, 2004). It could thus be that PMv activity did not emerge in our study because stimulations in both RHI and control condition were synchronous, and thus did not differ in terms of temporal attentional demands. However, we did not directly contrast the automated induction of the RHI with an induction performed by a human experimenter, or spatially incongruent with temporally asynchronous stroking, and therefore our explanation remains speculative.

While our design eliminated differences in temporal attentional demands between RHI and control condition, it could be argued that these now differed in terms of crossmodal spatial attention. Crossmodal

spatial and temporal attention seems to play an important role in tasks involving visuo-tactile interactions (Macaluso and Driver, 2001; Pavani et al., 2000; Spence et al., 2000). However, as noted recently by Macaluso and Maravita (2010), the effects of visuo-tactile interactions in peripersonal space "do not appear to be merely related to spatial attention control" and "can trigger specialized processes associated with embodiment and sense of body ownership." Also, one would not expect such a specific effect as we found in the left EBA (contralateral to the stimulated arm), because by the calculation of the joint contrast of both RHI and both control conditions, differences between them was averaged out. Finally, it should also be mentioned that the body- and motion-selective responses we found in lateral occipital cortex overlapped in some analyses. However, the effects of the illusory experience were specific to the body-selective regions as defined by the EBA localizer.

Conclusion

Using a novel, fully automated fMRI setup, we induced illusory limb ownership in healthy participants, isolated from social interaction. Thereby we have demonstrated for the first time that the extrastriate body area (EBA) not only shows a preference for seeing body parts (which we replicated using a functional localizer), but particularly when those are also experienced as part of one's own body. This interpretation is supported by the fact that (i) EBA activity is significantly stronger during the RHI versus control condition, (ii) activity differences between RHI and control condition in the left EBA correlate strongly positively with participants' respective behavioral illusion scores, and (iii) SI contralateral to the stimulated arm is more strongly coupled with bilateral EBA during the RHI versus control condition. Our results thus provide novel evidence for dynamic representations in EBA, and show that the RHI paradigm can be used to gain further insight into the functional role of the EBA.

Acknowledgments

References

Apps, M.A., Tsakiris, M., 2013. The free-energy self: a predictive coding account of self-recognition. Neurosci. Biobehav. Rev. http://dx.doi.org/10.1016/j.neubiorev.2013.01.029 (Epub ahead of print).

Apps, M.A., Tajadura-Jiménez, A., Sereno, M., Blanke, O., Tsakiris, M., 2013. Plasticity in unimodal and multimodal brain areas reflects multisensory changes in self-face identification. Cereb. Cortex. http://dx.doi.org/10.1093/cercor/bht199 (Epub ahead of print).

Armel, K.C., Ramachandran, V.S., 2003. Projecting sensations to external objects: evidence from skin conductance response. Proc. Biol. Sci. 270, 1499–1506.

Arzy, S., Thut, G., Mohr, C., Michel, C.M., Blanke, O., 2006. Neural basis of embodiment: distinct contributions of temporoparietal junction and extrastriate body area. J. Neurosci. 26, 8074–8081.

Astafiev, S.V., Stanley, C.M., Shulman, G.L., Corbetta, M., 2004. Extrastriate body area in human occipital cortex responds to the performance of motor actions. Nat. Neurosci. 7, 542–548.

Azañón, E., Longo, M.R., Soto-Faraco, S., Haggard, P., 2010. The posterior parietal cortex remaps touch into external space. Curr. Biol. 20, 1304–1309.

Bernhardt, B.C., Singer, T., 2012. The neural basis of empathy. Ann. Rev. Neurosci. 35, 1–23.

Blanke, O., 2012. Multisensory brain mechanisms of bodily self-consciousness. Nat. Rev. Neurosci. 13, 556–571.

Blanke, O., Metzinger, T., 2009. Full-body illusions and minimal phenomenal selfhood. Trends Cogn. Sci. 13, 7–13.

Blanke, O., Mohr, C., 2005. Out-of-body experience heautoscopy and autoscopic hallucination of neurological origin Implications for neurocognitive mechanisms of corporeal awareness and self-consciousness. Brain Res. Rev. 50, 184–199.

Blanke, O., Ortigue, S., Landis, T., Seeck, M., 2002. Stimulating illusory own-body perceptions. Nature 419, 269–270.

Blanke, O., Mohr, C., Michel, C.M., Pascual-Leone, A., Brugger, P., Seeck, M., Landis, T., et al., 2005. Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. J. Neurosci. 25, 550–557.

Botvinick, M., Cohen, J., 1998. Rubber hands "feel" touch that eyes see. Nature 391, 756.

Brainard, D.H., 1997. The Psychophysics Toolbox. Spat. Vis. 10, 433–436.

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. J. Neurosci. 31, 9023–9031.

Brozzoli, C., Gentile, G., Ehrsson, H.H., 2012. That's near my hand! Parietal and premotor coding of hand-centered space contributes to localization and self-attribution of the hand. J. Neurosci. 32, 14573–14582.

Carlsson, K., Petrovic, P., Skare, S., Peterson, K.M., Ingvar, M., 2000. Tickling expectations: neural processing in anticipation of a sensory stimulus. J. Cogn. Neurosci. 12, 691–703.

Costantini, M., Haggard, P., 2007. The Rubber Hand Illusion: sensitivity and reference frame for body ownership. Conscious. Cogn. 16, 229–240.

Costantini, M., Urgesi, C., Galati, G., Romani, G.L., Aglioti, S.M., 2011. Haptic perception and body representation in lateral and medial occipito-temporal cortices. Neuropsychologia 49, 821–829.

Craig, A.D.B., 2009. How do you feel-now? The anterior insula and human awareness. Nat. Rev. Neurosci. 10, 59–70.

Craig, A.D.B., 2010. The sentient self. Brain Struct. Funct. 214, 563–577.

Craig, A.D.B., 2011. Significance of the insula for the evolution of human awareness of feelings from the body. Ann. N. Y. Acad. Sci. 1225, 72–82.

Critchley, H.D., Wiens, S., Rotshtein, P., Ohman, A., Dolan, R.J., 2004. Neural systems supporting interoceptive awareness. Nat. Neurosci. 7, 189–195.

Davis, M.H., 1983. Measuring individual differences in empathy: evidence for a multidimensional approach. J. Pers. Soc. Psychol. 44, 113–126.

Decety, J., Sommerville, J.A., 2003. Shared representations between self and other: a social cognitive neuroscience view. Trends Cogn. Sci. 7, 527–533.

Dijkerman, H.C., de Haan, E.H.F., 2007. Somatosensory processes subserving perception and action. Behav. Brain Sci. 30, 189–239.

Downing, P.E., Peelen, M.V., 2011. The role of occipitotemporal body-selective regions in person perception. Cogn. Neurosci. 2, 37–41.

Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293, 2470–2473.

Ebisch, S.J.H., Perrucci, M.G., Ferretti, A., Del Gratta, C., Romani, G.L., Gallese, V., 2008. The sense of touch: embodied simulation in a visuotactile mirroring mechanism for observed animate or inanimate touch. J. Cogn. Neurosci. 20, 1611–1623.

Ehrsson, H.H., 2007. The experimental induction of out-of-body experiences. Science 317, 1048.

Ehrsson, H.H., Spence, C., Passingham, R.E., 2004. That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. Science 305, 875–877.

Ehrsson, H.H., Holmes, N.P., Passingham, R.E., 2005. Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. J. Neurosci. 25, 10564–10573.

Ehrsson, H.H., Wiech, K., Weiskopf, N., Dolan, R.J., Passingham, R.E., 2007. Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. PNAS 104, 9828–9833.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage 25, 1325–1335.

Evans, N., Blanke, O., 2012. Shared electrophysiology mechanisms of body ownership and motor imagery. NeuroImage 64, 216–228.

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 815–836. http://dx.doi.org/10.1098/rstb.2005.1622.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11, 127–138.

Friston, K., Kiebel, S., 2009. Predictive coding under the free-energy principle. Philos. Trans. R. Soc. B-Biol. Sci. 364, 1211–1221. http://dx.doi.org/10.1098/rstb.2008.0300.

Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6, 218–229.

Gallagher, S., 2000. Philosophical conceptions of the self: implications for cognitive science. Trends Cogn. Sci. 4, 14–21.

Gentile, G., Petkova, V.I., Ehrsson, H.H., 2011. Integration of visual and tactile signals from the hand in the human brain: an FMRI study. J. Neurophysiol. 105, 910–922.

Gläscher, J., 2009. Visualization of group inference data in functional neuroimaging. Neuroinformatics 7, 73–82.

Hilti, L.M., Hänggi, J., Vitacco, D.A., Kraemer, B., Palla, A., Luechinger, R., Jäncke, L., et al., 2012. The desire for healthy limb amputation: structural brain correlates and clinical features of xenomelia. Brain 136, 318–329.

Hohwy, J., 2007. The sense of self in the phenomenology of agency and perception. Psyche 13, 1–20.

Hohwy, J., 2010. The hypothesis testing brain: some philosophical applications. ASCS09: Proceedings of the 9th Conference of the Australasian Society for Cognitive Science, pp. 135–144.

Hohwy, J., 2012. Attention and conscious perception in the hypothesis testing brain. Front. Psychol. 3, 96. http://dx.doi.org/10.3389/fpsyg.2012.00096.

Holmes, N.P., Spence, C., 2004. The body schema and multisensory representation(s) of peripersonal space. Cogn. Process. 5, 94–105.

Ionta, S., Heydrich, L., Lenggenhager, B., Mouthon, M., Fornari, E., Chapuis, D., Gassert, R., et al., 2011. Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective. Neuron 70, 363–374.

Jeannerod, M., 2007. Being oneself. J. Physiol. Paris 101, 161–168.

Kammers, M.P.M., Van der Ham, I.J.M., Dijkerman, H.C., 2006. Dissociating body representations in healthy individuals: differential effects of a kinaesthetic illusion on perception and action. Neuropsychologia 44, 2430–2436.

Kanayama, N., Sato, A., Ohira, H., 2007. Crossmodal effect with Rubber Hand Illusion and gamma-band activity. Psychophysiology 44, 392–402.

Kanayama, N., Sato, A., Ohira, H., 2009. The role of gamma band oscillations and synchrony on Rubber Hand Illusion and crossmodal integration. Brain Cogn. 69, 19–29.

Keysers, C., Kaas, J.H., Gazzola, V., 2010. Somatosensation in social perception. Nat. Rev. Neurosci. 11, 417–428.

Klein, T.A., Ullsperger, M., Danielmeier, C., 2013. Error awareness and the insula: links to neurological and psychiatric diseases. Front. Hum. Neurosci. 7, 14. http://dx.doi.org/10.3389/fnhum.2013.00014.

Kuehn, E., Trampel, R., Mueller, K., Turner, R., Schütz-Bosbach, S., 2012. Judging roughness by sight-A 7-Tesla fMRI study on responsivity of the primary somatosensory cortex during observed touch of self and others. Hum. Brain Mapp. 34, 1882–1895. http://dx.doi.org/10.1002/hbm.22031.

Làdavas, E., Farnè, A., Zeloni, G., Di Pellegrino, G., 2000. Seeing or not seeing where your hands are. Exp. Brain Res. 131, 458–467.

Lamm, C., Decety, J., 2008. Is the extrastriate body area (EBA) sensitive to the perception of pain in others? Cereb. Cortex 18, 2369–2373.

Lamm, C., Singer, T., 2010. The role of anterior insular cortex in social emotions. Brain Struct. Funct. 214, 579–591.

Lenggenhager, B., Tadi, T., Metzinger, T., Blanke, O., 2007. Video ergo sum: manipulating bodily self-consciousness. Science 317, 1096–1099.

Lenggenhager, B., Halje, P., Blanke, O., 2011. Alpha band oscillations correlate with illusory self-location induced by virtual reality. Eur. J. Neurosci. 33, 1935–1943.

Limanowski, J., Blankenburg, F., 2013. Minimal self-models and the free energy principle. Front. Hum. Neurosci. 7, 547. http://dx.doi.org/10.3389/fnhum.2013.00547.

Longo, M.R., Schüür, F., Kammers, M.P.M., Tsakiris, M., Haggard, P., 2009. Self awareness and the body image. Acta Psychol. 132, 166–172.

Lutti, A., Thomas, D.L., Hutton, C., Weiskopf, N., 2012. High-resolution functional MRI at 3 T: 3D/2D echo-planar imaging with optimized physiological noise correction. Magn. Reson. Med. 69, 1657–1664. http://dx.doi.org/10.1002/mrm.24398.

Macaluso, E., Driver, J., 2001. Spatial attention and crossmodal interactions between vision and touch. Neuropsychologia 39, 1304–1316.

Macaluso, E., Maravita, A., 2010. The representation of space near the body through touch and vision. Neuropsychologia 48, 782–795.

Macey, P.M., Macey, K.E., Kumar, R., Harper, R.M., 2004. A method for removal of global effects from fMRI time series. NeuroImage 22, 360–366.

Makin, T.R., Holmes, N.P., Zohary, E., 2007. Is that near my hand? Multisensory representation of peripersonal space in human intraparietal sulcus. J. Neurosci. 27, 731–740.

Makin, T.R., Holmes, N.P., Ehrsson, H.H., 2008. On the other hand: dummy hands and peripersonal space. Behav. Brain Res. 191, 1–10.

Mazaika, P., Hoeft, F., Glover, G.H., Reiss, A.L., 2009. Methods and software for fMRI analysis for clinical subjects. Paper Presented at the Annual Meeting of the Organization for Human Brain Mapping.

Moseley, G.L., Olthof, N., Venema, A., Don, S., Wijers, M., Gallace, A., Spence, C., 2008. Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. PNAS 105, 13169–13173.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Orlov, T., Makin, T., Zohary, E., 2010. Topographic representation of the human body in the occipitotemporal cortex. Neuron 68, 586–600.

Pavani, F., Spence, C., Driver, J., 2000. Visual capture of touch: out-of-the-body experiences with rubber gloves. Psychol. Sci. 11, 353–359.

Peelen, M.V., Downing, P.E., 2007. The neural basis of visual body perception. Nat. Rev. Neurosci. 8, 636–648.

Petkova, V.I., Björnsdotter, M., Gentile, G., Jonsson, T., Li, T.Q., Ehrsson, H.H., 2011. From part- to whole-body ownership in the multisensory brain. Curr. Biol. 21, 1118–1122.

Pitcher, D., Charles, L., Devlin, J.T., Walsh, V., Duchaine, B., 2009. Triple dissociation of faces bodies and objects in extrastriate cortex. Curr. Biol. 19, 319–324.

Salomon, R., Lim, M., Pfeiffer, C., Gassert, R., Blanke, O., 2013. Full body illusion is associated with widespread skin temperature reduction. Front. Behav. Neurosci. 7, 65. http://dx.doi.org/10.3389/fnbeh.2013.00065.

Saygin, A.P., Chaminade, T., Ishiguro, H., Driver, J., Frith, C., 2012. The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. Soc. Cogn. Affect. Neurosci. 7, 413–422.

Schaefer, M., Heinze, H.-J., Rotte, M., 2012. Embodied empathy for tactile events: interindividual differences and vicarious somatosensory responses during touch observation. NeuroImage 60, 952–957.

Schubotz, R.I., von Cramon, D.Y., 2002. A blueprint for target motion: fMRI reveals perceived sequential complexity to modulate premotor cortex. NeuroImage 16, 920–935.

Schubotz, R.I., von Cramon, D.Y., 2003. Functional–anatomical concepts of human premotor cortex: evidence from fMRI and PET studies. NeuroImage 20, S120–S131.

Schubotz, R.I., von Cramon, D.Y., 2004. Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observation and imagery. J. Neurosci. 24, 5467–5474.

Schubotz, R.I., Anwander, A., Knösche, T.R., von Cramon, D.Y., Tittgemeyer, M., 2010. Anatomical and functional parcellation of the human lateral premotor cortex. NeuroImage 50, 396–408.

Seth, A.K., Suzuki, K., Critchley, H.D., 2011. An interoceptive predictive coding model of conscious presence. Front. Psychol. 2, 395.

Shimada, S., Hiraki, K., Oda, I., 2005. The parietal role in the sense of self-ownership with temporal discrepancy between visual and proprioceptive feedbacks. NeuroImage 24, 1225–1232.

Singer, T., Lamm, C., 2009. The social neuroscience of empathy. Ann. N. Y. Acad. Sci. 1156, 81–96.

Spence, C., Pavani, F., Driver, J., 2000. Crossmodal links between vision and touch in covert endogenous spatial attention. J. Exp. Psychol. Hum. Percept. Perform. 26, 1298.

Spiridon, M., Fischl, B., Kanwisher, N., 2006. Location and spatial profile of category-specific regions in human extrastriate cortex. Hum. Brain Mapp. 27, 77–89.

Tootell, R.B., Reppas, J.B., Kwong, K.K., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R., et al., 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J. Neurosci. 15, 3215–3230.

Tsakiris, M., 2010. My body in the brain: a neurocognitive model of body-ownership. Neuropsychologia 48, 703–712.

Tsakiris, M., Haggard, P., 2005. The Rubber Hand Illusion revisited: visuotactile integration and self-attribution. J. Exp. Psychol. Hum. Percept. Perform. 31, 80–91.

Tsakiris, M., Hesse, M.D., Boy, C., Haggard, P., Fink, G.R., 2007. Neural signatures of body ownership: a sensory network for bodily self-consciousness. Cereb. Cortex 17, 2235–2244.

Tsakiris, M., Costantini, M., Haggard, P., 2008. The role of the right temporo-parietal junction in maintaining a coherent sense of one's body. Neuropsychologia 46, 3014–3018.

Tsakiris, M., Tajadura-Jiménez, A., Costantini, M., 2011. Just a heartbeat away from one's body: interoceptive sensitivity predicts malleability of body-representations. Proc. R. Soc. B-Biol. Sci. 278, 2470–2476.

Urgesi, C., Candidi, M., Ionta, S., Aglioti, S.M., 2007. Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. Nat. Neurosci. 10, 30–31.

Weissenbacher, A., Kasess, C., Gerstl, F., Lanzenberger, R., Moser, E., Windischberger, C., 2009. Correlations and anticorrelations in resting-state functional connectivity MRI: a quantitative comparison of preprocessing strategies. NeuroImage 47, 1408–1416.

Zaki, J., Ochsner, K., 2012. The neuroscience of empathy: progress pitfalls and promise. Nat. Neurosci. 15, 675–680.

Study 2

Limanowski, J., Blankenburg, F. (2015). Network activity underlying the illusory self-attribution of a dummy arm. *Human Brain Mapping (Epub ahead of print).* doi: 10.1002/hbm.22770
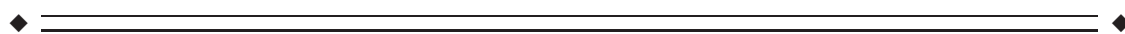
Network Activity Underlying the Illusory Self-Attribution of a Dummy Arm

Jakub Limanowski[1,2]* and Felix Blankenburg[1,2]

[1]*Neurocomputation and Neuroimaging Unit, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany*
[2]*Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany*

Abstract: Neuroimaging has demonstrated that the illusory self-attribution of body parts engages frontal and intraparietal brain areas, and recent evidence further suggests an involvement of visual body-selective regions in the occipitotemporal cortex. However, little is known about the principles of information exchange within this network. Here, using automated congruent versus incongruent visuotactile stimulation of distinct anatomical locations on the participant's right arm and a realistic dummy counterpart in an fMRI scanner, we induced an illusory self-attribution of the dummy arm. The illusion consistently activated a left-hemispheric network comprising ventral premotor cortex (PMv), intraparietal sulcus (IPS), and body-selective regions of the lateral occipitotemporal cortex (LOC). Importantly, during the illusion, the functional coupling of the PMv and the IPS with the LOC increased substantially, and dynamic causal modeling revealed a significant enhancement of connections from the LOC and the secondary somatosensory cortex to the IPS. These results comply with the idea that the brain's inference mechanisms rely on the hierarchical propagation of prediction error. During illusory self-attribution, unpredicted ambiguous sensory input about one's body configuration may result in the generation of such prediction errors in visual and somatosensory areas, which may be conveyed to parietal integrative areas. *Hum Brain Mapp 00:000–000, 2015.* © 2015 Wiley Periodicals, Inc.

Key words: extrastriate body area; body ownership; intraparietal sulcus; peripersonal space; predictive coding; rubber hand illusion; self-attribution

---
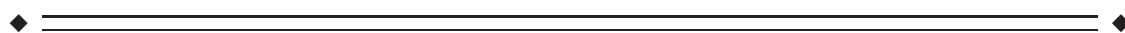
INTRODUCTION

The rubber hand illusion [RHI, Botvinick and Cohen, 1998] is an illusory self-attribution of a dummy hand induced via congruent touch on the dummy hand and one's real hand. Its classical explanation emphasizes the integration of conflicting information from vision, touch, and proprioception about one's hand's position in space [Botvinick and Cohen, 1998; Ehrsson 2012; Makin et al., 2008]. The RHI arises because the touch seen on the dummy hand and the touch felt on one's real hand are (falsely) bound together and taken to convey concurrent information about one and the same external event [Driver and Spence, 2000; Hohwy, 2012]. Due to the dominance of the visual modality, this event—the "felt" touch—is then attributed to the dummy hand [Armel and Ramachandran, 2003; Botvinick and Cohen, 1998]. Thus, by resolving the intersensory conflict via multisensory integration, the brain updates an internal multimodal body model to incorporate the dummy hand [Graziano and Botvinick, 2002; Maravita et al., 2003; Tsakiris, 2010].

In line with this explanation, brain imaging has demonstrated that the self-attribution of body parts during

the RHI activates parietal and frontal multimodal areas [Ehrsson et al., 2004; Petkova et al., 2011] and increases their functional coupling [Gentile et al., 2013; Guterstam et al., 2013]. These areas implement the processing of stimuli in the space surrounding the body [peripersonal space, (PPS); Graziano and Cooke, 2006; Graziano, 1999; Làdavas, 2002; Maravita et al., 2003; Rizzolatti et al., 1981, 1997], and it has been proposed that their increased (co)activation during the RHI reflects nonlinear responses to convergent multisensory information—indicative of multisensory integration—and a resulting recalibration of multimodal PPS coordinates onto the dummy hand and a coherent experience of body ownership [Brozzoli, Gentile et al. 2012; Ehrsson, 2012; Gentile et al., 2011; Makin et al., 2008]. Recent studies moreover suggest an important contribution of extrastriate visual areas in the lateral occipitotemporal cortex (LOC) to the RHI, most notably the body-selective extrastriate body area [EBA; Downing et al., 2001; Limanowski et al., 2014; Wold et al., 2014]. An involvement of visually body-selective areas in the RHI fits with previous reports of visuosomatosensory interactions in the EBA [Blanke, 2012; Costantini et al., 2011; Haggard et al., 2007], and with the fact that vision of body parts alone affects the processing of stimuli in the PPS [Graziano et al., 2000; Makin et al., 2007] and enhances the processing of tactile stimuli on the body itself [Haggard et al., 2007; Kennett et al., 2001; Taylor-Clarke et al., 2002]. In sum, frontoparietal and occipitotemporal brain areas seem to work in concert during the RHI to resolve intersensory conflicts by integrating multisensory information. However, which mechanisms within this network guide the brain's decision to self-attribute a body part or not is still largely based on speculations. Empirical evidence clarifying the nature of these mechanisms would substantially enrich existing models of body ownership and PPS processing [Brozzoli, Gentile et al., 2012; Graziano and Cooke, 2006; Makin et al., 2008; Tsakiris 2010].

A compelling speculation is that the interactions of the fronto-parietal network and body-selective occipitotemporal areas observed during the RHI may reflect reciprocal information exchange according to the principles of predictive coding [Friston and Kiebel, 2009; Friston and Stephan, 2007; Friston, 2005]. The predictive coding account of information processing within the brain is built upon the assumption that the brain constantly interprets its sensory information under a hierarchical generative model of the world. Thereby prediction errors are generated by a mismatch between predicted and actual sensory data at any given level and are passed on to the level above via bottom-up, feedforward connections. The guiding principle of the brain is to constantly minimize prediction error across all levels of the hierarchy to infer the causes of its current sensory state. This can be achieved by adjusting the model's predictions, which are conveyed to the level below via top-down, feedback connections to "explain away" the prediction error [Murray et al., 2002].

One important assumption derived from predictive coding is that stimulus-evoked neuronal activity increases reflect the computation and propagation of such prediction errors [Friston, 2005; Kok et al., 2012; Summerfield and Koechlin, 2008]. During the RHI, such prediction error-related activity increases should hence be observable in regions of the brain network subserving limb ownership that detect mismatches between the predictions of one's body model and the visuosomatosensory information provided [Apps and Tsakiris, 2014; Hohwy, 2012; Limanowski and Blankenburg, 2013].
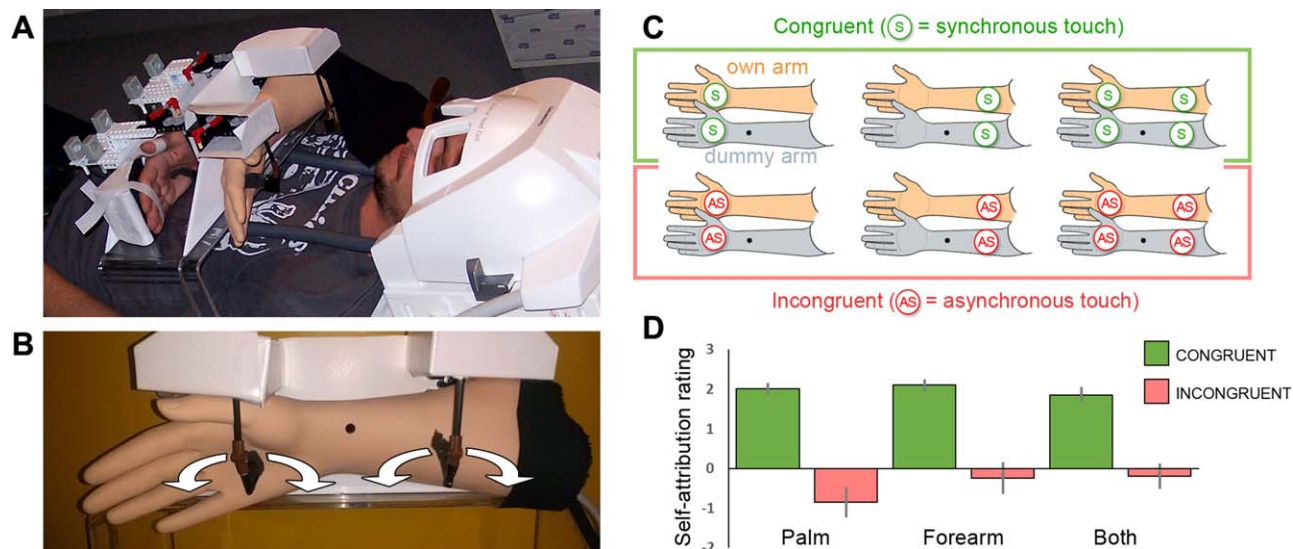
Here, we used an automated setup to deliver tactile stimulation to the participant's hidden right palm, forearm, or both, and synchronously or asynchronously to their counterpart on a realistic dummy arm inside an fMRI scanner. We hypothesized that the RHI would depend on the congruence of visuotactile stimulation, with neuronal mechanisms implemented within the same, potentially visually body-selective, brain areas for all locations. We therefore identified common effects of visuotactile congruence versus incongruence across stimulation locations, and independently tested for the visual body-selectivity of these brain areas. Moreover, we compared these effects to results of spatially (in)congruent stimulation, obtained using the same setup [Limanowski et al., 2014]. As expected, we observed increases of ventral premotor, intraparietal, and occipitotemporal activity during the RHI. Psychophysiological interaction analyses moreover demonstrated an increased functional coupling among these areas during the illusion. We further examined these illusion-related interactions using dynamic causal modeling [DCM, Friston et al., 2003]. Bayesian model comparison identified as the most parsimonious model one in which the RHI modulated the connectivity from lower-level visual and somatosensory areas to higher-level intraparietal areas. We interpret these results as support for a predictive coding account of hierarchical inference in the brain, whereby the probabilistic self-attribution of body parts during the RHI rests on the propagation of forward-flowing multisensory prediction errors to higher-level integrative brain areas.

MATERIALS AND METHODS

Participants

Twenty healthy volunteers [14 females, age 23–36 years, 19 right-handed, 1 left-handed as measured with the Edinburgh Handedness Inventory, Oldfield, 1971] participated in the main experiment. All participants issued informed consent before the experiment, and were paid after the scanning session. Participants were treated in accordance with the Human Subject Guidelines of the Declaration of Helsinki, and the experiment was approved by the local Ethical Committee of the Charité University Hospital Berlin.

Figure 1.

Experimental setup and design. We used an automatic setup to deliver tactile stimulation to two locations on the participant's right arm and a realistic ipsilateral dummy arm. **A:** The participants lay inside the MR scanner with their right arm hidden from view, while looking at the dummy arm. Two pairs of sponge brushes, driven by computer-controlled stepping motors, were installed at the palm and the forearm of each arm. **B:** View of the dummy arm from the participant's perspective. Participants were instructed to fixate the black dot in the middle of the dummy arm throughout the experiment. Visuo-tactile stimulation was delivered by 180° back-and-forth rotations of the brushes (symbolized with white arrows) at approx. 1.3 Hz. **C:** The experimental design for the main experiment was a 3 by 2 factorial block design, in which we manipulated the stimulation location (palm, forearm, or both), and the congruence of visuo-tactile stimulation (congruent versus incongruent). Thereby incongruent stimulation consisted of asynchronous touch, which was achieved by delaying the rotation of the felt brush strokes by 50 % relative to the rotation of the seen brush strokes. **D:** Participants' mean verbal ratings of the experienced self-attribution of the dummy arm during each condition. Error bars are standard errors of the mean. The congruent versus incongruent stimulation conditions produced a strong illusory self-attribution of the dummy arm at each anatomical location (Wilcoxon signed-ranks, $n = 20$, all $Zs > 3.5$, all $ps < 0.001$, see Results for details).

Experimental Setup and Design

We used a custom automatic setup to deliver tactile stimulation to a realistic, life-size right dummy arm and the participant's real arm (Fig. 1A). The dummy arm was mounted on a transparent acrylic console in full view atop the participant's chest, while her (fixated) real arm was occluded from view, placed approx. 13 cm behind the dummy arm in a corresponding posture. The participant's head was tilted and foam padded to guarantee full vision of the dummy arm. The gap between the dummy arm and the participant's shoulder was covered with black cloth. Two pairs of brushes were installed at corresponding anatomical locations of the dummy arm and the real arm—one at each palm, and one at each forearm (Fig. 1B). The brushes were adjusted and tested before the start of the experiment, and the perceived synchrony of brush strokes in the RHI condition was validated in a brief practice run. Participants were instructed to fixate a small black dot in the middle of the dummy arm throughout the whole experiment. Each brush could deliver touch via back and forth 180° rotations at a frequency of approx. 1.3 Hz with randomly varying inter-stimulus intervals (0 or 250 ms) to make stimulation less predictable. To eliminate any potential effects of human interaction or vision of another human's (the experimenter's) hands during the illusion induction, the experimental setup was completely automated. The brushes were driven by four identical stepping motors (1.8° stepping angle; RS Components GmbH, Mörfelden-Walldorf, Germany), which were controlled by a custom MATLAB (The Mathworks, Natick) script via a computer parallel port that also received the scanner-triggers to synchronize stimulation onsets with the fMRI acquisition. The motors' torque was transmitted from outside the scanner room to the brushes using nonmagnetic cables and gears.

The main experiment was conducted as a within-subject block design, comprising five experimental runs plus one separate run for visuotactile localizers, and a separate functional localizer for visually body part-selective brain areas (see below). During the experimental runs, tactile stimulation was applied to anatomically congruent locations of the real arm and the dummy arm, that is, to the palm, the forearm, or both locations

together (Fig. 1C) in two conditions: temporally congruent (synchronous touch at the same location) or incongruent (asynchronous touch at the same location, achieved by delaying the real arm brush strokes by 50%, i.e., approx. 400 ms). Each condition was presented twice per run and location in blocks of 20.16 s, followed by 12.6 s rest. The experiment also comprised two additional conditions where stimulation was synchronous at one location and asynchronous at the other, which we do not report here because of our explicit focus on clearly congruent versus incongruent multisensory information. Immediately after the functional runs, the verbal ratings of illusion intensity and its temporal onset were collected; for this purpose participants remained lying inside the scanner in the same way as during image acquisition, and the experimental conditions were presented again. Participants first indicated the strength of experienced illusory self-attribution of the dummy arm in the congruent and incongruent conditions ("During the stimulation, it felt as if the dummy arm was my own arm.", Botvinick and Cohen, 1998] on a 7-point Likert-scale ranging from −3 "completely disagree" to +3 "completely agree." Moreover, for the congruent condition only (as this was the condition in which we expected a RHI to occur), the elapsed time between the beginning of congruent stimulation and the participant's first verbal report of experienced illusory self-attribution of the dummy arm was measured with a stopwatch to represent the individual onset of the ownership illusion. Thus, in addition to the self-attribution ratings of each condition, we were able to calculate an illusion score that reflected both the strength and prevalence of the illusion: the ownership rating difference between the congruent and incongruent condition was multiplied by the difference between total stimulation duration minus the reported illusion onset [Ehrsson et al., 2004].

As part of our analysis, we compared our data to those reported in Limanowski et al. [2014; these data were obtained more than 6 months before the data acquired for the present experiment]. In that experiment ($N = 20$, two of which also took part in the present experiment), we have used the same experimental setup; in contrast to the present experiment, tactile stimulation was always synchronous, and applied either to congruent locations on the real arm and the dummy arm (palms or forearms) or to incongruent locations (one's palm together with the dummy forearm, or vice versa); only the former condition induced a RHI. By analyzing both datasets together, we were able to identify general effects of temporally (the present experiment) and spatially [Limanowski et al., 2014] congruent versus incongruent stimulation of the real arm and dummy arm underlying the illusion. Moreover, we were able to combine the functional localizer sessions for identification of visually body part-selective regions from the two samples, since the two protocols were identical.

Individual Localization of Visually Body Part-Selective Areas

The extrastriate body area (EBA) is a functionally defined area within the lateral occipital cortex, usually characterized by its preferential response to images of human body parts versus images of other objects [Downing et al., 2001; Downing and Peelen, 2011; Urgesi et al., 2007]. Therefore, for each participant, we implemented a standard functional localizer in a separate scanning session, in which we presented participants images of human hands and feet, and used images of motorcycle parts as control stimuli (following Urgesi et al., 2007, see Fig. 4 for sample stimuli). Images were color photographs presented on a blank white screen (18.7° × 13.7° visual angle) for 700 ms (150 ms interstimulus intervals) within blocks of 20 s duration and 20 s rest with a black fixation cross. Image categories were presented randomly, and the order of images within each category was randomized as well. Since the protocol for both datasets [present study and Limanowski et al., 2014] was identical, we were able to analyze all participants' data [$N = 36$ due to four dropouts in the dataset from Limanowski et al., 2014] in one group-level GLM, calculating the contrast Body parts vs Objects. The fMRI parameters, data preprocessing, and analyses used for the functional data obtained in this scanning session were identical as described in the following for the main experiment.

fMRI Data Acquisition, Preprocessing, and Analysis

The fMRI data were recorded using a 3 T scanner (Tim Trio, Siemens, Germany), equipped with a 32-channel head coil. T2*-weighted functional images were acquired using a customized high-resolution 3D-EPI sequence [Lutti et al., 2012]. Parallel imaging (GRAPPA image reconstruction) was used along the phase and partition directions with an acceleration factor of two, yielding a functional image resolution of 2.0 × 2.0 × 2.0 mm^3 at an acquisition time of 2520 ms per image volume (TR = 70 ms, matrix size [96, 96, 64], TE = 33 ms, flip angle = 20°, BW = 1408 Hz). A total of 1266 functional volumes were recorded for each participant (six runs à 211 volumes each), with an additional GRE field map (TE$_1$ = 10.00 ms, TE$_2$ = 12.46 ms) recorded after each scanning session. A high-resolution T1-weighted structural image was acquired for each participant (3D MPRAGE, voxel size = 1 × 1 × 1 mm^3, FOV = 256 mm × 256 mm, 176 slices, TR = 1900 ms, TE = 2.52 ms, flip angle = 9°). FMRI data were preprocessed and analyzed using SPM8 (Wellcome Department of Cognitive Neurology, London, UK: www.fil.ion.ucl.ac.uk/spm/). First, possible artifacts (i.e., artifacts that may be induced by abrupt head motion or spikes) in individual

slices of the functional image volumes were corrected via interpolation using the SPM ArtRepair toolbox [Mazaika et al., 2009; art_slice procedure, applying default settings]. To account for both dynamic (head motion-related) and static distortions, the functional images were realigned to the first image of each run using a least squares approach and a six-parameter rigid body transformation, and unwarped. To achieve better intersubject alignment, the functional images were spatially normalized using the DARTEL procedure as implemented in SPM8 [Ashburner, 2007]: Each participant's functional images were coregistered to the respective T1-weighted structural image. The structural images were segmented into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) components, and the GM and WM segments of all participants were aligned and warped together. The nonlinear deformation information from this step was then used to normalize each participant's functional images to MNI space. In the same way, we also normalized the individual structural images, and averaged them to obtain a sample-specific mean structural image onto which we projected our T-maps. Functional images were further spatially smoothed using a 6 mm full width at half maximum Gaussian kernel. To improve signal to noise ratio, global BOLD signal effects were removed from the functional images using a voxel-level linear model of the global signal [Macey et al., 2004], and subsequently the functional images were scanned for outlier volumes featuring excessive scan-to-scan movement, which were corrected via interpolation using the SPM ArtRepair toolbox (art_global procedure). Finally, to remove physiological noise from the BOLD signal of grey matter regions, we created participant-specific noise regressors using a component-based approach as described by Behzadi et al. [2007]: The WM and CSF structural segments were first smoothed with a 4 mm full width at half maximum Gaussian kernel. For each run, these segments were then multiplied with the first unwarped functional image (using the ImCalc function and thresholds of 80% for the CSF segment, and 90% for the WM segment). The resulting images were used as noise ROIs (regions in which the time signal is unlikely to reflect neural activity). From within these ROIs, we extracted the first five principal components that accounted for the most variance in the CSF or WM signal timecourse, respectively, for each functional run and participant. The ten extracted principal components were added to each run of the first-level general linear models (GLMs) as regressors of no interest. The fMRI data of Limanowski et al. [2014; acquired using the same scanner, head coil, and sequence, $2.0 \times 2.0 \times 2.0$ mm^3 at an acquisition time of 2240 ms per image volume] were preprocessed in the same way as described above for the present experiment.

On the first level, we fitted a GLM to each participant's fMRI data (microtime onset set to the middle slice, 300 s high-pass filter); each condition was modeled as a regressor with a boxcar function and convoluted with the standard hemodynamic response function of SPM. Contrast images of the regressors modeling each condition versus baseline were entered into flexible factorial GLMs on the second level (random effects, group analysis), including a factor modeling the subject constants. Thus we first set up a three-by-two factorial GLM in which we manipulated the stimulation location *(Palm, Forearm, Both)* and the congruence *(Cong, Incong)* of visuotactile stimulation (see Fig. 1C). Although we did not necessarily expect the same voxels to be activated by the different stimulation contrasts, brain regions that are universally involved in multisensory integration should still consistently respond to congruent versus incongruent stimulation. Therefore, we used the global conjunction analyses [Friston et al., 2005] on the respective congruent versus incongruent contrasts of each location and experiment to identify voxels that showed consistent effects across locations and type of comparison. As a complementary analysis, we tested for brain areas in which activity differences between the congruent and incongruent stimulations would additionally change over time, i.e., during the blocks of stimulation. This analysis was motivated by our assumptions based on predictive coding, whereby prediction error should be generated dynamically throughout the illusion. To test for such effects, we added a parametric modulation to each condition on the first-level. The parametric modulator was modeled as linearly increasing, and centered on the scan within the stimulation block in which the onset of the illusion was reported by the participant (obtained from the verbal illusion onset ratings, see above). The resulting contrast images were entered into a group-level design analogous to that of the main GLM analysis. We looked for stronger parametric activity increases over stimulation blocks (i.e., significantly higher beta values for the parametric regression slope) during the congruent versus incongruent conditions. We performed the analogous conjunction analysis of the three respective congruent versus incongruent contrasts for the palm, forearm, and both locations. Finally, to compare our data to those of Limanowski et al. [2014, see above], we set up a group-level GLM comprising a group factor and an experimental factor featuring four stimulation types: one congruent and one incongruent condition per stimulation location (the palm or the forearm). Thereby, to account for the differences between experiments in our design, all relevant contrasts were calculated on the interaction terms of the experimental factor with the group factor, i.e., on the following eight regressors: $Palm_{TEMP\_CONG}$, $Palm_{TEMP\_INCONG}$, $Forearm_{TEMP\_CONG}$, $Forearm_{TEMP\_INCONG}$ (present experiment, temporal (in)congruence), and $Palm_{SPATIAL\_CONG}$, $Palm_{SPATIAL\_INCONG}$, $Forearm_{SPATIAL\_CONG}$, $Forearm_{SPATIAL\_INCONG}$ [Limanowski et al., 2014, spatial (in)congruence]. For the analysis of correlations between brain activity and

between-participant differences in the reported intensity of the ownership illusion, we calculated the relevant congruent versus incongruent contrasts on the first-level, and entered the resulting contrast images into a group-level one-sample t-test (for the main experiment) or two-sample t-test [for the comparison of the current dataset with the dataset of Limanowski et al., 2014] including the behavioral illusion scores as a covariate (following Ehrsson et al., 2004). For activations obtained from this analysis, we also report the correlations (Pearson's r, using two-tailed significance) between the parameter estimate at the respective peak voxel, and the behavioral illusion scores.

Statistic images were assessed for significant activations in the whole brain using an initial voxel-wise threshold of $P < 0.001$, and a cluster-wise threshold of $P < 0.05$, family-wise error (FWE) corrected, was used to correct for false positives due to multiple comparisons. Based on extensive work done by others [Ehrsson et al., 2004; Petkova et al., 2011], we had specific a priori hypotheses about the involvement of the left PMv and the left IPS in the RHI; we hence corrected for multiple comparisons in these regions of interest (ROIs) using peak-FWE small volume correction within 10 mm radius spheres centered on the peak coordinates reported in the most recent fMRI study using a classical RHI paradigm [Petkova et al., 2011; left PMv: $x = -48$, $y = 6$, $z = 32$; left IPS: $x = -46$, $y = -48$, $z = 56$]. Likewise, we used the activations in the EBA as obtained from the joint visual body-selectivity localizer, and the activation in left primary somatosensory cortex (SI) and left secondary somatosensory cortex (SII) as obtained from the tactile localizer (touch to the real arm only), all thresholded at $P < 0.001$, uncorrected, as ROI mask images for small volume correction in the RHI experiments. We only report activations that survived either cluster-wise FWE-correction or small volume correction based on the pre-defined ROIs, unless explicitly stating so (in a few cases where activity in regions of interest did not survive correction, we explicitly report the statistics and uncorrected P-value). For a better interpretation of the results of the conjunction analyses, we report the corresponding peak T and Z values. For visualization of the results, we projected the resulting statistical parametric maps (thresholded at $P < 0.001$, uncorrected) onto the mean normalized structural image. All reported coordinates are in MNI space, and neuroanatomical labels are derived from the SPM Anatomy toolbox [Eickhoff et al., 2005] where possible. Activations that fell within the corresponding masks of EBA or FBA obtained from the visual body-selectivity localizer (which corresponded to published literature, see Results) were labeled accordingly.

Connectivity Analysis

We analyzed the brain connectivity (i.e., changes in the statistical dependencies of BOLD signal time-series under the illusion context) of the key cortical regions revealed by our standard GLM analysis by means of psycho-physiological interactions [PPIs, Friston et al., 1997], and the effective connectivity within the corresponding network (i.e., how influences of certain nodes onto others change under the illusion context) by means of dynamic causal modeling [DCM, Friston et al., 2003]. Both analyses (PPI and DCM) were performed based on the same seed regions, and separately for each stimulation location (palm, forearm, or both together) and type of congruence versus incongruence (spatial and temporal). The nodes of our model were chosen based on the results of our univariate analysis and their correspondence with previous literature on illusory body ownership: Our group-level GLM analysis identified three main sources of cortical activity related to congruent versus incongruent stimulation of the dummy arm and the real arm, whose involvement in the RHI is well documented: LOC/EBA, IPS, and PMv [e.g., Brozzoli, Gentile et al., 2012; Ehrsson et al., 2004; Gentile et al., 2013; Petkova et al., 2011]. Notably, our GLM analysis revealed a much stronger and more consistent response to spatiotemporal congruence of touches to a right arm within the respective areas of the left hemisphere (see Results). This result is consistent with previous evidence for predominantly left-hemispheric activation of the IPS and PMv to tactile stimulation of the right hand across the body midline [Lloyd et al., 2002], as well as with the implied general importance of the left IPS and PMv during the RHI [Ehrsson et al., 2004; Gentile et al., 2013; Guterstam et al., 2013; Petkova et al., 2011]; there is also evidence for a left-lateralized response to vision of hands in the IPS and LOC/EBA [Astafiev et al., 2004; Bracci et al., 2010, 2012; Zopf and Williams, 2013]. We therefore based the connectivity analyses on a left-hemispheric network comprising seed regions in the LOC, IPS, and PMv, and, since the RHI depends on visual and somatosensory information integration, we further included the left SII as the input area for somatosensory information from the contralateral hand in the DCM analysis. We did not include the cerebellum (our standard GLM analysis revealed some activity differences in the cerebellum) in the network, as we were interested in the interaction of secondary visual and somatosensory cortices with multisensory integrative areas in the IPS [following models of the RHI by Makin et al., 2008; Tsakiris, 2010] and the PMv. For each participant, the experimental runs, with their extracted physiological noise regressors, were concatenated into a single data sequence [Friston, 2004]. Region-specific BOLD time-series were extracted as the first eigenvariate of all significant voxels within a 6 mm radius sphere centered on each participant's local maxima in left SII, LOC, IPS, and PMv, as obtained from the respective T-contrast congruent versus incongruent stimulation (congruent plus incongruent stimulation versus baseline for SII); mean MNI coordinates and standard deviations of the centers of these VOIs were: SII ($x = -55.7 \pm 7.2$, $y = -24.3 \pm 5.4$, $z = 16.9 \pm 2.8$), LOC ($x = -45.9 \pm 5.3$, $y = -70.5 \pm 5.3$, $z = -1.5 \pm 4.6$), IPS ($x = -44.3 \pm 5.1$, $y = -45.3 \pm 6.5$, $z = 53.8 \pm 4.9$), PMv ($x = -45.9 \pm 5.5$, $y = 9.7 \pm 6.2$, $z = 31.3 \pm 5.7$).

First, we used PPI to examine changes in connectivity of each seed region under the RHI context (congruent versus

incongruent stimulation of the real and dummy arm). Taking the extracted BOLD time-series, we calculated the psycho-physiological interaction term with the experimental vector "congruent versus incongruent stimulation," and then performed a second GLM analysis including the seed region's time-series, the experimental vector, and the interaction term. For each seed region, the first-level contrast images of each stimulation location's PPI (palm, forearm, or both) were entered into a group-level within-subject GLM as three levels of one factor. Significant voxels were selected based on a global conjunction analysis of the three group-level T-contrasts, following the same logic as in our standard GLM analysis.

Next, we examined the mutual influences within this brain network involved in the RHI using DCM. The idea behind DCM is to construct a plausible model of interacting brain regions, whose parameters can be estimated from the data [Friston et al., 2003]. Thereby the connectivity targeted by DCM is the coupling of neuronal states among certain brain regions; more specifically, one typically tries to model how the influence of activity in one brain area on activity in other brain areas changes under a certain experimental factor. In DCM one distinguishes between the endogenous connectivity (the "architecture" of the model, i.e., the latent coupling of responses among brain areas that is independent of experimental manipulations, encoded in the DCM.A matrix) and the effective connectivity (changes in connectivity among brain areas due to experimental factors, DCM.B matrix). Experimental variables can affect the model in two ways, by directly changing activity in certain nodes (driving input), or by changing the coupling among two nodes (modulatory input). In a typical DCM analysis, one first constructs different plausible models varying in their connectivity among the nodes of the network, and then inverts these models (fits them to the data), to finally compare their evidence given the particular set of data using Bayesian inference. Classical statistical inference can then be performed on the different parameters of the "winning" model. Our DCM design matrices comprised a regressor modeling the sensory input (congruent and incongruent stimulation), and a regressor modeling the contextual effect of the RHI (congruent stimulation). After estimation, the individual models were compared using random-effects Bayesian model selection (RFX BMS) to determine the model with the overall highest evidence [Stephan et al., 2009]; we report each model's exceedance probability, which reflects how likely a model is compared with the competing models. Multisensory (visuo-tactile) stimulation was modeled with a boxcar function and defined as the driving input entering left SII and left LOC. We modeled all connections bidirectionally, and included self-connections on all nodes. In a first step, we tested various endogenous connectivity patterns (i.e., independent of experimental context) against each other, assuming connections between SII-IPS, LOC-IPS, and IPS-PMv, and evaluating all other possible connectivity patterns among SII, LOC, and PMv. This assumption was based upon previous investigations of multisensory integration and crossmodal effects,

where information transmission occurs between "lower-level" sensory areas (SII and LOC) and "higher-level" multisensory convergence zones in the PPC/IPS [Beauchamp et al., 2007, 2010; Macaluso and Driver, 2001]; such a hierarchy—with a central role of the IPS—is also implied by the RHI literature [Blanke, 2012; Brozzoli, Gentile et al., 2012]. Moreover, many studies investigating the RHI have put emphasis on a fronto-parietal brain network comprising the PMv and the IPS [e.g., Gentile et al., 2013], whereby some have argued that the PMv may have more complex functions [Ehrsson et al., 2004]. The winning endogenous connectivity model served as the basis for our comparison of models with differential effective connectivity, as follows: We tested whether the RHI context would affect feedforward and/or feedback connections across the network's hierarchy in a model space motivated by previous results on the RHI (see above). We defined a "bottom-up" model in which the RHI was allowed to modulate the connections from the SII and the LOC to the IPS, and a "top-down" model in which the RHI condition was allowed to modulate the respective reverse connections. Furthermore, for each model, modulations of the connectivity between the PMv and the IPS by the RHI were allowed (a) not at all, (b) from the IPS to the PMv (for the bottom-up model) or vice versa (for the top-down model), or (c) bidirectionally. We created two additional models: One in which the connections between all nodes were modulated by the RHI, and one in which bidirectional IPS–LOC/IPS–SII connections were modulated. Together with the null model (no modulatory effect of the RHI allowed) this resulted in a model space of nine models for each experiment and stimulation location (see Fig. 6A). Following RFX BMS to identify the most likely model given our data, we extracted all participants' parameter estimates of the winning model, and assessed them for significance using two-tailed t-tests and Bonferroni-adjusted alpha levels to account for the number of comparisons in each case.

RESULTS

Illusory Self-Attribution of the Dummy Arm During Congruent Versus Incongruent Visuotactile Stimulation

In this study, we stimulated the palm, forearm, or both locations on the participant's real hidden arm together with the corresponding location of a realistic ipsilateral dummy arm; To validate the induction of illusory self-attribution of the dummy arm by congruent (synchronous), but not incongruent (asynchronous) visuo-tactile stimulation of the two arms, we analyzed the behavioral self-attribution ratings obtained from our participants (using nonparametric tests since the ratings did not pass the Kolmogorov–Smirnov test for normality). As expected, the congruent versus incongruent visuo-tactile stimulation conditions produced a strong illusory self-attribution of the dummy arm, at the palm location (mean rating $\pm$ SD: congruent $= 2.00 \pm 0.73$, incongruent $= -0.85 \pm 1.79$, Wilcoxon signed-ranks, $n = 20$, $Z = 3.64$, $P = 0.00027$), the

TABLE I. Consistent effects of congruent versus incongruent visuo-tactile stimulation across touch locations, obtained from the conjunction of the contrasts Palm$_{CONG}$ vs Palm$_{INCONG}$, Forearm$_{CONG}$ vs Forearm$_{INCONG}$, and Both$_{CONG}$ vs Both$_{INCONG}$

	Peak MNI					
Anatomical region	x	y	z	Peak t	Peak z	P (corrected)
L middle occipital gyrus (LOC/EBA)	−40	−70	−2	2.59	5.11	< 0.001[a,b]
R cerebellum	24	−64	−36	2.26	4.60	< 0.001[a]
L middle frontal gyrus	−20	18	40	2.09	4.33	0.017[a]
L precentral sulcus (PMv)	−38	10	28	2.09	4.32	0.022[a,c]
L intraparietal sulcus	−42	−50	58	1.59	3.56	0.0497[c]

Significance at $P < 0.05$ (FWE-corrected) based on
[a]Cluster-wise correction.
[b]Small volume correction using ROIs from the visual body-selectivity localizer.
[c]Small volume correction using pre-defined ROIs based on published literature.

forearm location (mean rating ± SD: congruent = 2.10 ± 0.72, incongruent = −0.25 ± 1.86, Wilcoxon signed-ranks, $n = 20$, $Z = 3.54$, $P = 0.00039$), and both locations (mean rating ± SD: congruent = 1.85 ± 0.93, incongruent = −0.20 ± 1.51, Wilcoxon signed-ranks, $n = 20$, $Z = 3.60$, $P = 0.00032$), see Figure 1D.

Brain Activity in Fronto-Parietal and Occipito-Temporal Areas Increases during Congruent Versus Incongruent Visuotactile Stimulation Across Anatomical Locations

We expected that congruent visuotactile touch information (i.e., synchronous touch on the real arm and the corresponding location on the dummy arm) would selectively engage fronto-parietal and occipito temporal brain areas [Ehrsson et al., 2004; Gentile et al., 2013; Limanowski et al., 2014; Petkova et al., 2011]. The main effect of congruent versus incongruent visuo-tactile stimulation of the two arms revealed significantly ($P < 0.05$, corrected for multiple comparisons) increased BOLD responses during congruent versus incongruent stimulation in the left LOC ($x = −40$, $y = −60$, $z = −2$, $T = 4.64$), the cerebellum ($x = −2$, $y = −60$, $z = −34$, $T = 4.39$), the left PMv ($x = −40$, $y = 12$, $z = 28$, $T = 4.00$), and a strong trend in the left IPS ($x = −40$, $y = −54$, $z = 58$, $T = 3.22$, $P = 0.078$). We expected that the responses to congruent versus incongruent visuo tactile stimulation in these brain areas would generalize across the different stimulation locations, and therefore calculated a conjunction analysis of the congruent versus incongruent stimulation contrasts for each location: Palm$_{CONG}$ vs Palm$_{INCONG}$, Forearm$_{CONG}$ vs Forearm$_{INCONG}$, and Both$_{CONG}$ vs Both$_{INCONG}$. This analysis revealed significantly ($P < 0.05$, corrected, see Fig. 2 and Table I) higher activity during visuo tactile congruence versus incongruence in the left LOC, left PMv, left IPS, and the cerebellum across the three comparisons, thus supporting the observed main effect. Further activations

obtained from this conjunction (thresholded at $P < 0.001$, uncorrected) that corresponded to regions previously reported in RHI experiments were observed in the bilateral anterior insulae (L: $x = −34$, $y = 18$, $z = −2$, $T = 1.42$, $Z = 3.30$; R: $x = 32$, $y = 22$, $z = −2$, $T = 1.43$, $Z = 3.31$). Notably, the activations we observed within the left LOC, the left IPS, and the left PMv were also contained within the activations obtained from the visual body-selectivity localizer (see Fig. 4 and Table II), as revealed by masking the results of the conjunction analysis with the visual body-selectivity contrast (mask image thresholded at $P < 0.001$).

We also sought to identify brain areas whose activity would reflect individual differences in the reported strength and prevalence of the experienced illusory self-attribution, which we quantified via compound illusion scores reflecting both the rating difference between congruent and incongruent conditions, and the prevalence of the ownership illusion in the congruent condition [see Materials and methods and Ehrsson et al., 2004]. This analysis yielded voxels within the left ($x = −26$, $y = −86$, $z = 30$, $T = 3.71$, $P < 0.001$, uncorrected) and right ($x = 36$, $y = −72$, $z = 32$, $T = 4.10$, $P < 0.05$, corrected) posterior IPS (pIPS), and, at a more liberal threshold also in the left ($x = −46$, $y = −84$, $z = 4$, $T = 2.94$, $P = 0.002$, uncorrected) and right LOC/EBA ($x = 56$, $y = −62$, $z = −10$, $T = 2.93$, $P = 0.002$, uncorrected), both contained within the EBA as localized independently, and both showing significant positive correlations with the behavioral illusion scores (Pearson's $r = 0.36$, $P < 0.01$, for each location). An additional regression analysis looking for activity correlated with only the ownership ratings of each condition revealed a cluster in the left EBA ($x = −42$, $y = −68$, $z = −6$, $T = 3.38$, $P = 0.095$, corrected; Pearson's $r = .30$, $P < 0.001$).

Finally, we tested for brain areas in which activity in the congruent relative to the incongruent stimulation conditions would be differently modulated over time (i.e., during the stimulation blocks). We compared individually centered, linearly increasing parametric modulations of the
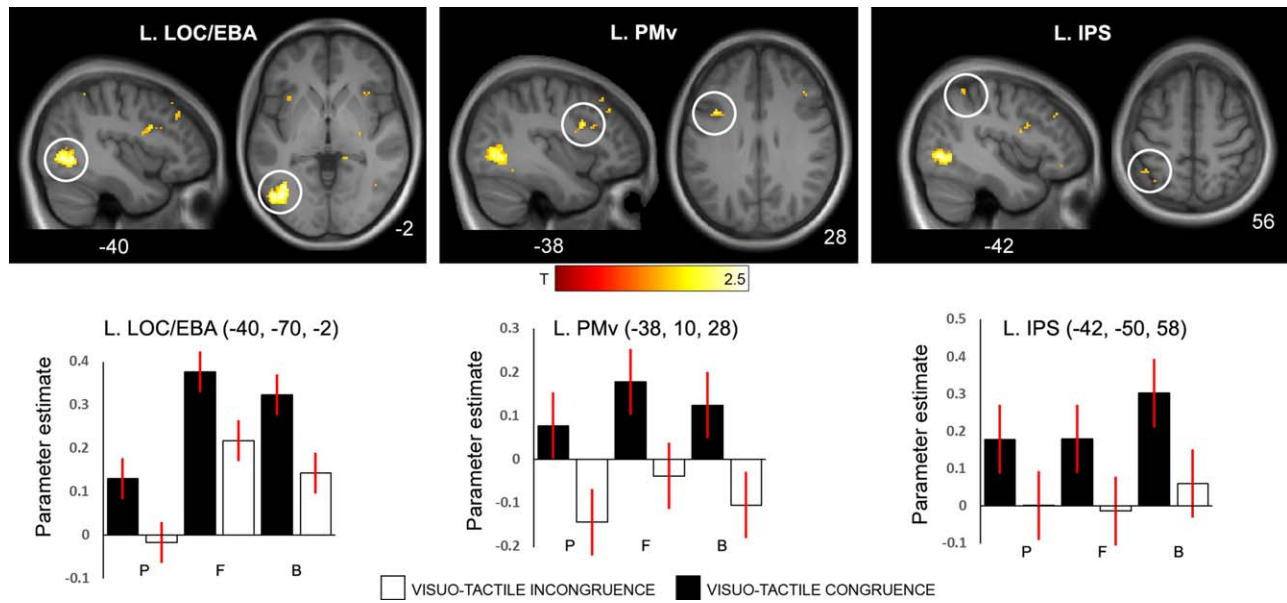
Figure 2.

Brain areas showing consistently stronger responses to congruent versus incongruent visuo-tactile stimulation across the different stimulation locations on the real arm and the dummy arm. A conjunction analysis of the differential contrasts $Palm_{CONG}$ vs $Palm_{INCONG}$, $Forearm_{CONG}$ vs $Forearm_{INCONG}$, and $Both_{CONG}$ vs $Both_{INCONG}$ revealed significant ($P < 0.05$, corrected) voxels within the left lateral occipitotemporal cortex (LOC, largely overlapping with the extrastriate body area, EBA), the left ven-

tral premotor cortex (PMv), and the left intraparietal sulcus (IPS). The SPM{T} maps are displayed at a threshold of $P < 0.001$, uncorrected, and superimposed onto the mean normalized structural image. Bar plots depict the parameter estimates and associated standard errors at the given MNI coordinates. P: Palm locations, F: Forearm location, B: Both locations. See Table I for details.

same regressors as in the main GLM. The main effect of congruent versus incongruent stimulation revealed significantly ($P < 0.05$, corrected) stronger modulations of activity in the left SII (L: $x = -62$, $y = -18$, $z = 24$, $T = 4.44$), the left LOC ($x = -42$, $y = -68$, $z = 6$, $T = 5.15$), and in the bilateral PMv (L: $x = -52$, $y = 2$, $z = 40$, $T = 3.33$, $P = 0.055$; R: $x = 48$, $y = 2$, $z = 38$, $T = 3.75$); these activations all reached corrected significance in a conjunction across stimulation locations.

Generalizable Effects of Spatial and Temporal Congruence Versus Incongruence of Visuotactile Stimulation

As part of our analysis, we aimed to show that the effects of the main analysis (increased BOLD activations during temporal congruence versus incongruence of visuotactile stimulation) would also generalize to spatial congruence versus incongruence of visuotactile stimulation. Therefore, we compared the data obtained in the current experiment to a previously acquired dataset featuring spatially (in)congruent visuotactile stimulation (see Materials and methods). We observed a significant ($P < 0.05$, corrected) main effect of congruent versus incongruent stimulation in the left LOC

($x = -44$, $y = -72$, $z = 2$, $T = 4.72$), the cerebellum ($x = -10$, $y = -62$, $z = -40$, $T = 4.76$), the left IPS ($x = -42$, $y = -46$, $z = 56$, $T = 3.94$) and superior parietal lobule ($x = -24$, $y = -50$, $z = 50$, $T = 4.90$), as well as statistical trends in the left PMv ($x = -44$, $y = 14$, $z = 34$, $T = 3.17$, $P = 0.083$, corrected) and the left IPS ($x = -40$, $y = -48$, $z = 58$, $T = 3.14$, $P = 0.052$, corrected, at a more liberal voxelwise threshold of $P < 0.005$). To test for the consistency of these results, we calculated a conjunction across the contrasts of the different experiments reflecting temporal or spatial (in)congruence at the palm or the forearm location, i.e., $Palm_{TEMP\_CONG}$ vs $Palm_{TEMP\_INCONG}$, $Forearm_{TEMP\_CONG}$ vs $Forearm_{TEMP\_INCONG}$, $Palm_{SPATIAL\_CONG}$ vs $Palm_{SPATIAL\_INCONG}$, and $Forearm_{SPATIAL\_CONG}$ vs $Forearm_{SPATIAL\_INCONG}$. This conjunction (Fig. 3 and Table II) revealed voxels showing consistently stronger activity ($P < 0.05$, corrected) to temporally *and* spatially congruent versus incongruent visuo-tactile stimulation across the locations in the left LOC/EBA and the left IPS and inferior parietal lobule (the activations in the LOC and the IPS thereby again fell within the corresponding body-selective clusters, see Table II), further in the bilateral fusiform gyri and the cerebellum. These results strongly support the involvement of the visual body-selective left LOC/EBA and IPS in the illusory self-attribution of the dummy arm.

TABLE II. Brain regions showing consistently stronger responses to both temporal and spatial congruence versus incongruence of visuo-tactile stimulation, obtained from the conjunction of the contrasts $Palm_{TEMP\_CONG}$ vs $Palm_{TEMP\_INCONG}$, $Forearm_{TEMP\_CONG}$ vs $Forearm_{TEMP\_INCONG}$, $Palm_{SPATIAL\_CONG}$ vs $Palm_{SPATIAL\_INCONG}$, and $Forearm_{SPATIAL\_CONG}$ vs $Forearm_{SPATIAL\_INCONG}$

Anatomical region	Peak MNI			Peak t	Peak z	P (corrected)
	x	y	z			
L middle occipital gyrus (EBA)	−42	−70	2	1.83	4.67	<0.001[a,b]
L cerebellum	−10	−62	−42	1.81	4.63	<0.001[a]
R fusiform gyrus (FBA)	42	−48	−20	1.71	4.45	0.007[b]
L inferior parietal lobule	−26	−44	48	1.61	4.28	0.002[a]
L fusiform gyrus (FBA)	−36	−48	−20	1.48	4.06	0.032[b]
L intraparietal sulcus	−42	−48	58	1.34	3.81	0.021[b,c]

Significance at $P < 0.05$ (FWE-corrected) based on
[a]Cluster-wise correction.
[b]Small volume correction using ROIs from the visual body-selectivity localizer.
[c]Small volume correction using pre-defined ROIs based on published literature.

Next, we compared the behavioral effects of the two experiments. There were no significant differences between the reported ownership ratings or the respective rating differences reported for temporal versus spatial (in)congruence (Wilcoxon signed-ranks, $n = 20$, all $Zs < 1.71$., all $ps > 0.2$): Participants reported significantly higher self-attribution of the dummy arm following temporally congruent (mean ± SD = 2.30 ± 0.66) versus incongruent (mean ± SD = 0.25 ± 1.29) and spatially congruent (mean ± SD = 1.98 ± 0.70) versus incongruent (mean ± SD = −0.43 ± 1.43) stimulation (Wilcoxon signed-ranks, $n = 20$, $Z = 3.83$, $P = 0.0001$) stimulation (Wilcoxon signed-ranks, $n = 20$, $Z = 3.99$, $P = 0.00006$). Hence temporal and spatial congruence versus incongruence of visuo-tactile stimulation of the dummy arm and the real arm both successfully induced an illusory self-attribution of the dummy arm. A conjunction analysis across the participants' behavioral illusion scores as separate covariates in a two-sample t-test revealed significant ($P < 0.05$, corrected) voxels within the left LOC/EBA ($x = −40$, $y = −74$, $z = −2$, $T = 2.81$, $Z = 4.27$), whose activity was consistently positively correlated with the behavioral illusion scores (Fig. 3D). Further significant correlations were observed in voxels in the left PPC ($x = −18$, $y = −66$, $z = 64$, $T = 3.21$, $Z = 4.76$), and in the left ($x = −14$, $y = −84$, $z = 40$, $T = 2.88$, $Z = 4.13$) and right ($x = 18$, $y = −78$, $z = 48$, $T = 2.84$, $Z = 4.30$) PPC, both spanning to the posterior IPS. Similar correlations were also observed in voxels in the left Thalamus ($x = −12$, $y = −16$, $z = 0$, $T = 2.25$, $Z = 3.56$), the right SMG ($x = 60$, $y = −18$, $z = 30$, $T = 2.33$, $Z = 3.66$), and the right LOC/EBA ($x = 50$, $y = −72$, $z = 8$, $T = 2.04$, $Z = 3.29$), but these activations did not survive correction for multiple comparisons. Correspondingly, an analogous regression analysis on the ownership ratings revealed a cluster in the left EBA ($x = −58$, $y = −66$, $z = 6$, $T = 2.16$, $Z = 3.47$, $P = 0.089$, corrected), where activity was significantly positively correlated with the ownership ratings across experiments (Pearson's $r = .29$, $P < 0.01$ and $r = 0.23$, $P < 0.05$); further ($P < 0.001$, uncorrected) activations were observed in the right LOC/EBA ($x = 54$, $y = −60$, $z = −12$, $T = 1.89$, $Z = 3.11$) and the left pIPS ($x = −22$, $y = −74$, $z = 26$, $T = 2.44$, $Z = 3.84$). These results imply that the left EBA (and IPS) reflected individual differences in the experienced intensity of the illusory self-attribution of the dummy arm across both touch locations and types of (in)congruence.

Brain Regions Showing a Preferential Response to Vision of Human Body Parts

We hypothesized that brain areas involved in the visual processing of the body play an important role in the self-attribution of body parts. To identify such brain areas, we implemented an independent functional localizer run, in which we presented our participants pictures of human hands and feet versus motorcycle parts [following Urgesi et al., 2007]. The contrast *Body parts vs Objects* (Fig. 4 and Table III) revealed the strongest activity ($P < 0.05$, corrected for multiple comparisons) in bilateral LOC, matching coordinates previously reported for the body-selective extrastriate body area [EBA; e.g., Astafiev et al., 2004; Costantini et al., 2011; Downing et al., 2001]. These large clusters each spanned to more inferior parts of the temporal and fusiform gyri, thus including locations reported for the body-selective fusiform body area [FBA; e.g., Schwarzlose et al., 2005]. Interestingly, significant activations were also observable in frontal and parietal areas, namely, bilaterally in the SMG, anterior parts of the IPS and superior parietal cortex, the PMv, and in the right inferior frontal gyrus.
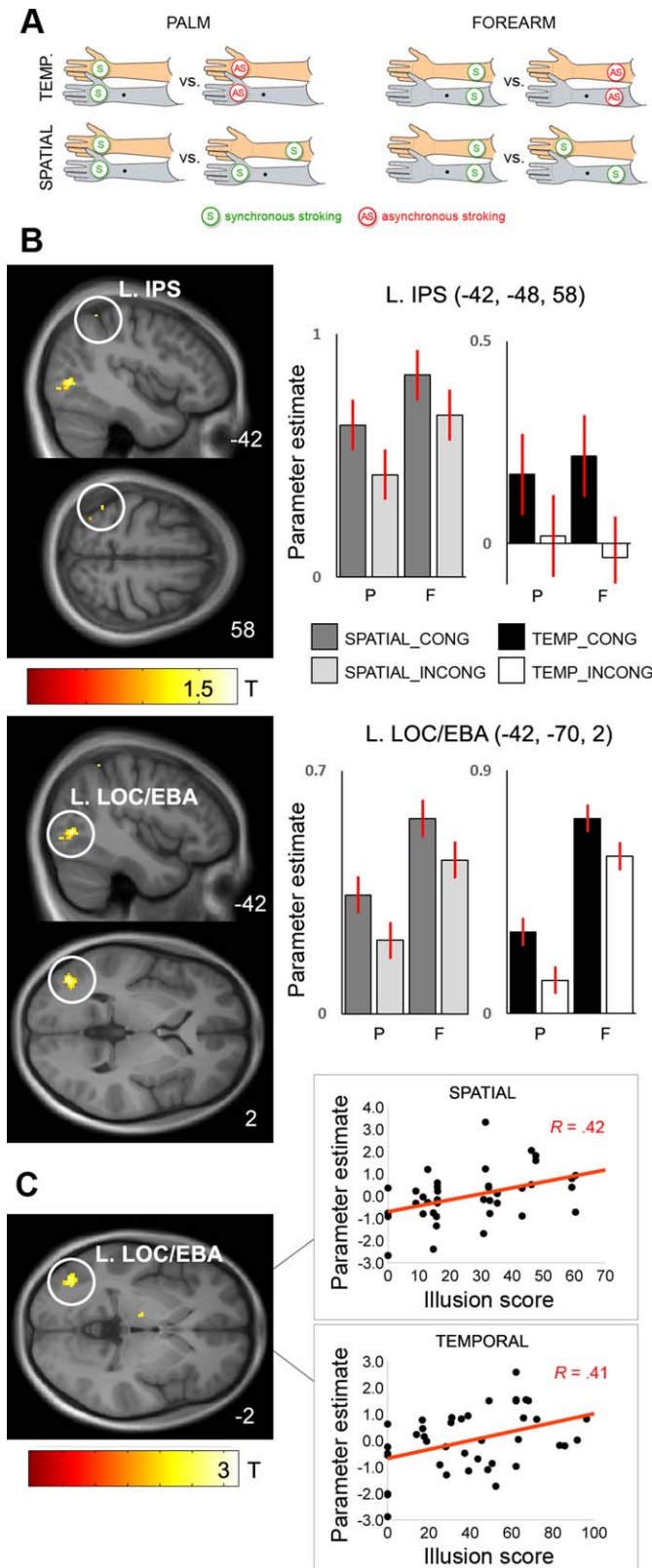
Figure 3.

Consistent Increases in Functional Coupling during Congruent Versus Incongruent Visuotactile Stimulation of each Anatomical Location

In the next step, we sought to illuminate the illusion-related connectivity changes in the network implied by our standard GLM analyses. To this end, we first examined changes in functional coupling of the key nodes of this network using psycho-physiological interaction analyses (PPIs) with seed regions in the left LOC, left IPS, and left PMv. To test for the consistency of these changes, we calculated a conjunction analysis of the congruent versus incongruent stimulation contrasts for each location, that is: $Palm_{CONG}$ vs $Palm_{INCONG}$, $Forearm_{CONG}$ vs $Forearm_{INCONG}$, and $Both_{CONG}$ vs $Both_{INCONG}$. This (see Fig. 5A and Table IV) revealed consistent, significantly ($P < 0.05$, corrected) increased functional coupling during congruent versus incongruent stimulation of the left IPS with regions in the bilateral LOC/EBA, the left PMv, and the left SI. The left PMv and the left LOC each showed similar significant ($P < 0.05$, corrected) increases in functional coupling with the bilateral LOC/EBA, and the left SI and SII during congruent versus incongruent touch across all locations. We also compared the connectivity of each seed region across stimulation locations and type of (in)congruence [temporal, present experiment, and spatial, Limanowski et al., 2014] by calculating a group-level conjunction analysis of the PPIs corresponding to the contrasts $Palm_{TEMP\_CONG}$ vs $Palm_{TEMP\_INCONG}$, $Forearm_{TEMP\_CONG}$ vs $Forearm_{TEMP\_IN-CONG}$, $Palm_{SPATIAL\_CONG}$ vs $Palm_{SPATIAL\_INCONG}$, and $Forearm_{SPATIAL\_CONG}$ vs $Forearm_{SPATIAL\_INCONG}$, for each seed region. The results of this analysis (Fig. 5B, see Table IV for details) replicated the connectivity pattern observed for the main experiment: The left IPS also showed consistent,

The effects of congruent versus incongruent visuo-tactile stimulation generalized across touch locations for spatial and temporal (in)congruence. A: Schematic depiction of the four differential contrasts examined in the conjunction analysis: $Palm_{TEMP\_CONG}$ vs $Palm_{TEMP\_INCONG}$, $Palm_{SPATIAL\_CONG}$ vs $Palm_{SPATIAL\_INCONG}$, $Forearm_{TEMP\_CONG}$ vs $Forearm_{TEMP\_INCONG}$, and $Forearm_{SPATIAL\_CONG}$ vs $Forearm_{SPATIAL\_IN-CONG}$. B: SPM{T} maps of the significant activations obtained from the conjunction of all four contrasts located in the left IPS and the left LOC/EBA ($P < 0.05$, corrected, displayed at a threshold of $P < 0.001$, uncorrected, and superimposed onto the mean normalized structural image). Bar plots depict the parameter estimates and associated standard errors at the given MNI coordinates for each stimulation type and location (P: Palm, F: Forearm). See Table II for details. C: A conjunction analysis of the effects of the illusion scores as separate covariates for each type of (in)congruence revealed significant positive correlations with the reported illusory self-attribution within left LOC/EBA ($P < 0.05$, corrected). The plots show significant correlations of left LOC/EBA response differences between congruent and incongruent stimulation, and the respective illusion scores (Pearson's $r = .42$, $P < 0.01$, and $r = .41$, $P < 0.01$).
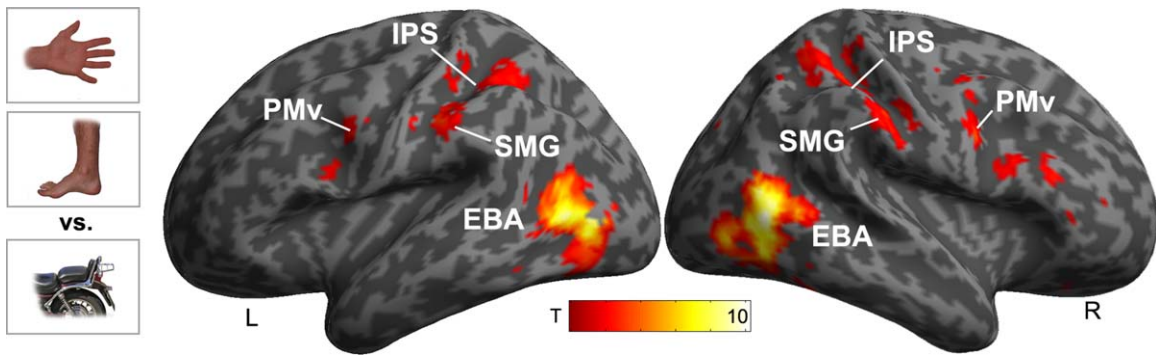
Figure 4.

Significant activations obtained from the joint visual body-selectivity localizer (contrast *Body parts vs Objects*, sample stimuli shown left). Voxels that showed a strong preferential response to vision of human body parts versus objects were located bilaterally in the extrastriate visual cortex, matching reported locations of the extrastriate body area (EBA). Further significant activations were located bilaterally in anterior parts of the intraparietal sulcus (IPS), the supramarginal gyrus (SMG), and the ventral premotor cortex (PMv). The group-level surface render is displayed at a threshold of $P < 0.001$, uncorrected; labels mark signficant activations ($P < 0.05$, corrected for multiple comparisons), see Table III for details. L/R: left/right hemisphere.

significantly increased functional coupling with the bilateral LOC/EBA, the left PMv, and the left SI and SII. The left PMv and the left LOC again showed significant ($P < 0.05$, corrected) increases in functional coupling with the bilateral LOC/EBA and the left SI and SII. Notably, in all analyses, the activations in LOC always fell within the visual body-selective EBA ($P < 0.05$, small volume corrected within mask images thresholded at $P < 0.001$).

In sum, the results of the PPI analyses suggest an overall increased connectivity between left-hemispheric fronto-parietal areas—the IPS and the PMv—with the LOC/EBA during the illusory self-attribution of a dummy arm induced by congruent visuotactile stimulation. Crucially, this connectivity pattern generalized across stimulation locations and type of visuotactile (in)congruence, thus supporting the results of our standard GLM analysis.

Dynamic Causal Modeling

Our standard GLM and PPI connectivity analyses together suggest that the self-attribution of the right dummy arm depended on interactions of left-hemispheric fronto-parietal areas with occipitotemporal areas. However, from correlation analyses like the PPI, one cannot clearly infer any directionality of the mutual influences among the nodes of a network. Therefore, we next constructed a dynamic causal model, comprising the left LOC, IPS, and PMv, and, as an input area for somatosensory information, the left SII. To ensure the generalizability of our results, we performed the same DCM analysis for each location, i.e., for the contrasts $Palm_{CONG}$ vs $Palm_{INCONG}$, $Forearm_{CONG}$ vs $Forearm_{INCONG}$, and $Both_{CONG}$ vs $Both_{INCONG}$. In a first step, we examined a model space of all possible combinations of endogenous connectivity

TABLE III. Brain regions preferentially responding to vision of human body parts: Significant activations obtained from the visual body-selectivity localizer (contrast *Body parts vs Objects*)

Anatomical region	Peak MNI			Peak t	Peak z	P (corrected)
	x	y	z			
R middle temporal gyrus (EBA)	54	−68	2	10.22	Inf.	<0.001
L middle temporal gyrus (EBA)	−50	−74	6	10.11	Inf.	<0.001
L inferior temporal gyrus (FBA)	−42	−48	−18	6.88	6.57	0.003
R precentral gyrus (PMv)	52	10	32	5.47	5.31	<0.001
R postcentral gyrus / superior parietal lobule	30	−38	52	5.08	4.95	<0.001
R supramarginal gyrus and intraparietal sulcus	50	−24	38	5.03	4.90	<0.001
L supramarginal gyrus and intraparietal sulcus	−56	−24	36	4.98	4.86	<0.001
L precentral gyrus (PMv)	−52	8	28	4.76	4.65	0.002
R inferior frontal gyrus	50	28	22	4.22	4.14	0.043

Significance at $P < 0.05$ based on cluster-wise FWE-correction.
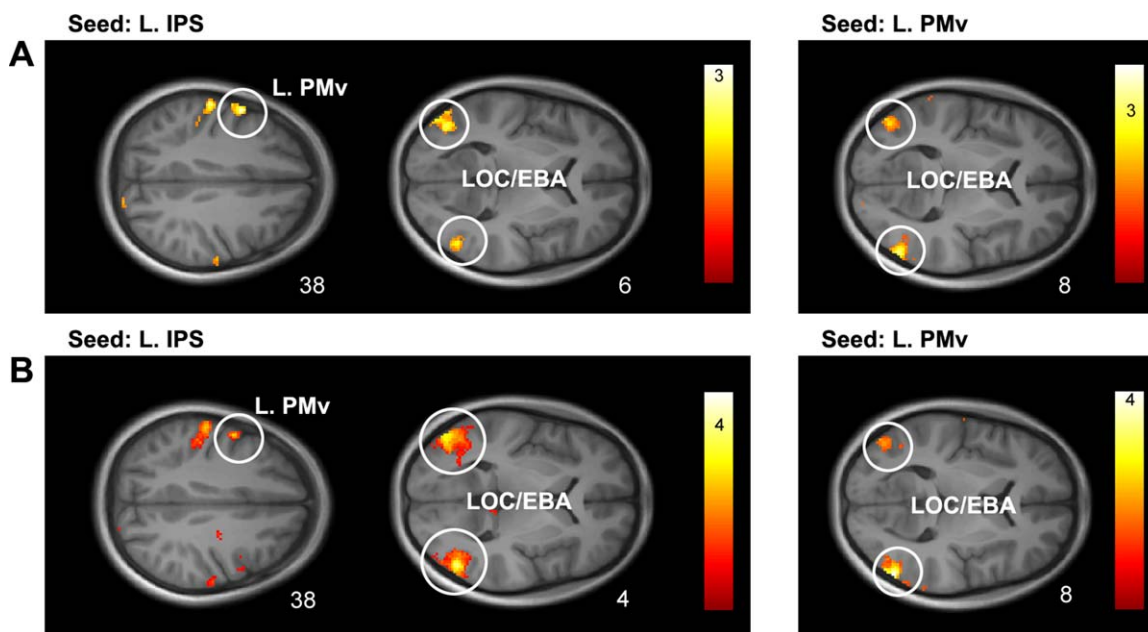
Figure 5.

Enhanced functional coupling of frontal and parietal regions with occipitotemporal regions during the rubber hand illusion. Psycho-physiological interaction analyses (PPIs) revealed consistently increased connectivity of the left IPS and the left PMv with the LOC/EBA during congruent versus incongruent visuo-tactile stimulation across stimulation locations and type of (in)congruence, as revealed by conjunction analyses across the individual PPI analyses. **A:** Results of the conjunction across the three PPIs for temporally congruent versus incongruent visuo-tactile stimulation at the palm, forearm, and both locations (main experiment). **B:** Results of the conjunction across the corresponding PPIs from each seed region during spatially and temporally congruent versus incongruent visuo-tactile stimulation at the palm and forearm location. SPM{T} maps displayed at a threshold of $P < 0.001$, uncorrected, superimposed onto the mean normalized structural image. White circles mark significant ($P < 0.05$, corrected) activations. See Table IV for details.

among these nodes (i.e., connectivity independent of the illusion context, but related to visuotactile stimulation *per se*), thereby assuming connections of the IPS to all other nodes. Random-effects Bayesian model selection (RFX BMS) yielded the same winning model in each case, featuring bidirectional connections between SII-IPS, LOC-IPS, and IPS-PMv (model excedance probabilities: 96.6% at the palm location, 99.8% at the forearm location, and 78.6% at both locations). This model therefore served as the architectural basis for the examination of RHI-evoked modulations of connectivity within this network (the resulting effective connectivity model space is described in Figure 6A, see Materials and methods for details).

In each of these comparisons, the model with the highest model exceedance probability (Model 4: 49.2% at the palm location, 64.2% at the forearm location, 44.9% at both locations) was the model featuring a modulation of the connections from the SII to the IPS and from the LOC to the IPS under congruent versus incongruent stimulation. The analysis of the parameter estimates of the winning model's DCM.B matrix revealed that in each case (palm, forearm, or both locations stimulated) the connectivity from the SII and LOC to the IPS was signifi-

cantly enhanced by congruent as compared with incongruent stimulation (two-tailed *t*-tests, all $ps < 0.025$, using Bonferroni-adjusted alpha levels to account for the number of comparisons). Additionally, we tested whether this model would also best fit our previous dataset (spatially congruent versus incongruent stimulation), with which we had compared the present data in the standard GLM analyses (see above). Therefore, we created and compared the same model space for spatially congruent versus incongruent stimulation at the palm and the forearm location. Notably, RFX BMS identified the same winning endogenous connectivity (model exceedance probabilities 82.2% and 55.0%), and more importantly, also the same winning effective connectivity (Model 4; model exceedance probabilities 52.6% at the palm location and 53.2% at the forearm location), as for the current experiment. Thus the RHI context also modulated the bottom-up connections from the SII and LOC to IPS; the DCM.B parameter estimates of these modulations reached significance for the SII to IPS connection (palm location: +0.76, $P = 0.0030$; forearm location: +0.62, $P = 0.0019$) but not for the LOC to IPS connection (palm location: −0.09, $P = 0.68$; forearm location: +0.12, $P = 0.29$). In sum, our
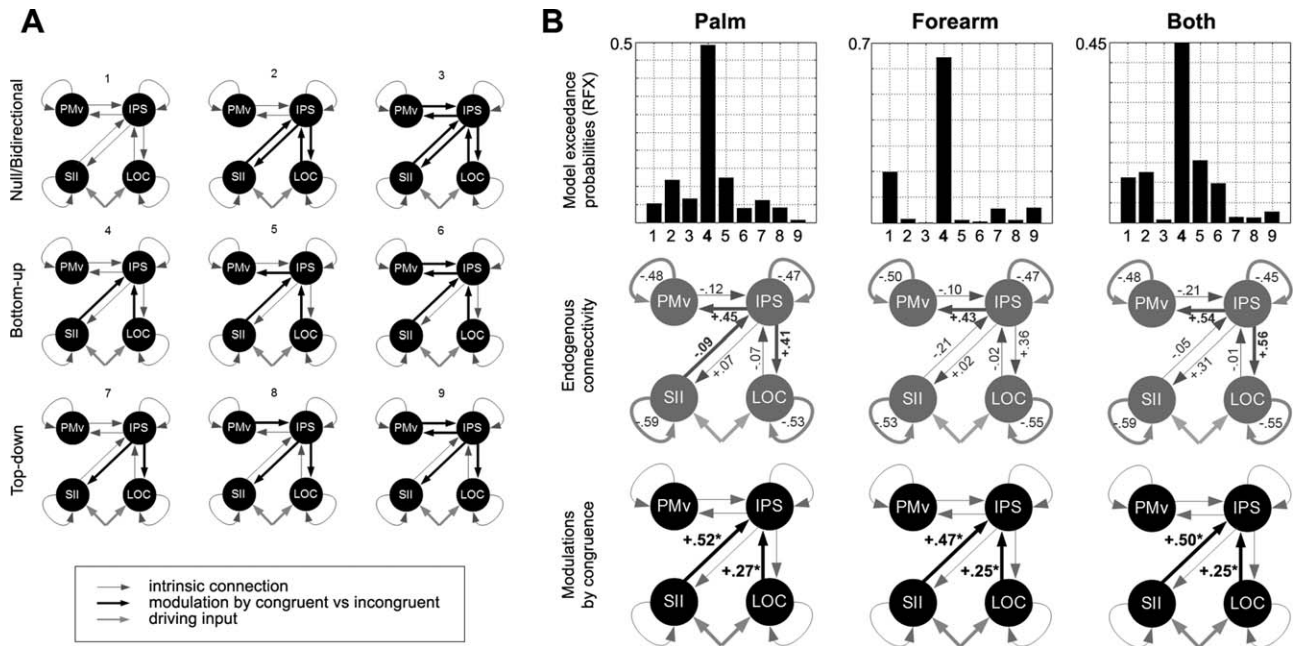
Figure 6.

Dynamic causal modeling results. **A:** The effective connectivity model space for Bayesian model selection was defined based on our GLM results and according to hypotheses about the mechanisms underlying the RHI (see Materials and methods), and consisted of nine models; top row: a "null model" without any allowed modulations (model 1, this model was identified as the most likely endogenous connectivity pattern of the nodes in a previous step), a model with bidirectional modulations between SII-IPS and LOC-IPS, and a model with additional bidirectional modulations of IPS-PMv connectivity; middle row: "bottom-up" models allowing modulations of connections from SII to IPS and LOC to IPS, and additional modulation of bottom-up or bidirectional IPS-PMv connections; bottom row: "top-down" models allowing the modulation of the respective reverse connections.

Driving input entered SII and LOC. **B:** RFX BMS identified the same winning model (Model 4) in each case (congruent vs incongruent stimulation at the palm, forearm, or both locations). Shown are the model exceedance probabilities, the endogenous connectivity of the winning model with averaged coefficients extracted from the DCM.A matrix (connectivity regardless of context), and the significant modulations of this model's connectivity from the SII to the IPS and from the LOC to the IPS, with averaged coefficients extracted from the DCM.B matrix. Bold arrows mark significant connections or modulations; the significance of all coefficients was assessed using two-tailed t-tests and Bonferroni-adjusted alpha levels ($0.05/10 = 0.005$ for the endogenous connections and $0.05/2 = 0.025$ for the modulated connections).

DCM results strongly suggest that the processing of visuo tactile stimuli in peripersonal space depends on information exchange within a hierarchical network involving the SII and LOC at lower levels, and the IPS and PMv at higher levels, whereby visuo tactile temporal and spatial congruence modulates the bottom-up connections from the SII and LOC to the IPS.

pathways from earlier sensory areas in the LOC and SII to higher-level integrative brain areas in the IPS during illusory self-attribution. These results offer important new insights into the nature of the interactions within the brain network underlying body ownership [Blanke, 2012; Ehrsson, 2012; Gallagher, 2000; Graziano and Botvinick, 2002; Makin et al., 2008; Tsakiris, 2010], as discussed in the following.

DISCUSSION

This study demonstrated the involvement of a predominantly left-lateralized network comprising PMv, IPS, and visually body-selective LOC in the self-attribution of a right dummy arm. Our analyses revealed consistent increases in the functional coupling of the PMv and the IPS with the LOC, and consistent modulations of

Processing of Congruent Visuotactile Information in Body-Selective Brain Areas Enables the Rubber Hand Illusion across Different Anatomical Locations

Congruent (synchronous) versus incongruent (asynchronous) visuo-tactile stimulation of the palm, forearm, or

both locations together on a realistic dummy arm and the real counterpart successfully induced an illusory self-attribution of the dummy arm, as demonstrated by the behavioral ratings. Conjunction analyses of the corresponding BOLD contrasts revealed consistently stronger responses to congruent versus incongruent stimulations in the left LOC, left IPS, and in the left PMv, as well as in the cerebellum. Temporal congruence of stimuli from multiple sensory modalities [as typically used in RHI experiments, e.g., Botvinick and Cohen, 1998; Ehrsson et al., 2004; Tsakiris and Haggard, 2005] is one crucial rule that the brain uses to decide whether to integrate these stimuli into a coherent percept—another such rule is the spatial co-occurrence of stimuli [Driver and Spence, 2000; Meredith and Stein, 1986]. Importantly, we replicated the involvement of the left LOC and the left IPS in the illusory self-attribution of the dummy arm by comparing the data of the current experiment with a dataset acquired using the same setup and fMRI scanner, in which the spatial, but not temporal, congruence of stimulation was manipulated [Limanowski et al., 2014]. Moreover, left LOC/EBA activity was consistently positively correlated with the behavioral illusion scores across locations and type of (in)congruence, although due to the post-hoc assessment of the behavioral ratings these correlations have to be considered with some caution. In sum, our results generalized across anatomical locations according to the spatiotemporal principles of multisensory integration [Ehrsson, 2012; Meredith and Stein, 1986].

The observed frontal, parietal, and occipitotemporal activations fit well with previous findings: The IPS and PMv are involved in processing stimuli in the PPS surrounding the upper limb [Bremmer et al., 2001; Graziano and Cooke, 2006; Grefkes and Fink, 2005; Lloyd et al., 2002; Rizzolatti et al., 1997]. Activations of the IPS and PMv by the RHI have hence been interpreted as reflecting the remapping of the PPS onto the dummy hand and the production of a coherent body ownership experience by multisensory integration [Brozzoli et al., 2012; Ehrsson et al., 2004; Gentile et al., 2013; Petkova et al., 2011]. Makin et al. [2007] reported a preferential response to visual stimuli in the space around real or dummy hands in the IPS, and notably, also in the LOC. Correspondingly, a body-selective area in the LOC has been shown to play a role during the RHI [Limanowski et al., 2014; Wold et al., 2014].

Using an independent functional localizer, we were able to show that across all comparisons, the areas within the left LOC and the left IPS (and partly also the PMv) that responded more strongly to congruent versus incongruent stimulation fell within areas that also preferentially respond to mere vision of the body: The visual presentation of human body parts versus objects produced strong bilateral BOLD responses in the LOC, corresponding to reported locations of the EBA, and frontoparietal areas including the IPS, the SMG, and the PMv. These results suggest that the visual body-selectivity of RHI-related

regions is indeed no coincidence, in line with previous speculations: The processing of visual stimuli on a hand is enhanced compared with on an object, which may be caused by the "attribution of these stimuli to the body" [Whiteley et al., 2004]. PPS processing in frontoparietal areas may also be modulated by mere vision of body parts [Graziano et al., 2000; Làdavas, 2002], and correspondingly, a visual body-selectivity of the IPS has been suggested [Konen and Haggard, 2014; Zopf and Williams, 2013]. A particularly promising candidate where modulatory effects of vision of the body on processing of visuosomatosensory body-related information could be implemented is the EBA [Costantini et al., 2011; Haggard et al., 2007], which has traditionally been considered to process unimodal visual features of the human body [Downing et al., 2001]. However, recent studies have demonstrated a modulation of EBA activity in congenitally blind people during haptic exploration of body-shaped objects (Kitada et al., 2014, 2009), even when such haptic information was transmitted via an auditory sensory-substitution algorithm [Striem-Amit and Amedi, 2014], thus arguing against a purely unimodal visual function of the EBA. Crucially, the EBA is activated during planning and producing actions with one's hand even in the absence of vision [Astafiev et al., 2004], suggesting that it also processes somatosensory (proprioceptive) information. Correspondingly, the EBA has been implied in processing action knowledge [Bracci et al., 2010], in predicting actions from visual action cues [van Elk, 2014], and in action imitation versus observation [Jackson et al., 2006]. These findings extend on previous proposals of cross-modal interactions in LOC [Beauchamp et al., 2007] in that they suggest a role of the EBA in a perception-action system that differentiates between oneself and others [Astafiev et al., 2004; Jeannerod, 2004]. Specifically, it has been speculated that the EBA might contribute to self-other differentiation via the detection of "violations of internal, possibly multimodal, body or action representations and incoming visual signals" [David et al., 2007, 2009]. Notably, David et al. [2007] reported increased functional coupling between the EBA and PPC when participants correctly identified mismatches between their movements and manipulated visual feedback, and Möhring et al. [2014] have speculated that the EBA may contribute to sensory-motor integration in the inferior parietal lobule. The fact that disrupting EBA activity using rTMS increased the proprioceptive mislocalization of one's hand during the RHI [Wold et al., 2014] could be due to the interference with the EBA's detecting multisensory mismatches (and signaling them to the IPS). In conjunction with these reports, our results show that body-selective regions in the IPS and LOC are essentially involved in the self-attribution of a dummy arm during the RHI, and also suggest a similar involvement of the PMv. Building up on this, our connectivity analysis sheds new light on the interplay of the frontoparietal circuit with the LOC/EBA during the RHI.

TABLE IV. Results of the PPI analysis

Anatomical region	Peak MNI			Peak t	Peak z	P (corrected)
	x	y	z			

Brain regions showing consistently increased functional connectivity with left LOC, left IPS, or left PMv during temporal visuo-tactile congruence versus incongruence across stimulation locations (main experiment)

Seed region: L IPS

L middle occipital gyrus (EBA)	−42	−72	14	3.06	5.64	< 0.001[a,b]
L precentral gyrus (PMv)	−56	4	38	3.02	5.59	0.012[a,c]
R middle temporal gyrus (EBA)	58	−66	−2	2.84	5.34	< 0.001[a,b]
L postcentral gyrus (SI)	−60	−20	36	2.59	4.98	0.012[a,d]
L precentral gyrus (PMd)	−34	−4	50	2.39	4.70	< 0.001[a]

Seed region: L PMv

R middle temporal gyrus (EBA)	54	−68	10	3.67	6.47	< 0.001[a,b]
R inferior frontal gyrus	52	38	−6	2.82	5.31	0.001[a]
L middle occipital gyrus (EBA)	−46	−76	8	2.36	4.66	0.006[a,b]
L superior temporal gyrus (SII)	−62	−30	16	2.24	4.48	0.003[d]
L postcentral gyrus (SI)	−60	−18	32	1.80	3.83	0.034[d]

Seed region: L LOC

L middle occipital gyrus (EBA)	−38	−76	2	2.47	4.82	< 0.001[a,b]
R middle temporal gyrus (EBA)	50	−66	4	1.93	4.02	0.038[a,b]
L supramarginal/postcentral gyrus (SI)	−62	−20	40	1.89	3.96	0.022[d]
L supramarginal gyrus (SII)	−52	−24	18	1.68	3.66	(0.060)[d]

Brain regions showing consistently increased functional connectivity with left LOC, left IPS, or left PMv during temporal and spatial congruence versus incongruence of visuo-tactile stimulation across stimulation locations

Seed region: L IPS

R middle temporal gyrus (EBA)	50	−68	0	4.51	Inf.	<0.001[a,b]
L middle occipital gyrus (EBA)	−52	−74	2	3.33	7.06	<0.001[a,b]
L precentral gyrus (PMv/d)	−54	2	38	2.64	5.96	0.026[a,c]
L supramarginal gyrus (SI/SII)	−60	−20	38	2.47	5.67	<0.001[a,d]
L precentral grus	−40	−8	58	2.25	5.32	<0.001[a]
R superior parietal lobule	28	−66	62	1.92	4.77	<0.001[a]
R postcentral gyrus (SI)	64	−16	34	1.82	4.60	0.008[a]
R superior temporal gyrus (SII)	62	−30	14	1.81	4.58	<0.001[a]
R precentral gyrus	44	2	34	1.43	3.94	0.018[a]
L inferior frontal gyrus (PMv)	−44	8	22	1.34	3.78	0.038[c]

Seed region: L PMv

R middle temporal gyrus (EBA)	56	−68	8	4.01	Inf.	<0.001[a,b]
L middle frontal gyrus	−44	14	50	2.34	5.46	0.008[a]
L superior temporal gyrus (SII)	−60	−32	16	2.28	5.37	0.003[a]
L precentral gyrus	−42	−8	48	2.17	5.19	<0.001[a]
L inferior frontal gyrus	−46	44	−8	2.02	4.94	<0.001[a]
L middle occipital gyrus (EBA)	−50	−78	8	1.97	4.85	<0.001[a,b]
R postcentral gyrus	36	−40	58	1.91	4.76	<0.001[a]
R middle frontal gyrus	38	−6	54	1.76	4.50	0.022[a]
L intraparietal sulcus	−38	−54	58	1.70	4.39	0.032[a,c]
R postcentral gyrus	62	−12	36	1.69	4.37	0.0021[a]

Seed region: L LOC

L inferior occipital gyrus (EBA)	−48	−78	−8	2.11	5.08	<0.001[a,b]
R inferior occipital gyrus (EBA)	54	−72	−14	2.01	4.92	<0.001[a,b]
L supramarginal/postcentral gyrus (SI)	−62	−20	40	1.80	4.56	0.009[d]
L postcentral gyrus (SII)	−56	−20	22	1.64	4.30	(0.051)[d]

Significance at $P < 0.05$ (FWE-corrected) based on
[a]Cluster-wise correction.
[b]Small volume correction using ROIs from the visual body-selectivity localizer.
[c]Small volume correction using predefined ROIs based on published literature.
[d]Small volume correction using ROIs from the tactile localizer.

The Rubber Hand Illusion Consistently Modulates Bottom-Up Connectivity within a Hierarchical Cortical Network

Our PPI analyses revealed a significantly increased functional coupling of the left IPS and the left PMv with the LOC/EBA under congruent versus incongruent visuotactile stimulation, consistent across all stimulation locations and type of (in)congruence. These results demonstrate an increased communication within a network comprising frontoparietal and occipitotemporal regions during the RHI, replicating previous reports of increased functional coupling of the left IPS with the PMv and the LOC during paradigms inspired by the RHI [Gentile et al., 2013; Guterstam et al., 2013]. Furthermore, our results show that the left PMv also increases its communication with the IPS and the LOC during the RHI.

Crucially, our DCM analysis extends these findings by demonstrating an information exchange between hierarchically lower areas in the SII and the LOC with the IPS, and between the IPS and the PMv. Our most important finding is that in this model, the connections from both the SII and the LOC to the IPS were significantly enhanced under the RHI, meaning that activity within the IPS was more strongly causally influenced by modulations of the bottom-up connections from the SII and the LOC. Notably, RFX BMS yielded the same winning models (featuring the same endogenous and effective connectivity) for stimulation at the palm, forearm, or both locations, and even when this analysis was repeated on the dataset implementing spatial (in)congruence. Although the model exceedance probabilities may perhaps not be considered very high, no other model lent itself as an alternative; this consistency across all comparisons suggests that the identified modulations of bottom-up connectivity might reflect a general process underlying the RHI.

There is one intriguing interpretation of these results, which fits nicely with our knowledge about the RHI and with the assumptions about cortical information flow made by predictive coding [Friston, 2005; Friston and Stephan, 2007], namely, that the brain tries to infer the most likely causes of its sensory input via the inversion and optimization of the current hierarchical generative model of the world. The latent architecture of the DCM selected by our formal Bayesian model comparison is compatible with a predictive coding scheme, in which hierarchically lower areas in the secondary somatosensory and extrastriate visual cortex communicate with higher-level multimodal convergence zones in the IPS, from where information is potentially exchanged with the PMv. According to the principles of predictive coding, cortical activity at each processing stage of a hierarchical network is determined by bottom-up, forward-flowing information and its comparison with the feedback provided by the predictions of the next-highest level about the current state, conveyed via top-down connections. Importantly, the informative quantity passed from lower to higher levels is the *error* that

arises from unpredicted data. It has been speculated that such information exchange may account for the RHI [Apps and Tsakiris, 2014; Hohwy, 2013; Limanowski and Blankenburg, 2013]. During the RHI, there is conflicting information about one's limb's position in space: Whereas vision of touch seems to recalibrate and represent the reference frame of the PPS in dummy arm-centered coordinates, somatosensation represents touch information still in PPS coordinates centered on the real arm [Makin et al., 2008; Macaluso and Maravita, 2010]. This unpredicted incompatibility of information about the location of the touched arm is likely to elicit corresponding errors—in contrast to the incongruent (control) conditions, where seen and felt touch are clearly dissociable and attributable to different arms based on their spatiotemporal incongruence, and therefore no such intersensory conflict and related prediction error should arise. The results of our DCM analysis may offer an explanation for how such illusion-evoked prediction errors are generated and propagated within the brain network involved in establishing body ownership and the peripersonal space.

We propose that first, the LOC/EBA associates the touch seen on the dummy arm and the touch felt on one's real arm due to their spatio-temporal congruence on a body part within the peripersonal space—by this cross-modal interaction, the seen touch becomes self-relevant [Macaluso and Maravita, 2010]. The observed significant positive correlation of LOC/EBA activity and the behavioral illusion scores supports this interpretation, since the stronger the interaction of vision and touch (the interplay of seen and felt touches), the higher the chances of the RHI actually emerging. However, following this cross-modal interaction, there is a mismatch between the coordinates in which the seen touch is now represented (i.e., centered onto the dummy arm), and the predictions about one's arm's location made by the higher-level IPS. The increases in the BOLD signal within the LOC/EBA revealed by our standard GLM analysis may reflect the generation of such a prediction error, and the correspondingly significantly enhanced connections from LOC/EBA to IPS under the RHI may reflect the feedforward communication of this prediction error to the hierarchically higher IPS. Given the knowledge we have about the RHI [Ehrsson, 2012; Makin et al., 2008; Tsakiris, 2010] and the functions of the IPS [Avillac et al., 2007; Beauchamp et al., 2010; Bremmer et al., 2001; Graziano and Cooke, 2006], we propose that the IPS tries to counter this error by integrating the multisensory touch information and recalibrating the coordinates of the somatosensory reference frame onto the visual reference frame. This could suppress (some) prediction error in the lower-level LOC/EBA. However, the IPS would also signal these "adjusted" predictions to the SII, where they do not match the incoming somatosensory information (i.e., the proprioceptive information about the position of one's real arm and the skin-based information about touches on it), and hence generate

another (somatosensory) prediction error. In line with this speculation, our winning model also shows an enhancement of the connections from the SII to the IPS, which could indicate the forward-propagation of such a prediction error.

In all models, the endogenous connectivity pattern revealed enhanced connections from the IPS to the LOC. This could reflect enhanced top-down attention to the visual modality during the processing of stimuli in the PPS. According to predictive coding, attention interacts with predictions via top-down connections by enhancing the precision of the prediction errors in a relevant modality [Feldman and Friston, 2010; Kok et al., 2012]. Vision is more informative than somatosensation for spatial and temporal judgments [Ernst and Bülthoff, 2004], thus it seems reasonable that the brain should try to put more weight on the incoming information from the visual cortex during processing visuo-tactile information in PPS [Ernst and Banks, 2002; Ma and Pouget, 2008; Pavani et al., 2000]. Using a paradigm similar to the RHI, Beauchamp et al. [2010] indeed showed that the connectivity between the IPS and the secondary somatosensory or extrastriate visual cortex strengthened depending on which modality was more reliable.

Similarly, we observed enhanced endogenous connections from the IPS to the PMv, which could indicate a potential information transfer about the PPS from parietal to frontal areas, in accordance with previous speculations [Graziano and Cooke, 2006; Makin et al., 2008]. Predictive coding states that representations at higher levels are encoded more abstractly, and at longer timescales, and some authors have argued for a more complex role of the PMv than the IPS in the RHI [Ehrsson et al., 2004]. The information transmitted from the IPS to the PMv could thus perhaps be a higher-level prediction error about current PPS representations that change depending on the type of (co)stimulation. The fact that our winning model did not feature modulations of PMv connectivity could perhaps be due to the duration of illusion induction being not intensive enough, or simply not long enough to fully engage the PMv [e.g., Petkova et al., 2011 used 30 s blocks]. Interestingly, in the left SII, left LOC, and in the bilateral PMv, activity in the congruent versus incongruent stimulation blocks was also differently modulated over time. This could hint toward a possible formation of prediction errors over time in brain areas that cannot (fully) adjust to the incorporation of the dummy arm. For example, the stronger PMv activity modulations could point towards a relatively slow formation of a high-level prediction error, whereas the stronger modulations in the SII and the LOC could mean that the prediction errors from vision and touch indeed mutually influence each other via the IPS.

Although in conjunction, our results comply with the idea that the information flow within the brain network underlying body ownership follows a predictive coding scheme (i.e., the forward-propagation of multisensory prediction errors), predictive coding is only one candidate explanation for the mechanisms underlying the brain's hierarchical inference about the world and the body. For example, the activation of the IPS and the PMv may also be interpreted as reflecting processes of multisensory integration that produce the coherent ownership experience [Ehrsson, 2012], since there is evidence for neuronal populations in these areas showing non-linear responses (i.e., enhanced or depressed relative to the sum of unimodal input) to spatially or temporally congruent multisensory stimuli, which may indicate such integrative processes (Ehrsson, 2012; Gentile et al., 2011; Maravita et al., 2003). An alternative interpretation of the enhanced connectivity to the IPS could be that activity in the SII and the LOC is evoking multisensory integration processes in the IPS when visual and tactile information is congruent, but not when it is incongruent. Note that this interpretation does not necessarily contradict a predictive coding scheme—the resolution of intersensory conflict via multisensory integration could also be seen as an updating of higher-level representations in these areas to suppress prediction errors at lower levels. Next, it should also be pointed out that the parameter estimates of the DCMs calculated on the Limanowski et al. [2014] dataset showed some differences to those of the main experiment. This may be attributable to differences of the spatially-incongruent experimental design: Unlike the temporally incongruent design, this design involved touching different body parts (i.e., different parts of the real arm) during the control condition than the RHI condition, which might be differently "surprising" for the brain or involve more spatial attention than the RHI condition, and could explain the difference in LOC-IPS modulations in these models. However, both datasets were best fit by the same model, featuring the same endogenous coupling and modulation of bottom-up connectivity from the SII and the LOC to the IPS under the illusion context.

Our findings could be extended by future studies implementing online behavioral measures, which would allow assessing illusory body ownership before, during, and after stimulation. Such designs could be used to clarify the specific role of the PMv during illusory body ownership and thus perhaps shed light on why our DCM analysis did not imply a strong modulation of the IPS-PMv pathway by the RHI, somewhat in contrast to previous assumptions about the importance of fronto-parietal interplay during the illusion [Gentile et al., 2013; Guterstam et al., 2013]. Likewise, although the most parsimonious endogenous connectivity model identified by all of our RFX BMS comparisons did not suggest PMv-SII connections, possible interactions between these areas due to their anatomical connections (Cipolloni and Pandya, 1999) might still be worth investigating in future experiments. Further, we observed increased cerebellar activity during congruent visuo-tactile stimulation, although none of the seed regions of our PPI analysis increased its functional coupling with the cerebellum. Although most models of

body ownership do not include subcortical structures [e.g., Makin et al., 2008; Tsakiris, 2010], there is some evidence for a role of the cerebellum [Gentile et al., 2013; Guterstam et al., 2013] or the putamen [Petkova et al., 2011] in multi-sensory processes underlying body ownership, and this role should be addressed by future research. Finally, our results emphasize areas in the left hemisphere during the self-attribution of a right dummy arm, consistent with the results of other right-hand RHI paradigms [Ehrsson et al., 2004; Gentile et al., 2013; Petkova et al., 2011], BOLD responses touch to the right hand across the body midline [Lloyd et al., 2002], and with evidence for a left-lateralized hand-preference within the LOC [Astafiev et al., 2004; Bracci et al., 2010, 2012; Zopf and Williams, 2013]. Future research should investigate whether such a lateralization is due to stimulation of the contralateral hand, or whether it generalizes to the ipsilateral hand.

CONCLUSION

This study demonstrated that visually body-selective areas in the frontal, parietal, and occipito temporal cortex work in concert during the illusory self-attribution of a dummy arm. The information flow revealed by the DCM analysis supports the importance of visuosomatosensory interaction and integration during the illusory self-attribution of body parts, and the hypothesis that the underlying multisensory integration mechanisms are implemented in the IPS and its connections. Crucially, our results suggest that the processing of touch information within the PPS during the RHI is not restricted to fronto-parietal areas, but involves information exchange with the occipitotemporal cortex, most likely with the EBA. The hierarchical inference underlying illusory self-attribution during the RHI could be implemented within this brain network according to the principles of predictive coding: Thus congruent visuotactile information on a body part interacts in hierarchically lower areas in the LOC, which may then detect violations of multisensory predictions about one's arm's position in space and pass the corresponding prediction error on to the higher-level IPS. Adjusting these predictions by the IPS may explain away this error, while eliciting another (somatosensory) prediction error about the touched arm's location in the SII. Importantly, the interpretation of our results in terms of predictive coding rests on the assumption of a pre-existing hierarchical model that constantly generates predictions about the self and evaluates the incoming data in the light of these predictions, thus acknowledging the necessary interaction between current sensory input and the constraints of an internal body model proposed by traditional accounts of the RHI. In sum, our findings lend support to the multisensory hypothesis of body ownership and propose predictive coding as a plausible implementation of the underlying cortical information exchange and integration.

REFERENCES

Apps MA, Tsakiris M (2014): The free-energy self: a predictive coding account of self-recognition. Neurosci Biobehav Rev 41: 85–97.

Armel KC, Ramachandran VS (2003): Projecting sensations to external objects: Evidence from skin conductance response. Proc R Soc London B, 270:1499–1506.

Ashburner J (2007): A fast diffeomorphic image registration algorithm. Neuroimage 38:95–113.

Astafiev SV, Stanley CM, Shulman GL, Corbetta M (2004): Extrastriate body area in human occipital cortex responds to the performance of motor actions. Nat Neurosci 7:542–548.

Avillac M, Hamed SB, Duhamel JR (2007): Multisensory integration in the ventral intraparietal area of the macaque monkey. J Neurosci 27:1922–1932.

Beauchamp MS, Yasar NE, Kishan N, Ro T (2007): Human MST but not MT responds to tactile stimulation. J Neurosci 27:8261–8267.

Beauchamp MS, Pasalar S, Ro T (2010): Neural substrates of reliability-weighted visual-tactile multisensory integration. Front Syst Neurosci 4:25.

Behzadi Y, Restom K, Liau J, Liu TT (2007): A component based noise correction method CompCor for BOLD and perfusion based fMRI. NeuroImage 37:90–101.

Blanke O (2012): Multisensory brain mechanisms of bodily self-consciousness. Nat Rev Neurosci 13:556–571.

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

Bracci S, Ietswaart M, Peelen MV, Cavina-Pratesi C (2010): Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. J Neurophysiol 103:3389–3397.

Bracci S, Cavina-Pratesi C, Ietswaart M, Caramazza A, Peelen MV (2012): Closely overlapping responses to tools and hands in left lateral occipitotemporal cortex. J Neurophysiol 107:1443–1456.

Bremmer F, Schlack A, Duhamel JR, Graf W, Fink GR (2001): Space coding in primate posterior parietal cortex. Neuroimage 14:S46–S51.

Brozzoli C, Gentile G, Ehrsson HH (2012): That's near my hand! Parietal and premotor coding of hand-centered space contributes to localization and self-attribution of the hand. J Neurosci 32:14573–14582.

Cipolloni PB, Pandya DN (1999): Cortical connections of the frontoparietal opercular areas in the rhesus monkey. J Comp Neurol 403:431–458.

Costantini M, Urgesi C, Galati G, Romani GL, Aglioti SM (2011): Haptic perception and body representation in lateral and medial occipito-temporal cortices. Neuropsychologia 49:821–829.

David N, Cohen MX, Newen A, Bewernick BH, Shah NJ, Fink GR, Vogeley K (2007): The extrastriate cortex distinguishes between the consequences of one's own and others' behavior. Neuroimage 36:1004–1014.

David N, Jansen M, Cohen MX, Osswald K, Molnar-Szakacs I, Newen A, Paus T (2009): Disturbances of self–other distinction after stimulation of the extrastriate body area in the human brain. Soc Neurosci 4:40–48.

Downing PE, Jiang Y, Shuman M, Kanwisher N (2001): A cortical area selective for visual processing of the human body. Science 293:2470–2473.

Downing PE, Peelen MV (2011): The role of occipitotemporal body-selective regions in person perception. Cogn Neurosci 2: 37–41.

Driver J, Spence C (2000): Multisensory perception: beyond modularity and convergence. Curr Biol 10:R731–R735.

Ehrsson HH (2012): The concept of body ownership and its relation to multisensory integration. In: The New Handbook of Multisensory Processes, B.E. Stein Ed, MA: MIT Press Cambridge.

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

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K (2005): A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage 25:1325–1335.

Ernst MO, Banks MS (2002): Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433.

Ernst MO, Bülthoff HH (2004): Merging the senses into a robust percept. Trends Cogn Sci 8:162–169.

Feldman H, Friston KJ (2010): Attention, uncertainty, and free-energy. Front Hum Neurosci 4:215.

Friston KJ (2004): Dynamic Causal Modelling. In: Human Brain Function, 2nd ed. Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, Zeki S, Ashburner JT, Penny WD. Burlington: Academic Press, 2004, p. 1063–1090.

Friston KJ, Harrison L, Penny W (2003): Dynamic causal modelling. Neuroimage 19:1273–1302.

Friston KJ (2005): A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biol Sci 360:815–836.

Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997): Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6:218–229.

Friston KJ, Penny WD, Glaser DE (2005): Conjunction revisited. NeuroImage 25:661–667.

Friston KJ, Stephan KE (2007): Free-energy and the brain. Synthese 159:417–458.

Friston KJ, Kiebel S (2009): Predictive coding under the free-energy principle. Philos Trans R Soc B: Biol Sci 364:1211–1221.

Gallagher S (2000): Philosophical conceptions of the self: implications for cognitive science. Trends Cogn Sci 4:14–21.

Gentile G, Petkova VI, Ehrsson HH (2011): Integration of visual and tactile signals from the hand in the human brain: an FMRI study. J Neurophysiol 105:910–922.

Gentile G, Guterstam A, Brozzoli C, Ehrsson HH (2013): Disintegration of multisensory signals from the real hand reduces default limb self-attribution: An fMRI study. J Neurosci 33: 13350–13366.

Graziano MS (1999): Where is my arm? The relative role of vision and proprioception in the neuronal representation of limb position. Proc Natl Acad Sci USA 96:10418–10421.

Graziano MS, Cooke DF, Taylor CS (2000): Coding the location of the arm by sight. Science 290:1782–1786.

Graziano MS, Botvinick MM (2002): How the brain represents the body: Insights from neurophysiology and psychology. Common mechanisms in perception and action: Attention and performance XIX, pp 136–157.

Graziano MS, Cooke DF (2006): Parieto-frontal interactions, personal space, and defensive behavior. Neuropsychologia 44:845–859.

Grefkes C, Fink, GR (2005): REVIEW: The functional organization of the intraparietal sulcus in humans and monkeys. J Anat 207: 3–17.

Guterstam A, Gentile G, Ehrsson HH (2013): The invisible hand illusion: multisensory integration leads to the embodiment of a discrete volume of empty space. J Cogn Neurosci 25:1078–1099.

Haggard P, Christakou A, Serino A (2007): Viewing the body modulates tactile receptive fields. Exp Brain Res 180:187–193.

Hohwy J (2012): Attention and conscious perception in the hypothesis testing brain. Front Psychol 3:96.

Hohwy J (2013): The predictive mind. Oxford: Oxford University Press.

Jackson PL, Meltzoff AN, Decety J (2006): Neural circuits involved in imitation and perspective-taking. Neuroimage 31:429–439.

Jeannerod M (2004): Visual and action cues contribute to the self–other distinction. Nat Neurosci 7:422–423.

Kennett S, Taylor-Clarke M, Haggard P (2001): Noninformative vision improves the spatial resolution of touch in humans. Curr Biol 11:1188–1191.

Kitada R, Johnsrude IS, Kochiyama T, Lederman SJ (2009): Functional specialization and convergence in the occipito-temporal cortex supporting haptic and visual identification of human faces and body parts: an fMRI study. J Cogn Neurosci 21: 2027–2045.

Kitada R, Yoshihara K, Sasaki AT, Hashiguchi M, Kochiyama T, Sadato N (2014): The Brain network underlying the recognition of hand gestures in the blind: The supramodal role of the extrastriate body area. J Neurosci 34:10096–10108.

Kok P, Jehee JF, de Lange FP (2012): Less is more: Expectation sharpens representations in the primary visual cortex. Neuron 75:265–270.

Konen CS, Haggard P (2014): Multisensory parietal cortex contributes to visual enhancement of touch in humans: A single-pulse TMS study. Cerebral Cortex 24:501–507.

Làdavas E (2002): Functional and dynamic properties of visual peripersonal space. Trends in cognitive Sci 6:17–22.

Limanowski J, Blankenburg F (2013): Minimal self-models and the free energy principle. Frontiers in Human Neuroscience 7:547.

Limanowski J, Lutti A, Blankenburg F (2014): The extrastriate body area is involved in illusory limb ownership. NeuroImage 86:514–524.

Lloyd DM, Shore DI, Spence C, Calvert GA (2002): Multisensory representation of limb position in human premotor cortex. Nat Neurosci 6:17–18.

Lutti A, Thomas DL, Hutton C, Weiskopf N (2012): High-resolution functional MRI at 3T: 3D/2D echo-planar imaging with optimized physiological noise correction. Magn Reson Med 69: 1657–1664.

Ma WJ, Pouget A (2008): Linking neurons to behavior in multisensory perception: A computational review. Brain Res 1242:4–12.

Macaluso E, Driver J (2001): Spatial attention and crossmodal interactions between vision and touch. Neuropsychologia 39: 1304–1316.

Macaluso E, Maravita A (2010): The representation of space near the body through touch and vision. Neuropsychologia 48:782–795.

Macey PM, Macey KE, Kumar R, Harper RM (2004): A method for removal of global effects from fMRI time series. NeuroImage 22:360–366.

Makin TR, Holmes NP, Zohary E (2007): Is that near my hand? Multisensory representation of peripersonal space in human intraparietal sulcus. J Neurosci 27:731–740.

Makin TR, Holmes NP, Ehrsson HH (2008): On the other hand: dummy hands and peripersonal space. Behav Brain Res191:1–10.

Maravita A, Spence C, Driver J (2003): Multisensory integration and the body schema: close to hand and within reach. Curr Biol 13:R531–R539.

Mazaika P, Hoeft F, Glover GH, Reiss AL (2009): Methods and Software for fMRI Analysis for Clinical Subjects. San Francisco, CA: Annual Meeting of the Organization for Human Brain Mapping.

Meredith MA, Stein BE (1986): Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J Neurophysiol 56:640–662.

Möhring N, Shen C, Neuhaus AH (2014): Spatiotemporal dynamics of early cortical gesture processing. NeuroImage 99:42–49.

Murray SO, Kersten D, Olshausen BA, Schrater P, Woods DL (2002): Shape perception reduces activity in human primary visual cortex. Proc Natl Acad Sci USA 99:15164–15169.

Oldfield RC (1971): The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113.

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

Petkova VI, Björnsdotter M, Gentile G, Jonsson T, Li TQ, Ehrsson HH (2011): From part-to whole-body ownership in the multisensory brain. Curr Biol 21:1118–1122.

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

Rizzolatti G, Fadiga L, Fogassi L, Gallese V (1997): The space around us. Science 277:190–191.

Schwarzlose RF, Baker CI, Kanwisher N (2005): Separate face and body selectivity on the fusiform gyrus. J Neurosci 25:11055–11059.

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009): Bayesian model selection for group studies. Neuroimage 46: 1004–1017.

Striem-Amit E, Amedi A (2014): Visual cortex extrastriate body-selective area activation in congenitally blind people "seeing" by using sounds. Curr Biol 24:687–692.

Summerfield C, Koechlin E (2008): A neural representation of prior information during perceptual inference. Neuron 59:336–347.

Taylor-Clarke M, Kennett S, Haggard P (2002): Vision modulates somatosensory cortical processing. Curr Biol 12:233–236.

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

Tsakiris M, Haggard P (2005): The rubber hand illusion revisited: visuotactile integration and self-attribution. J Exp Psychol Hum Percept Perform 31:80–91.

Urgesi C, Candidi M, Ionta S, Aglioti SM (2007): Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. Nat Neurosci 10:30–31.

van Elk M (2014): The left inferior parietal lobe represents stored hand-postures for object use and action prediction. Front Psychol 5:333.

Wold A, Limanowski J, Walter H, Blankenburg F (2014): Proprioceptive drift in the rubber hand illusion is intensified following 1 Hz TMS of the left EBA. Front Hum Neurosci 8:390.

Whiteley L, Kennett S, Taylor-Clarke M, Haggard P (2004): Facilitated processing of visual stimuli associated with the body. Perception 33:307–314.

Zopf R, Williams MA (2013): Preference for Orientations Commonly Viewed for One's Own Hand in the Anterior Intraparietal Cortex. PloS One 8:e53812.

Study 3

Wold, A., **Limanowski, J.**, Walter, H., & Blankenburg, F. (2014). Proprioceptive drift in the rubber hand illusion is intensified following 1 Hz TMS of the left EBA. *Frontiers in Human Neuroscience, 8*, 390. doi: 10.3389/fnhum.2014.00390

Proprioceptive drift in the rubber hand illusion is intensified following 1 Hz TMS of the left EBA

*Andrew Wold[1,2,3] *, Jakub Limanowski[1,2] , Henrik Walter[1,3] and Felix Blankenburg[1,2]*

[1] Berlin School of Mind and Brain, Humboldt University, Berlin, Germany
[2] Neurocomputation and Neuroimaging Unit, Freie Universität Berlin, Berlin, Germany
[3] Division of Mind and Brain Research, Charité University of Medicine, Berlin, Germany

The rubber hand illusion (RHI) is a paradigm used to induce an illusory feeling of owning a dummy hand through congruent multisensory stimulation. Thus, it can grant insights into how our brain represents our body as our own. Recent research has demonstrated an involvement of the extrastriate body area (EBA), an area of the brain that is typically implicated in the perception of non-face body parts, in illusory body ownership. In this experiment, we sought causal evidence for the involvement of the EBA in the RHI. Sixteen participants took part in a sham controlled, 1 Hz repetitive transcranial magnetic stimulation (rTMS) experiment. Participants received (RHI condition) or asynchronous (control) stroking and were asked to report the perceived location of their real hand, as well as the intensity and the temporal onset of experienced ownership of the dummy hand. Following rTMS of the left EBA, participants misjudged their real hand's location significantly more toward the dummy hand during the RHI than after sham stimulation. This difference in "proprioceptive drift" provides the first causal evidence that the EBA is involved in the RHI and subsequently in body representation and further supports the view that the EBA is necessary for multimodal integration.

Keywords: rubber hand illusion, transcranial magnetic stimulation, extrastriate body area, body representation, proprioceptive drift

INTRODUCTION

The rubber hand illusion (RHI) is a well-established paradigm to manipulate the sense of body ownership in healthy individuals (Botvinick and Cohen, 1998). When one's own occluded hand and an anatomically congruent dummy hand are stroked synchronously, this leads to a feeling of ownership over the dummy hand that is generally interpreted as a momentary incorporation of the seen dummy hand into the participant's body representation (Ehrsson et al., 2004; Tsakiris and Haggard, 2005). The RHI has been explained as a result of multisensory information integration in a hierarchically organized cortical network that ultimately constructs and maintains one's body representation (Hohwy, 2007; Makin et al., 2007; Tsakiris, 2010; Blanke, 2012; Limanowski and Blankenburg, 2013; Apps and Tsakiris, 2014). Thus, it is assumed that visual, proprioceptive, and somatosensory input feed into higher-order multimodal integration areas (Tsakiris, 2010; Blanke, 2012). Previous research, however, has indicated that visual areas work merely on low level processing such as representing visual form (Makin et al., 2007), whereas recent research has emphasized some of these visual areas as playing a more sophisticated role in body representation (Ionta et al., 2011; Gentile et al., 2013; Limanowski et al., 2014), particularly focusing on the so-named extrastriate body area (EBA; Downing et al., 2001).

The EBA is an occipito-temporal visual region that has gained considerable attention in recent literature because of its selective, strong response to non-face body parts (Downing et al., 2001; Peelen and Downing, 2007; Downing and Peelen, 2011) and

contribution to explicit representations of identity (Urgesi et al., 2007), body configurations (Pitcher et al., 2009), and goal-directed actions (Wiggett and Downing, 2011). Along these lines, a recent fMRI study by Limanowski et al. (2014) demonstrated an involvement of the EBA in illusory body ownership. The involvement of the EBA in the RHI provides support for previous speculations about a role of this region in the representation of one's body (Costantini et al., 2011; Apps and Tsakiris, 2014).

In this experiment, we sought causal evidence for the EBA's role in body ownership by applying repetitive transcranial magnetic stimulation (rTMS) over the left EBA during the RHI (following Limanowski et al., 2014, who found ownership-related activity changes in the left EBA, contralateral to the arm subjected to the RHI). TMS to the EBA has causally proven the EBA's pivotal role in non-face body part perception (Urgesi et al., 2004). Previous studies applying TMS over the left inferior parietal lobe (IPL; Kammers et al., 2009) and right temporo-parietal junction (TPJ; Tsakiris et al., 2008) have successfully modulated behavioral measurements of the RHI, but until now no one has explored this combination with regard to the EBA.

We hypothesized that, if the EBA is indeed involved in the processing of one's body representation, interfering with its neural activity during illusions of body ownership should result in significant changes on the behavioral measures of the RHI, namely verbal reports, and the so-named "proprioceptive drift," the relative displacement of the perceived location of one's own hand toward the location of the rubber hand after the RHI, compared with a pre-stimulation baseline (e.g., Tsakiris and Haggard, 2005).

MATERIALS AND METHODS

PARTICIPANTS

Nineteen participants (11 female, median age: 25, range: 21–42) took part in this experiment. Three participants (two female) did not experience the illusion, and were therefore excluded after the first session. That 16 of 19 participants did experience is in accordance with classic RHI literature estimating the illusion to function in approximately 80% of the population (Botvinick and Cohen, 1998). All participants gave written informed consent to participate in the study, which had been approved by the local University Hospital Ethics Committee (Charité - Universitätsmedizin Berlin) and was within limits of safety guidelines (Rossi et al., 2009).

DESIGN AND PROCEDURE

In a two session, single-blind, sham-controlled, counterbalanced crossover design, participants received either 20 min of real or sham 1 Hz rTMS (1200 pulses) over their left EBA, which was functionally defined by a standard EBA localizer in a separate fMRI session (see **Figure 1**). The participants were seated in front of a meter long table with a window that gave them full view of a realistic right dummy hand. The table was set at an angle of 15°, had an additional opaque layer to cover the entire surface. The participant's right hand was positioned at 20 cm distance from the dummy hand. Participants were instructed to keep their right hand still throughout the experiment and were observed by the experimenter, who only proceeded if this was actually the case. The experiment began with nine proprioceptive judgments (as a calibration to obtain a subject-specific baseline against which the following judgments were compared), then four randomly assigned stroking blocks (two synchronous, two asynchronous) comprised the pre- and post-stimulation sessions. Each stroking session lasted 3 min beginning by participant's performing a button click with a computer mouse under their left hand. During stroking, participants were instructed to click again when they experienced the RHI onset. Hand stroking was delivered with paintbrushes by the experimenter, at an approximate frequency of 1 Hz, and included vertical stroking (from knuckle to fingertip) of fingers and horizontal stroking (from left most knuckle to right most). Asynchronous stroking was displaced both temporally and spatially. Following the stroking block, participants were asked for three drift measurements and to rate the intensity of the illusion. Participants were asked to maintain focus on the dummy hand during stroking and to look away from the measuring tape between proprioceptive trials; so that they would not focus on any particular spot on the set-up. Total time for an entire block was approximately 3.5 min. After the four pre-stimulation blocks, participants moved into a different chair in the same room to receive real or sham TMS, and then repeated the same procedure of calibration followed by another four 3.5-min blocks. Motor threshold (MT) was only assessed during the first session.

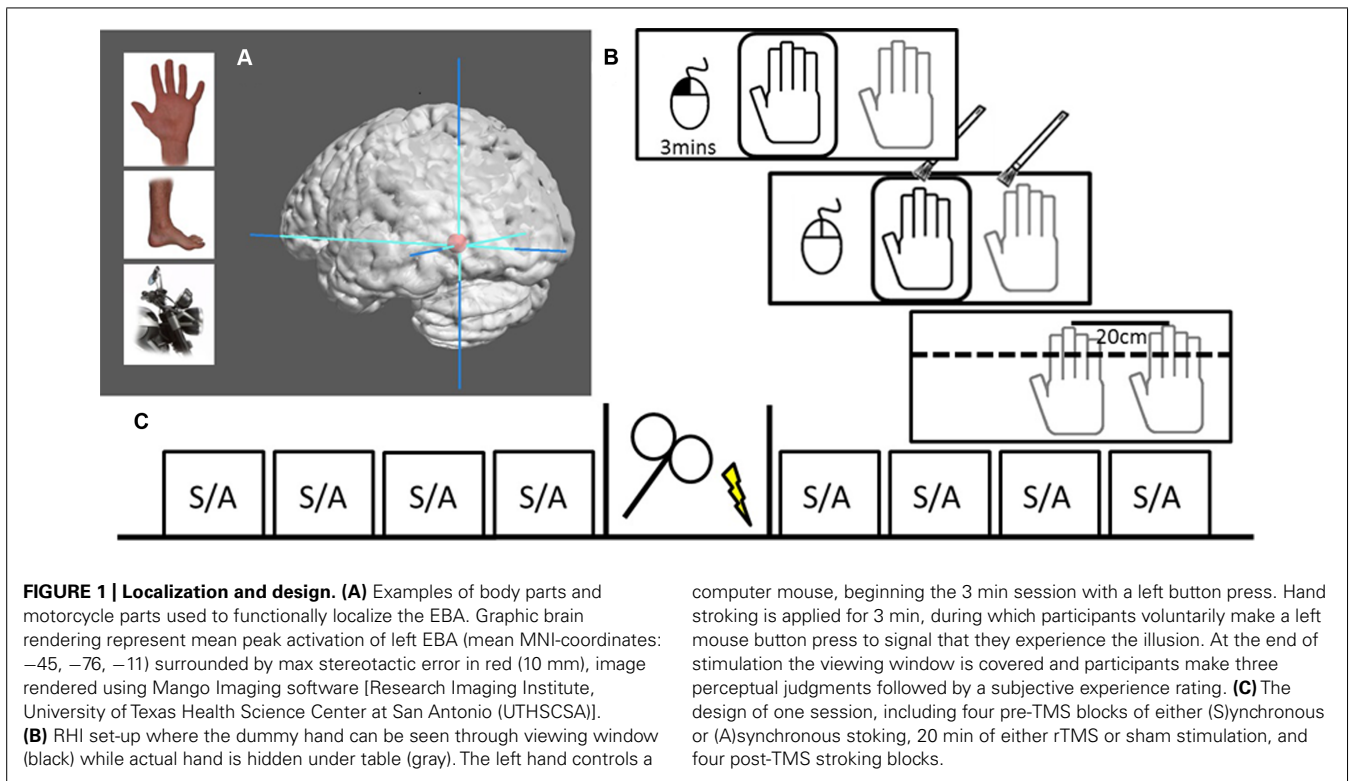### rTMS TARGET SELECTION AND STEREOTACTIC NAVIGATION

All participants had their left EBA's localized via a 3 Tesla MRI scanner (Magnetom Tim Trio, Siemens, Erlangen, Germany) with a 32-channel head coil. Using a standard EBA localizer task, two 6-min sessions of eight randomly assigned blocks of body parts or motorcycle parts were presented to participants. In each task session, 175 functional volumes were acquired with a T2*-weighted EPI-sequence, each consisting of 37 oblique-axial slices (voxel size = 3 mm × 3 mm × 3 mm, 64 × 64 matrix, FOV = 192 mm, TR = 2000 ms, TE = 30 ms, flip angle = 70°). After the functional runs, a high-resolution T1-weighted structural image was acquired for each participant (3D MPRAGE, voxel size = 1 mm × 1 mm × 1 mm, FOV = 256 mm × 256 mm, 176 slices, TR = 1900 ms, TE = 2.52 ms, flip angle = 9°). Six subjects already had their EBA localized in an fMRI experiment we conducted recently (Limanowski et al., 2014). The functional as well as anatomical data for these subjects were taken from this previous study. Data were preprocessed and analyzed using SPM 8 (Wellcome Department of Cognitive Neurology, London, UK) and BrainVoyager (Brain Innovation B.V., Netherlands).

After motion correction, participant's functional images were coregistered to their respective structural images and smoothed with a 5-mm full width at half maximum Gaussian kernel, but were not normalized. For each subject, the two sessions were included in one general linear model and the contrast body parts minus motorcycle parts was computed to localize EBA (i.e., activity specific to vision of body parts). The localization was individually derived for each participant. Six of our participants came from a previous study (Limanowski et al., 2014) but utilized non-normalized data for navigation. Statistical parametric maps of the body parts versus motorcycle parts were thresholded at a significance threshold of $p < 0.01$ to $p < 0.001$ uncorrected. This anatomical and functional data were imported from SPM into BrainVoyager QX, aligned in the AC–PC space, and marked with a target file corresponding to the individual left EBA (via peak activation of the EBA localizer). Using this target file and head mesh reconstructions of the participants brain, we then coregistered TMS coil and head position using a Zebris CMS20S tracking device (Zebris Medical GmbH, Isny, Germany), allowing us to navigate coil position relative to target, in real-time. After the localization session the data were converted into Talairach space using the normalization procedure in BrainVoyager. The Talairach coordinates where then transformed to MNI-coordinates using the TAL2MNI Matlab code (tal2icbm_spm; retrieved from www.brainmap.org/icbm2tal/tal2icbm_spm.m). The mean MNI-coordinates (±SEM) for the EBA of the 16 participants corresponded to $x = -45 \pm 0.95$, $y = -76 \pm 1.9$, $z = -11 \pm 1.7$, which also align to previously published locations of the EBA (Downing et al., 2001).

TRANSCRANIAL MAGNETIC STIMULATION

Transcranial magnetic stimulation was applied using Magstim Rapid[2] device (Magstim, Whitland, UK). In order to establish the appropriate intensity of stimulation, resting MT of the left hemisphere was determined according to the standard MT procedure (mean ± SEM = 56.2 ± 1.3% of maximum stimulator output; Schutter and van Honk, 2006). Two stimulation intensities were used: a high- and a low-intensity rTMS (80 and 40% of the MT) for a total of 1200 pulses at a frequency of 1 Hz. The figure-of-eight coil was turned tangentially to the scalp and the handle aligned along the rostrocaudal plane. The low-intensity rTMS application

FIGURE 1 | Localization and design. (A) Examples of body parts and motorcycle parts used to functionally localize the EBA. Graphic brain rendering represent mean peak activation of left EBA (mean MNI-coordinates: −45, −76, −11) surrounded by max stereotactic error in red (10 mm), image rendered using Mango Imaging software [Research Imaging Institute, University of Texas Health Science Center at San Antonio (UTHSCSA)]. **(B)** RHI set-up where the dummy hand can be seen through viewing window (black) while actual hand is hidden under table (gray). The left hand controls a computer mouse, beginning the 3 min session with a left button press. Hand stroking is applied for 3 min, during which participants voluntarily make a left mouse button press to signal that they experience the illusion. At the end of stimulation the viewing window is covered and participants make three perceptual judgments followed by a subjective experience rating. **(C)** The design of one session, including four pre-TMS blocks of either (S)ynchronous or (A)synchronous stoking, 20 min of either rTMS or sham stimulation, and four post-TMS stroking blocks.

(assumed to be neurally ineffective) served as the sham condition for non-specific effects of rTMS, such as the "click" sound and the scalp sensation inevitably associated with rTMS delivery (Auksztulewicz et al., 2011). Full stimulation was lowered from 100% of MT to reduce the current spread into neighboring cortical regions (Fitzgerald et al., 2006).

MEASUREMENTS

First, proprioceptive drift was assessed by having participants report a number on a measuring tape that best corresponded to the perceived location of the index finger of their hidden right hand. Three measurements were always taken while the dummy hand was hidden from view by an additional opaque layer, on top of which the measuring tape was placed. The experimenter had a list of randomized displacements (ranging 1–20 cm) at which to hold a standard measuring tape and instructed the participant to verbalize the felt position of the right index finger to the nearest centimeter. Before the RHI was induced, the set-up was calibrated with nine perceptual judgments before the pre- and post-stimulation block. Mean calibrations (±SEM) for all participants did not differ significantly between the pre- and the post-stimulation sessions (−0.63 ± 0.66 cm, and −0.63 ± 0.82 cm, relative to actual finger position). Second, participants were asked to rate the intensity of the ownership illusion after each stroking block. The question was "How strongly do you feel that this hand could belong to your own body?" on a 7-point Likert scale from −3 (strongly disagree) to 0 (unsure of what I felt) to +3 (strongly agree). Third, for each stimulation block, participants were instructed to make a mouse click with their left hand as soon as they experienced the ownership illusion, if they experienced it at all. This time point was taken to represent the temporal onset of illusory ownership (see Ehrsson et al., 2004).

CALCULATIONS AND DATA ANALYSIS

All behavioral data were analyzed with SPSS (version 12.2) software. Proprioceptive drift was calculated per block as the mean of the three measurements minus the individual's calibration value (relative to the edge of set-up), making pre-stroking subjectively felt finger location the proprioceptive reference. The rating was reported as a value from −3 to +3. Onset was calculated as the total duration minus the onset (i.e., 180 s − onset), so that higher values represent faster responses to RHI. For each measurement of each pre- and post-stimulation session, the two synchronous and asynchronous blocks were separately averaged.

Data were analyzed using repeated measures MANOVA and ANOVA (all measurements passed a Shapiro–Wilk test for normality). *Post hoc* paired *t*-tests were calculated to compare real TMS to sham TMS, significance was assessed using a Bonferroni corrected alpha level of $\alpha = 0.05$ and two-tailed distributions.

RESULTS

All participants completed the entire experiment and had no adverse side effect associated with TMS (for descriptive statistics, see **Figure 2B**). The repeated measures MANOVA including three measurements (proprioceptive drift, rating, and illusion onset), two levels of stimulation (rTMS/sham), two levels of session (pre/post stimulation), and two levels of stroking condition (synchronous/asynchronous) showed a significant main effect for
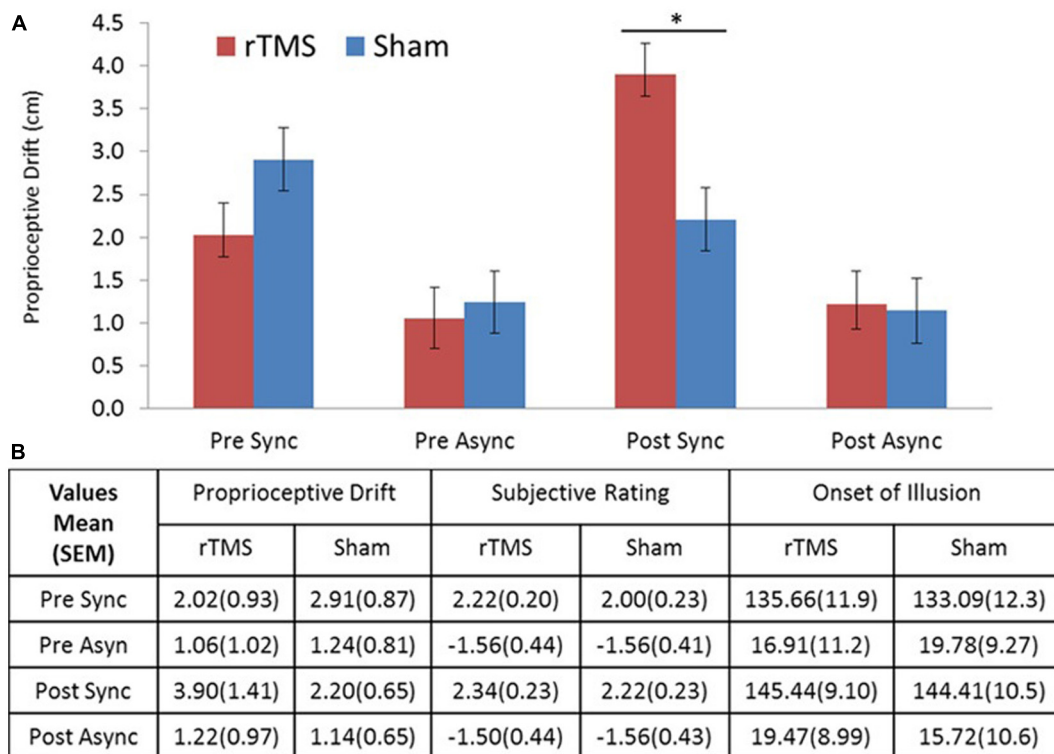
FIGURE 2 | (A) Drift results. Proprioceptive drift in centimeters showed a significant difference between rTMS and sham stimulation for post-TMS sync stroking. **(B)** Table of descriptive statistics. Means and standard errors of the mean (SEM) for proprioceptive drift, ownership rating, and illusion onset (s). Sync, synchronous; Async, asynchronous stroking. All error bars represent SEM, star represents significance at α = 0.05 (Bonferroni corrected).

measurement $[F(2,30) = 108.9, p < 0.001]$ and stroking condition $[F(2,30) = 107.8, p < 0.001]$. In addition, this analysis revealed a significant four-way interaction between stimulation test × session × stroking condition $[F(2,14) = 8.475, p = 0.004]$. To make the analysis more comprehensible, we computed three-way repeated measure ANOVAs for each of the individual measures. These analyses revealed that the three-way interaction (stimulation × session × stroking condition) in proprioceptive drift was significant $[F(1,15) = 15.29, p < 0.001]$, whereas the corresponding three-way interactions in ownership rating or illusion onset were not $[F(1,15) = 0.256, p = 0.620$ and $F(1,15) = 0.409, p = 0.532$, respectively]. A closer inspection of the significant interaction in proprioceptive drift (see **Figure 2**) elucidates the effect of increased drift in the synchronous stroking condition following rTMS. Based on our assumption that rTMS stimulation would produce an effect compared to sham stimulation during synchronous stroking but not asynchronous stroking, we compared proprioceptive drift in the post-stimulation session using *post hoc* paired *t*-tests, which revealed a significantly $(p < 0.025)$ stronger effect of rTMS versus sham stimulation on proprioceptive drift during synchronous stroking $[t(15) = 2.578, p = 0.021]$, but not during asynchronous stroking $[t(15) = 0.161, p = 0.876]$.

Since it has been shown that men and women express different EBA functional lateralization effects in body perception (Aleong and Paus, 2010), we ran the repeated measure MANOVA with gender as a between subjects factor. This analysis revealed no significant effect of gender $[F(1,15) = 0.001, p = 0.976]$, neither did further repeated measures ANOVAs with gender as between subjects variable for the individuals RHI measurements [proprioceptive drift: $[F(1,15) = 0.103, p = 0.753$; ownership rating: $F(1,15) = 0.022, p = 0.883$; illusion onset: $F(1,15) = 0.001, p = 0.995]$.

DISCUSSION

Repetitive TMS over the left EBA, compared with sham stimulation, resulted in an increased proprioceptive drift in the RHI (synchronous stroking of the rubber hand and real hand) versus the control condition (asynchronous stroking). Ownership rating and illusion onset, like drift, could be differentiated by stroking condition (synchronous > asynchronous), but unlike proprioceptive drift were not differently affected by rTMS versus sham stimulation. Together, our results suggest a causal involvement of the left EBA in the processing of an own body representation, as discussed in the following.

The EBA has been shown to be causally involved in non-face body perception (Urgesi et al., 2004) and the processing of haptic and visual information (Costantini et al., 2011). In this way, the EBA may be part of a hierarchical network of brain areas – and due to its potentially multimodal processing capacities likely located at intermediate levels in this hierarchy – representing the own body in a probabilistic fashion (Hohwy, 2013; Limanowski

and Blankenburg, 2013; Apps and Tsakiris, 2014). Moreover, Jackson et al. (2006) propose that the EBA is not only important for the visual processing of body parts but mapping that representation of another body onto one's own body. Our results – an increased behavioral effect of illusory limb ownership – may be seen as support for this view. The causal modulation of EBA activity by rTMS may increase the drift toward the rubber hand, not by confusion of the location of one's own hand, but by allowing an illusory body part to be incorporated into one's body representation.

Proprioceptive drift is a multimodal measure combining the processing of visual, tactile, and proprioceptive information (Botvinick and Cohen, 1998; Tsakiris and Haggard, 2005). Tsakiris (2010) proposes that there is a pre-existing body model containing a reference description of the visual, anatomical and structural properties of the body, and that during the RHI certain multimodal brain regions (e.g., TPJ) act as comparators, matching the (dummy) viewed hand with the body model reference. Modulating this comparator process has an effect on proprioceptive drift, as is evident in rTMS experiments stimulating the rTPJ during the RHI (Tsakiris et al., 2008). In our study, rTMS over EBA also affected drift, but led to an increase (versus a decreased drift found by Tsakiris et al., 2008). This result suggests that the EBA might be also involved in this process of body model-comparison, potentially integrating visual body representations with somatosensory and proprioceptive information with regard to body parts (Costantini et al., 2011; Blanke, 2012; Apps and Tsakiris, 2014). It should be noted that we only found an effect of rTMS over left EBA on proprioceptive drift but not the verbal ownership ratings or onsets. This could be because of ceiling effects of the reported illusion onsets or the ownership rating scale not being sensitive enough to discriminate fine differences in subjective experience. However, as has been previously reported, mechanisms of RHI measurements are different and may not actually be associated, as is documented in the case of the dissociation between drift and ownership (Holmes et al., 2006; Rohde et al., 2011). Future work will have to investigate the exact role of EBA in the construction and maintenance of one's body model – most importantly it will have to be clarified whether EBA acts as a multimodal integration region and/or as a region creating a body reference in terms of visual appearance and location.

As mentioned above our study is not the only one to combine RHI with TMS. However, this study differs from previous studies by stimulating the (left) EBA during RHI and by showing an increase in proprioceptive drift due to a TMS intervention. Kammers et al. (2009) stimulated the IPL showing that IPL affects the perceptual, rather than the sensorimotor, representation of the body. This was in accordance with neuroimaging studies showing the IPL to be involved in the RHI and hypothesized to play a role in the perception of size and location (Ehrsson et al., 2005). The rTMS results of Kammers et al. (2009) yielded a similar magnitude of effect but in the opposite direction (i.e., reduced drift). Tsakiris et al. (2008) stimulated the rTPJ, an area known for its role in perspective taking (Saxe and Wexler, 2005) and multisensory integration (Tsakiris and Haggard, 2005). Following online single pulse TMS, Tsakiris et al. (2008) also found reduced drift compared to sham stimulation. It should be noted that the RHI controls and

calculations differ among these studies, but we suggest that the EBA is not acting on the RHI through perception of size and location, like the IPL. Moreover, the difference in TMS effect on the EBA and regions such as the TPJ could reflect a difference in the underlying cortical mechanism used to process a sense of bodily self. Arzy et al. (2006) distinguishes between the left EBA and right TPJ for neural mechanisms of embodiment (shortly defined as the sense of being localized within one's physical body), showing that the EBA is more present for embodied processing while the TPJ is distinct to disembodied processing. This difference could be the basis for the TMS effect observed, which we would assume is a reduced ability to embody one's own hand or to visually discriminate the rubber hand from one's own (i.e., increasing drift due to improved acceptance of rubber hand) when stimulating the EBA and a reduced ability to actualize a disembodied perspective (in the case of the RHI: reduced ability to self-attribute a foreign body part) when stimulating the TPJ. It should be noted that research shows bilateral TPJ activation in out-of-body-experience (Blanke et al., 2005) and self-location tasks (i.e., first-person perspective; Ionta et al., 2011). When distinguishing the TPJ and EBA in terms of their role for embodiment processing, one should therefore pay attention to hemispheric lateralization; our findings hence have to be interpreted with some caution as we only stimulated left EBA and cannot therefore reveal the relative role of EBA functional specialization and lateralization. Although time restraints limited the amount of rating questions in our studies, future work may benefit from including additional ratings pertaining to other aspects of the RHI besides ownership (e.g., increased visual similarity between real and fake hand; discussed as "perceptual assimilation" in Longo et al., 2008). Our null finding with regard to gender suggests that our data were not affected by the known gender difference in EBA functional specialization (Aleong and Paus, 2010), a larger sample would be needed to address the role of gender specific effects of the EBAs. Future studies focusing on these issues may help to uncover the differential EBA effects noted in our RHI experiment.

In sum, our study shows an increase in proprioceptive drift due to an rTMS intervention over left EBA; this increased mislocalization of one's real hand position provides evidence for a causal involvement of the EBA in changes in one's own body representation.

AUTHOR CONTRIBUTIONS

Andrew Wold, Jakub Limanowski, and Felix Blankenburg developed the concept, Andrew Wold, Henrik Walter, Felix Blankenburg, and Jakub Limanowski contributed to the design, Andrew Wold and Jakub Limanowski conducted the experiment, Andrew Wold did the analysis, and all co-authors contributed to the manuscript.

REFERENCES

Aleong, R., and Paus, T. (2010). Neural correlates of human body perception. *J. Cogn. Neurosci.* 22, 482–495. doi: 10.1162/jocn.2009.21211

Apps, M. A., and Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neurosci. Biobehav. Rev.* 41C, 85–97. doi: 10.1016/j.neubiorev.2013.01.029

Arzy, S., Thut, G., Mohr, C., Michel, C. M., and Blanke, O. (2006). Neural basis of embodiment: distinct contributions of temporoparietal junction and extrastriate body area. *J. Neurosci.* 26, 8074–8081. doi: 10.1523/JNEUROSCI.0745-06.2006

Auksztulewicz, R., Spitzer, B., Goltz, D., and Blankenburg, F. (2011). Impairing somatosensory working memory using rTMS. *Eur. J. Neurosci.* 34, 839–844. doi: 10.1111/j.1460-9568.2011.07797.x

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nat. Rev. Neurosci.* 13, 556–571. doi: 10.1038/nrn3292

Blanke, O., Mohr, C., Michel, C. M., Pascual-Leone, A., Brugger, P., Seeck, M., et al. (2005). Linking out-of-body experience and self processing to mental own-body imagery at the temporoparietal junction. *J. Neurosci.* 25, 550–557. doi: 10.1523/JNEUROSCI.2612-04.2005

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

Costantini, M., Urgesi, C., Galati, G., Romani, G. L., and Aglioti, S. M. (2011). Haptic perception and body representation in lateral and medial occipito-temporal cortices. *Neuropsychologia* 49, 821–829. doi: 10.1016/j.neuropsychologia.2011.01.034

Downing, P. E., Jiang, Y., Shuman, M., and Kanwisher, N. G. (2001). A cortical area selective for visual processing of the human body. *Science* 293, 2470–2473. doi: 10.1126/science.1063414

Downing, P. E., and Peelen, M. V. (2011). The role of occipitotemporal body-selective regions in person perception. *Cogn. Neurosci.* 2, 186–203. doi: 10.1080/17588928.2011.582945

Ehrsson, H. H., Holmes, N. P., and Passingham, R. E. (2005). Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. *J. Neurosci.* 25, 10564–10573. doi: 10.1523/JNEUROSCI.0800-05.2005

Ehrsson, H. H., Spence, C., and Passingham, R. E. (2004). That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science* 305, 875–877. doi: 10.1126/science.1097011

Fitzgerald, P. B., Fountain, S., and Daskalakis, Z. J. (2006). A comprehensive review of the effects of rTMS on motor cortical excitability and inhibition. *Clin. Neurophysiol.* 117, 2584–2596. doi: 10.1016/j.clinph.2006.06.712

Gentile, G., Guterstam, A., Brozzoli, C., and Ehrsson, H. H. (2013). Disintegration of multisensory signals from the real hand reduces default limb self-attribution: an fMRI study. *J. Neurosci.* 33, 13350–13366. doi: 10.1523/JNEUROSCI.1363-13.2013

Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche* 13, 1–20.

Hohwy, J. (2013). *The Predictive Mind.* Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199682737.001.0001

Holmes, N. P., Snijders, H. J., and Spence, C. (2006). Reaching with alien limbs: visual exposure to prosthetic hands in a mirror biases proprioception without accompanying illusions of ownership. *Percept. Psychophys.* 68, 685–701. doi: 10.3758/BF03208768

Ionta, S., Heydrich, L., Lenggenhager, B., Mouthon, M., Fornari, E., Chapuis, D., et al. (2011). Multisensory mechanisms in temporo-parietal cortex support self-location and first-person perspective. *Neuron* 70, 363–374. doi: 10.1016/j.neuron.2011.03.009

Jackson, P. L., Meltzoff, A. N., and Decety, J. (2006). Neural circuits involved in imitation and perspective-taking. *Neuroimage* 31, 429–439. doi: 10.1016/j.neuroimage.2005.11.026

Kammers, M. P., Verhagen, L., Dijkerman, H. C., Hogendoorn, H., De Vignemont, F., and Schutter, D. J. (2009). Is this hand for real? Attenuation of the rubber hand illusion by transcranial magnetic stimulation over the inferior parietal lobule. *J. Cogn. Neurosci.* 21, 1311–1320. doi: 10.1162/jocn.2009.21095

Limanowski, J., and Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Front. Hum. Neurosci.* 7:547. doi: 10.3389/fnhum.2013.00547

Limanowski, J., Lutti, A., and Blankenburg, F. (2014). The extrastriate body area is involved in illusory limb ownership. *Neuroimage* 86, 514–524. doi: 10.1016/j.neuroimage.2013.10.035

Longo, M. R., Schüür, F., Kammers, M. P., Tsakiris, M., and Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition* 107, 978–998. doi: 10.1016/j.cognition.2007.12.004

Makin, T. R., Holmes, N. P., and Zohary, E. (2007). Is that near my hand? Multisensory representation of peripersonal space in human intra-parietal sulcus. *J. Neurosci.* 27, 731–740. doi: 10.1523/JNEUROSCI.3653-06.2007

Peelen, M. V., and Downing, P. E. (2007). The neural basis of visual body perception. *Nat. Rev. Neurosci.* 8, 636–648. doi: 10.1038/nrn2195

Pitcher, D., Charles, L., Devlin, J. T., Walsh, V., and Duchaine, B. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Curr. Biol.* 19, 319–324. doi: 10.1016/j.cub.2009.01.007

Rohde, M., Di Luca, M., and Ernst, M. O. (2011). The rubber hand illusion: feeling of ownership and proprioceptive drift do not go hand in hand. *PLoS ONE* 6:e21659. doi: 10.1371/journal.pone.0021659

Rossi, S., Hallett, M., Rossini, P. M., and Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin. Neurophysiol.* 120, 2008–2039. doi: 10.1016/j.clinph.2009.08.016

Saxe, R., and Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia* 43, 1391–1399. doi: 10.1016/j.neuropsychologia.2005.02.013

Schutter, D. J., and van Honk, J. (2006). A standardized motor threshold estimation procedure for transcranial magnetic stimulation research. *J. ECT* 22, 176–178. doi: 10.1097/01.yct.0000235924.60364.27

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

Tsakiris, M., Costantini, M., and Haggard, P. (2008). The role of the right temporo-parietal junction in maintaining a coherent sense of one's body. *Neuropsychologia* 46, 3014–3018. doi: 10.1016/j.neuropsychologia.2008.06.004

Tsakiris, M., and Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 80–91. doi: 10.1037/0096-1523.31.1.80

Urgesi, C., Berlucchi, G., and Aglioti, S. M. (2004). Magnetic stimulation of extrastriate body area impairs visual processing of nonfacial body parts. *Curr. Biol.* 14, 2130–2134. doi: 10.1016/j.cub.2004.11.031

Urgesi, C., Candidi, M., Ionta, S., and Aglioti, S. M. (2007). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nat. Neurosci.* 10, 30–31. doi: 10.1038/nn1815

Wiggett, A. J., and Downing, P. E. (2011). Representation of action in occipito-temporal cortex. *J. Cogn. Neurosci.* 23, 1765–1780. doi: 10.1162/jocn.2010.21552

Paper 4

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

Minimal self-models and the free energy principle

Jakub Limanowski[1]* and Felix Blankenburg[1,2,3]

[1] Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany
[2] Dahlem Institute for Neuroimaging of Emotion, Freie Universität Berlin, Berlin, Germany
[3] Center for Adaptive Rationality (ARC), Max Planck Institute for Human Development, Berlin, Germany

The term "minimal phenomenal selfhood" (MPS) describes the basic, pre-reflective experience of being a self (Blanke and Metzinger, 2009). Theoretical accounts of the minimal self have long recognized the importance and the ambivalence of the body as both part of the physical world, and the enabling condition for being in this world (Gallagher, 2005a; Grafton, 2009). A recent account of MPS (Metzinger, 2004a) centers on the consideration that minimal selfhood emerges as the result of basic self-modeling mechanisms, thereby being founded on pre-reflective bodily processes. The free energy principle (FEP; Friston, 2010) is a novel unified theory of cortical function built upon the imperative that self-organizing systems entail hierarchical generative models of the causes of their sensory input, which are optimized by minimizing free energy as an approximation of the log-likelihood of the model. The implementation of the FEP via predictive coding mechanisms and in particular the active inference principle emphasizes the role of embodiment for predictive self-modeling, which has been appreciated in recent publications. In this review, we provide an overview of these conceptions and illustrate thereby the potential power of the FEP in explaining the mechanisms underlying minimal selfhood and its key constituents, multisensory integration, interoception, agency, perspective, and the experience of mineness. We conclude that the conceptualization of MPS can be well mapped onto a hierarchical generative model furnished by the FEP and may constitute the basis for higher-level, cognitive forms of self-referral, as well as the understanding of other minds.

Keywords: free energy principle, predictive coding, active inference, self, minimal phenomenal selfhood, ownership, agency, self-model

INTRODUCTION

What lets an organism be a self? Throughout philosophical attempts to understand the enabling conditions of minimal self-awareness (Zahavi, 1999), or *minimal phenomenal selfhood* (MPS)[1] (Blanke and Metzinger, 2009), the special status of the body among all other physical things has long been apparent (Merleau-Ponty, 1962; Bermúdez et al., 1998; Anderson and Perlis, 2005; Legrand, 2006; Blanke, 2012). Recently, the role of the human body for cognition has been re-emphasized in the field of embodied cognition (Varela et al., 1994; Clark, 1999; Gallagher, 2005a; Grafton, 2009; Gallese and Sinigaglia, 2011).

The body lets us interact with the world via perception and action (Legrand, 2006; Friston, 2011; Farmer and Tsakiris, 2012), leading to a whole new form of intelligence that is different from, for example, mere computation (Frith, 2007; Grafton, 2009). One's everyday experience is enabled and structured through a body that is "always there" (James, 1890), and hence the body—*my* body—is not just part of the physical world, but also the "vehicle" that enables being a self in this world (Merleau-Ponty, 1962; Varela et al., 1994; Gallagher, 2005a). Minimal, or pre-reflective selfhood emerges from this experience of a unified, situated living body as a "sensorimotor unity anchored to its world" (Bermúdez et al., 1998; Anderson and Perlis, 2005; Gallagher, 2005a; Legrand, 2006; Hohwy, 2010; Blanke, 2012; Apps and Tsakiris, 2013).

In this review, we will particularly consider an account of the mechanisms giving rise to minimal selfhood that has recently been proposed by Metzinger (2003, 2004a,b, 2005). Central to the theory is the premise that minimal selfhood emerges as the result of pre-reflective self-modeling, i.e., through an organism's model of the world that is phenomenologically centered onto the self. Thereby, Metzinger's account builds on the proposition that the brain is a representational system that needs to interpret the

---

[1]In general, this approach is concerned with "global aspects of bodily self-consciousness" (Blanke and Metzinger, 2009), where a *global* property is something that can only be ascribed to a system as a whole, and *self-consciousness* refers to "the ability to become aware of one's own mental and bodily states ... as one's own mental and bodily states" (Vogeley and Fink, 2003). The kind of self-consciousness meant here is not cognitive but "immediate, *pre-reflective* and non-observational" (see also Zahavi, 1999; Gallagher, 2005a; Legrand, 2006; Hohwy, 2007), where the term pre-reflective is referring to levels of self-awareness that are independent of explicit cognition and linguistic abilities (Blanke and Metzinger, 2009). In its simplest form, this is the *minimal phenomenal self*, the "fundamental conscious experience of being someone" (Blanke and Metzinger, 2009).

world (Gallese and Metzinger, 2003), and thus constructs and simulates a model in order to reduce ambiguity originating from the external world (Metzinger, 2005). For this system-model to be successful, i.e., of adaptive value, "the self needs to be embedded into the causal network of the physical world" (Knoblich et al., 2003; Metzinger, 2004a, 2005). The model thus also has to include as part of itself the physical body—"the part of the simulation that represents the system itself" (Edelman, 2008, p. 419). Metzinger (2004a) emphasizes that this self-representation of the system is special in that it (i.e., the body) is the only representational structure that constantly generates and receives internal input via its different intero- and proprioceptive systems. Notably, a resulting structural property of the system-model is the spatiotemporal centeredness of the model onto a coherent phenomenal subject, described by Metzinger with the term *perspectivalness* (Metzinger, 2004a, 2005; Blanke and Metzinger, 2009). Throughout this review, we will return to this, and propose to understand it as an instance of "perspective taking", whereby the brain assigns the subjective, first-person perspective (1PP) to its self-model.

Following their emphasis of self-modeling mechanisms for minimal selfhood, Metzinger and colleagues (Knoblich et al., 2003) have argued that an analysis of selfhood should focus on the underlying *functional* properties of the system, i.e., the brain. In this review, we will examine one promising candidate brain theory for this analysis: over the last years, a general theoretical account of cortical function based on the "free energy principle" (FEP) has been put forth by Friston (Friston et al., 2006; Friston, 2009, 2010; Clark, 2013), based on the conclusive assumption that the brain entails hierarchical dynamical models to predict the causes of its sensory data (Hohwy, 2007; Frith, 2007; Friston and Kiebel, 2009; Bubic et al., 2010).

The key premise of the FEP is that self-organizing organisms have to resist the natural tendency to disorder that is implied by the second law of thermodynamics, i.e., they have to "maintain their states and form in the face of a constantly changing environment" (Friston, 2010). Organisms do so by avoiding *surprise* associated with their sensory states (Friston et al., 2011, 2012; Friston, 2012a,b), which in turn will result in a (desired) state where the world is highly predictable. The FEP proposes that the brain infers the hidden causes of the environment via the inversion of hierarchical generative models that predict their sensory consequences (Friston, 2010; Bastos et al., 2012), with higher levels encoding increasingly abstract and information-integrating conceptions of the world (Fotopoulou, 2012; Clark, 2013). Importantly, as biological organisms are embodied in the environment, the "world-model" of a self-organizing system also has to include the sensory apparatus (the body) of the organism (Friston, 2012b; Friston et al., 2012; Clark, 2013). In agreement with the Good Regulator theorem (Conant and Ashby, 1970; Edelman, 2008; Friston et al., 2012), which states that every good regulator of a system will ultimately become a model of that system, the FEP thus proposes as a consequence of hierarchical predictive modeling that "I model myself as existing" (Friston, 2011, 2013b). We will later highlight that this conforms nicely to accounts of minimal selfhood, whereby the self is perceived as a result of dynamic self-modeling mechanisms (Metzinger, 2004a; Hohwy, 2007).

Conceptually, the FEP is based on the evaluation of the improbability of some sensory data under a hierarchical generative model, where the (model-conditional) improbability of the data is commonly referred to as *surprise* (Friston et al., 2006; Friston, 2010, 2011). The theory builds on *free energy* as an information-theoretical quantity on the upper bound of surprise that can be formally assessed (Friston et al., 2006, 2012; Friston, 2010, 2011). By minimizing free energy within a model, biological agents thus always also minimize surprise. In principle, this can be done in two ways: By changing the *predictions of the model* by means of perception, or by changing *what is predicted* by selectively sampling those sensations that confirm the model's predictions by means of action (a "systematic bias in input sampling", Verschure et al., 2003; Friston, 2011).

Minimizing surprise associated with sensory data by the inversion of the hierarchical generative model (and the dynamic optimization of its parameters) has been established as *predictive coding* (Srinivasan et al., 1982; Mumford, 1992; Rao and Ballard, 1999; Friston, 2005a; Friston and Stephan, 2007; Kilner et al., 2007; Friston and Kiebel, 2009). Thereby, the predictive coding scheme infers the hidden causes of its sensory input by minimizing the difference between the predictions about sensory data and the actual sensory data at any level of the model's hierarchy, which is encoded by the *prediction error* (Friston and Kiebel, 2009; Bubic et al., 2010; Friston, 2010; Brown and Brüne, 2012; Friston, 2012a). Thus the *feedforward* signal is not the sensory information *per se*, but the associated prediction error that is passed up the hierarchy (Hohwy, 2012; Clark, 2013), while the generative model's predictions are the *feedback* signal (Friston, 2010; Bastos et al., 2012; Edwards et al., 2012). The second form of prediction error minimization via interaction with the environment is described under the *active inference* principle (Friston, 2012a, 2013a). Reminiscent of "affordances", Gibson's (1977) famous description of the fact that the environment is "co-perceived" depending on the perceiver's bodily endowment, active inference thus emphasizes the bi-directional role of embodiment such that "not only does the agent embody the environment but the environment embodies the agent" (Friston, 2011). Interestingly, the computational assumptions of predictive coding are surprisingly well reflected by neuroanatomical organization of the cortex (Bastos et al., 2012; Friston, 2012a), suggesting that neuronal populations indeed encode probabilities, i.e., uncertainty (Clark, 2013). In sum, predictive coding and active inference are neurobiologically plausible, "action-oriented" (Bastos et al., 2012; Clark, 2013) implementations of free energy minimization (Friston, 2011; Bastos et al., 2012; Friston, 2012a; Clark, 2013).

In this review, we summarize recently formulated free energy accounts of key aspects of minimal selfhood: multisensory integration, interoception, agency, ownership or "mineness" of experience, the perspectivity of self-models and models of other selves. Common to these FEP applications is the focus on "self modeling" (Friston, 2012a). We hence consider these approaches in the light of the proposal that the minimal self is the result of an ongoing predictive process within a generative model that is centered onto the organism (Metzinger, 2004a; Hohwy, 2007; Friston, 2011).

ASPECTS OF THE MINIMAL SELF IN THE FREE ENERGY FRAMEWORK

A number of publications have recently put forward the idea that (minimal) selfhood is based on the neurobiological implementation of hierarchical generative models in the brain (Hohwy, 2007, 2010; Seth et al., 2011; Fotopoulou, 2012; Friston, 2012a,b; Apps and Tsakiris, 2013; Clark, 2013). In one sentence, these accounts propose to "understand the elusive sense of minimal self in terms of having internal models that successfully predict or match the sensory consequences of our own movement, our intentions in action, and our sensory input" (Hohwy, 2007). In accordance with Friston (2011, 2012b, 2013b), who has already emphasized the fundamental, bi-directional role of embodiment in the FEP, these accounts also embrace the body as a central part of the self-model. The aspects of the minimal self that these approaches formalize in the FEP all follow as consequences from this embodied self-modeling (Metzinger, 2004a; Hohwy, 2007; Friston, 2011): The body predicts and integrates multisensory information in a way that no other physical object does (Hohwy, 2007, 2010; Apps and Tsakiris, 2013), it is the only source of internally generated input (Seth et al., 2011; Critchley and Seth, 2012), it is crucial for interaction with the environment and a sense of agency (Kilner et al., 2007; Frith, 2007; Friston et al., 2011). From the phenomenological and spatiotemporal centeredness of experience onto the body (Friston, 2011) emerges the 1PP, and ultimately, the "mineness" of experience (Hohwy, 2007; Apps and Tsakiris, 2013).

MULTISENSORY INTEGRATION

A very important implication of the free energy framework is that sensory information is processed probabilistically, and thus it follows that the representation of the self is also probabilistic (Friston, 2011). This conceptualization fits comfortably with Metzinger's (2004b) theory, where the content of the self-model is probabilistic, i.e., it is "simply the best hypothesis about the current state of the system, given all constraints and information resources currently available" (see also Hohwy, 2010; Clark, 2013; Friston, 2013b). However, sensory information is not *per se* specific to the self, which implies that there must be additional levels of information processing in which information is related to the self (Apps and Tsakiris, 2013).

Previous accounts of bodily self-awareness, inspired by work on illusions of body ownership and related paradigms, have emphasized the role of multimodal, hierarchical cortical networks in processing self-related information (Hohwy, 2007, 2010; Tsakiris, 2010; Petkova et al., 2011a; Blanke, 2012). In a recent paper, Apps and Tsakiris (2013) propose that hierarchical prediction error minimization can explain processes of self-recognition and self-representation: for the processing of information relating to the self, free energy minimization happens via the integration of various streams of surprise from unimodal sensory information in hierarchically higher multimodal areas, where information from any system can be used to "explain away" surprise in any other system (Hohwy, 2010; Apps and Tsakiris, 2013; Clark, 2013). This corresponds to the basic claim of predictive coding about crossmodal information processing, according to which hierarchically higher levels form amodal concepts that

generate multimodal predictions and prediction errors (Friston, 2012a). Following this logic, higher-level multisensory areas must predict input in multiple sensory modalities, which according to Apps and Tsakiris (2013) implies "a high level representation (of self) that elaborates descending predictions to multiple unimodal systems" (see also Clark, 2013; Friston, 2013b). This self-model can thus be seen as the most accurate, immediately available explanation of the bottom-up surprise from incoming multisensory information (Apps and Tsakiris, 2013; thereby the model need not be "true", just a *sufficient* explanation of the sensory input, Schwabe and Blanke, 2008; Hohwy and Paton, 2010; Hohwy, 2012). The predictive coding account suggests that, at the hierarchically highest level, such a self-model will encode, as model evidence, the evidence for the existence of the agent in the present form (Hohwy, 2010; Friston, 2011).

A particularly intriguing example of how self-representation is constructed in a probabilistic way is the rubber hand illusion (RHI; Botvinick and Cohen, 1998): observing a dummy hand being touched, while receiving synchronous tactile stimulation at the anatomically congruent location of one's real, hidden hand typically leads to an illusory experience of feeling the touch on the dummy hand (Botvinick and Cohen, 1998; Ehrsson et al., 2004, 2005; Makin et al., 2008). This usually results in a self-attribution, or "incorporation" (Holmes and Spence, 2004) of the fake hand as a part of one's own body (Tsakiris and Haggard, 2005; Hohwy and Paton, 2010; Tsakiris, 2010; Petkova et al., 2011a). A number of behavioral measures such as a fear response to the dummy hand being threatened (Armel and Ramachandran, 2003; Ehrsson et al., 2007), or the mislocalization of one's real hand towards the location where the dummy hand is seen (Botvinick and Cohen, 1998; Tsakiris and Haggard, 2005), suggest that the brain indeed seems to treat the dummy hand as part of the body as a result of the multisensory stimulation (see Tsakiris, 2010, or Blanke, 2012, for detailed reviews). Using virtual reality techniques, the RHI paradigm has been extended to induce an illusory self-identification with a whole dummy body located at a different position in space (Ehrsson, 2007; Lenggenhager et al., 2007). In those cases, participants exhibited a bias in judging their own spatial location towards the location where the dummy body was positioned in space, just as the mislocalization of the own hand during the RHI (see Blanke, 2012, for a review). These findings thus impressively demonstrate that perceived self-location can be manipulated with appropriate stimulation.

Generally, illusory percepts are well explained as a result of Bayes-optimal inference, i.e., arising from an interpretation of ambiguous sensory input under strong prior hypotheses (Friston, 2005b; Brown and Friston, 2012; Apps and Tsakiris, 2013; Clark, 2013). Correspondingly, a combination of bottom-up input and modulatory top-down factors has been suggested to drive illusory ownership of body parts as experienced during the RHI (de Vignemont et al., 2005; Tsakiris and Haggard, 2005; de Preester and Tsakiris, 2009; Hohwy and Paton, 2010; Tsakiris, 2010). While congruent multisensory input seems crucial for the RHI (Botvinick and Cohen, 1998; Armel and Ramachandran, 2003; Ehrsson et al., 2004, 2005; Hohwy and Paton, 2010; Petkova et al., 2011a), there have been strong arguments for top-down "body representations" that define which objects

(namely, only anatomically plausible hand-shaped objects, see e.g., Tsakiris and Haggard, 2005) can be incorporated during the RHI (de Vignemont et al., 2005; IJsselsteijn et al., 2006; Costantini and Haggard, 2007; Tsakiris et al., 2007; de Preester and Tsakiris, 2009). However, various inconsistent definitions of body representations may have lead to some confusion and thus prevented the emergence of a unifying theoretical account (de Vignemont, 2007; Longo et al., 2008; Apps and Tsakiris, 2013).

As a solution to this problem, several authors have endorsed a predictive coding approach (Hohwy, 2007, 2010; Apps and Tsakiris, 2013). Consider that, under normal circumstances, observed touch on our skin is accompanied by a corresponding, temporally congruent tactile sensation—in predictive coding terms, the underlying generative model of our physical self predicts a somatosensory sensation when touch is about to occur on the body, because associations between events that have a high probability of predicting events in another system lead to the formation of beliefs, or priors on a hierarchically higher level (Apps and Tsakiris, 2013). Note that it are not *per se* the associations between different kinds of sensory input that are of importance here, but the parallel predictions of the generative model. Among all physical objects in the world, it is only our body that will evoke (i.e., predicts) this kind of multisensory sensation—congruence of multisensory input has (not surprisingly) been called "self-specifying" (Botvinick, 2004) and has been ascribed a crucial role in self-representation (Botvinick and Cohen, 1998; Armel and Ramachandran, 2003; Ehrsson et al., 2005; Hohwy and Paton, 2010). Following this logic, during the RHI, surprise[2] or prediction error is evoked by the simultaneous occurrence of observed touch on an external object (the dummy hand) together with a somatosensory sensation, because such congruence is not predicted by the brain's initial generative model.

The predictive coding account suggests that, as stimuli can usually be caused "in an infinite number of ways" (Brown and Friston, 2012), there are several competing explanations of the sensory input between which the brain needs to decide. In the case of the RHI, these are coded by the probabilities of the actual hand, or the dummy hand being "me" (Apps and Tsakiris, 2013). One explanation, or model, of the sensory input is that vision and touch occur at different locations (the "true" model, Hohwy, 2010). However, during the RHI, spatially distributed observed and felt touch are "bound together" by causal inference (Hohwy, 2012): this "false" model (that observed and felt touch occur at the same location, namely, one's own hand) is selected because it more successfully explains the incoming prediction error in favor of a unified self (see also Schwabe and Blanke, 2008; Hohwy, 2010; Hohwy and Paton, 2010). This is a crucial point, because

predictive coding is a "winner takes all" strategy (Hohwy, 2007, 2010): there is always one model that has the lowest amount of free energy (the highest model evidence) among all possible models of the sensory input (Friston et al., 2012; Apps and Tsakiris, 2013; Clark, 2013), and this model is selected as the explanation for the world. This model does not have to be "true", just a better explanation of the sensory input than competing models (Friston et al., 2012). As minimizing surprise is the same as maximizing model-evidence (where model-evidence is evidence for the agent's existence), the agent, or self, in its present form will cease to exist if another model has to be chosen as a better explanation of sensory input (Hohwy, 2010; Friston, 2011): "I" (i.e., the embodied model of the world) will only exist "iff (sic) I am a veridical model of my environment" (Friston, 2011).

Applied to the RHI example, this means that if prediction error could not be explained away in this way, the system might have to dismiss its current self-model in favor of a better explanation of the input—which would result in the representation of a "disunified self" (Hohwy, 2010). The FEP states that, if prediction error can be explained away at lower levels, there is no need to adjust higher-level representations (Friston, 2012a). Apps and Tsakiris (2013) propose that, as the prediction error is passed up the hierarchy during the RHI, it can be explained away at multimodal cortical nodes. Thereby "explaining away" means an updating of the generative model's predictions about the physical features of the self to minimize the overall level of surprise in the system. This results in a different posterior probabilistic representation of certain *features* of the self (Hohwy and Paton, 2010; Apps and Tsakiris, 2013), however, without any necessity to change the actual generative self-*model* (Hohwy, 2010). Specifically, the dummy hand is now probabilistically more likely to be represented as part of one's body, which in turn is accompanied by a decrease in the probability that one's actual hand will be represented as "self". This manifests as a self-attribution of the dummy hand, and a partial rejection of the real limb (de Preester and Tsakiris, 2009; Tsakiris, 2010).

Indeed, there is compelling experimental evidence in support of such a probabilistic integration process underlying the RHI. For example, the mislocalization of one's real hand towards the location of the dummy hand is never absolute, but relative; participants usually judge the location of their hand several centimeters closer to the dummy, but not at the same location (Tsakiris and Haggard, 2005). Lloyd (2007) showed that the RHI gradually decreases with increasing distance between the own and the dummy hand. Furthermore, a drop in skin temperature of the stimulated real hand was found to accompany the RHI (Moseley et al., 2008), which has been interpreted as evidence for top-down regulations of autonomic control and interoceptive prediction error minimization during the RHI (Moseley et al., 2008; Seth et al., 2011; Suzuki et al., 2013). Also, after the illusion, the dummy hand is frequently perceived as more similar to one's real hand (Longo et al., 2009). These findings suggest that in fact, explaining away prediction error from ambiguous multisensory stimulation may lead to changes in the encoded features of the self (Hohwy and Paton, 2010).

The idea of a probabilistic self-representation in the brain benefits from the fact that the free energy account is relatively

---

[2]Although the illusory experience of feeling the touch on the dummy hand is certainly surprising, one has to distinguish this cognitive surprise of the agent from "surprise" on a neurobiological level ("surprisal", see Clark, 2013), as defined by prediction error. In fact, here these two notions may be somewhat opposing: the dummy hand is accepted as a part of one's body as a result of successfully *explaining away* the surprise evoked by the ambiguous multisensory stimulation (Hohwy, 2010; Hohwy and Paton, 2010). However, the agent experiences exactly this state—owning a lifeless dummy hand—as surprising.

unconstrained and thus not as heavily dependent on conceptual assumptions as other theories (Hohwy, 2007, 2010; Friston, 2008; Friston and Kiebel, 2009; Friston et al., 2012). Thus the FEP does not need to treat information relating to the self as a distinct class of information (Apps and Tsakiris, 2013), because it is concerned with information flow and system structure. For example, the matching of sensory predictions based on corollary discharge with actual sensory input has been previously proposed as a basis for self-awareness (see Gallagher, 2000; Brown et al., 2013). In the free energy account, however, self-awareness is not restricted to the integration of sensorimotor efference and re-afference. Rather, *any* type of sensory information can be integrated within a multimodal, abstract representation of the self, and explain away surprise in another system (Apps and Tsakiris, 2013). The RHI example demonstrates that, as claimed by the FEP (Friston, 2012a), if prediction error can be explained away in the periphery (e.g., adjusting the encoded location of one's real hand), there is no need to adjust higher-level representations (the unified self-model). The FEP is thus a parsimonious, and hence inherently flexible, formal description of how multisensory information integration underpins minimal forms of self-awareness (Hohwy, 2010; Blanke, 2012).

INTEROCEPTION

A special case of information that the self-model receives is input from interoceptive senses: within the world-model, the (own) body is special among all physical objects in that it constantly receives a "background buzz" of somatosensory input, including input from somato-visceral and mechanoreceptors, and higher-level feeling states (Metzinger, 2004a, 2005; see Friston, 2011). Acknowledging the importance of interoception, recent work by Seth (Critchley and Seth, 2012; Seth et al., 2011; Suzuki et al., 2013) has promoted interoceptive prediction error minimization as a mechanism for self-representation. Specifically, Seth et al. provide a predictive coding account of "presence", where presence means the subjective experience of being in the here and now (see Metzinger, 2004a). Presence is hence a structural property of conscious experience (Seth, 2009) that is transparent in the sense that Metzinger (2003) uses the term (Seth et al., 2011). According to Seth et al. (2011), interoceptive predictions arise from autonomic control signals and sensory inputs evoked by motor control signals. The generative model of the causes of interoceptive input gives rise to "interoceptive self-representations" and "emotional feeling states" (Suzuki et al., 2013). Presence results as the successful suppression of the associated prediction error (Seth et al., 2011), more specifically, "self-consciousness is grounded on the feeling states that emerge from interaction of interoceptive predictions and prediction errors" (Critchley and Seth, 2012). The emphasis on subjective feeling states (Critchley et al., 2004; Seth et al., 2011) as a key component of interoceptive predictive coding links this account to emotion frameworks like the somatic marker hypothesis (Damasio, 1999; Bechara et al., 2000).

Half a century ago, Schachter and Singer (1962) showed that people seek explanations for their bodily sensations after having become aware of them. Reversing this argument, Pennebaker and Skelton (1981) showed that the perception of bodily sensations

depended on the hypotheses held by the participants, and was thus not different from the processing of any other ambiguous information. More recently, Moseley et al. (2008) found that the RHI led to a cooling of participants' real hand (and only the hand affected by the illusion), and concluded that there is a causal link between self-awareness and homeostatic regulation, where bodily self-awareness regulates physiological processing in a top-down manner. In accordance with these results, the FEP indicates that interoceptive predictions are "one—among many—of multimodal predictions that emanate from high-level hypotheses about our embodied state." (Friston, 2013b; Suzuki et al., 2013). Interestingly, as we will see later (see *Modeling Others*), these predictions can also be used to model others' internal states (Bernhardt and Singer, 2012). In sum, although predictive coding accounts of interoception still need detailed work, the corresponding emphasis of interoceptive signals by predictive coding (Seth et al., 2011) and philosophical (Metzinger, 2004a) accounts of the self promises many insightful studies to come.

ACTION AND AGENCY

Agency as a "sense of initiative" (Edelman, 2008) has been emphasized as a key component of MPS (Gallagher, 2000; Metzinger, 2004a; Frith, 2007). Distinguishing between self-initiated actions and actions of other organisms is crucial for being a self. The importance of the motor system in the brain's ontology (interpretation) of the world (Gallese and Metzinger, 2003) has been promoted by forward models of agency based on corollary discharge (Blakemore et al., 2002; Gallagher, 2005a; Frith, 2012), which have also been applied to describe disturbances of agency resulting from a failure of these mechanisms (Gallagher, 2000). Advancing on these accounts, action and the phenomenology of agency have both been accounted for in terms of hierarchical generative models (Hohwy, 2007).

The active inference principle is of central importance in the FEP (Friston and Stephan, 2007; Hohwy, 2007, 2010; Kilner et al., 2007; Brown et al., 2013; Friston, 2013a): action changes the sensory input of an organism so that it better corresponds to the current generative model, without having to revise the model parameters (Friston and Stephan, 2007; Hohwy, 2010). This validation of the current generative system-model is a confirmation of the agent's existence (Friston, 2011). However, for active inference to be feasible, the agent has to be able to predict which actions will lead to a better confirmation of its predictions. Friston (2012b) thus states that "implicit in a model of sampling is a representation or *sense of agency*", since the effects of selective sampling of sensations as through active inference have to be known—modeled—as well. Thus, by selectively sampling sensations so that they confirm the model's predictions, action is a form of "reality testing" (Hohwy, 2007). For instance, consider that the induction of illusory limb or body ownership via multisensory stimulation (like in the RHI) only works because this kind of active inference is suppressed.[3] If allowed, participants would

---

[3]But, as pointed out by Hohwy (2007, 2010), active inference is still happening at a more subtle level, as participants focus their attention on the rubber hand to detect potential mismatches of observed and felt touch.

probably instantaneously move their hand to *test* whether the rubber hand moves as well. The illusion will be immediately abolished once participants see that the rubber hand does not move according to their intentions (IJsselsteijn et al., 2006; Slater et al., 2009; Maselli and Slater, 2013), because now there is a clear mismatch between predicted and actual sensory outcome, which cannot be explained away.

It is noteworthy that failures in basic inference mechanisms are a likely cause of many symptoms connected to a disturbed sense of agency (Gallagher, 2000; Frith, 2007). As stated by the FEP, probabilistic inference under uncertainty underlies all perception, and it thus seems reasonable to explain abnormal experiences in the same framework (Fletcher and Frith, 2008; Hohwy, 2013). Predictive coding schemes and Bayesian inference have been successfully applied to explain symptoms like delusion formation (Fletcher and Frith, 2008; Hohwy, 2013) or failures in sensory attenuation occurring in schizophrenia (Brown et al., 2013), hysteria or functional symptoms (Edwards et al., 2012), out-of-body experiences (Schwabe and Blanke, 2008), and depersonalization (Seth et al., 2011). In many of these cases, basic mechanisms of active inference fail (Brown et al., 2013), but it is not yet clear whether these symptoms can be explained by failures at low levels alone, or rather by a failure of mechanisms across the hierarchy (Fletcher and Frith, 2008). For instance, a noisy prediction error signal has been suggested as the cause for positive symptoms in schizophrenia (Fletcher and Frith, 2008), while delusions are seen as the result of false inference "at a conceptual level" (Brown et al., 2013), which may be characterized by a "lack of independent sources of evidence for reality testing" (Hohwy, 2013).

In conclusion, action and agency are of fundamental importance for the experience of normal minimal selfhood. However, although a sense of agency (Gallagher, 2000) is sufficient for MPS, it may not be the most basal constituent (Blanke and Metzinger, 2009). What matters is that I experience the action as *mine* (Gallagher, 2000), which brings us to the most important aspect of the generative self-model: the experience of "mineness" (Hohwy, 2007).

MINENESS

The phenomenal experience of "mineness" is a key property of MPS (Metzinger, 2004a). The idea that the living body is experienced as mine ("owned") can be traced back to early phenomenologists like Merleau-Ponty or Husserl (see Gallagher, 1986, 2009). It has been claimed that this "self-ownership" (Gallagher, 2000) is the most fundamental sense of phenomenal selfhood (Aspell et al., 2009; Blanke and Metzinger, 2009). Similarly, Hohwy (2007) equates experienced mineness of actions and perceptions with the experience of a minimal self.

In Hohwy's (2007) FEP account of the self, mineness is a general phenomenon, resulting from successful predictions of actions and perceptions. It is hereby important to keep in mind that prediction is more than mere anticipation (Hohwy, 2007; Bubic et al., 2010), but describes predictive *modeling* as a fundamental principle of the brain, and that what is informative in predictive coding is the prediction *error*. Following Hohwy's (2007) logic, phenomenal selfhood thus arises as a consequence of successfully

having predicted incoming sensory input across the hierarchy of the self-model. Within predictive coding, prediction error is not explained away post-hoc, but constantly, and across all levels of the model (Friston, 2012a). Thus mineness is always *implicit* in the flow of information within the hierarchical generative self-model, and can correspondingly be experienced for actions and perceptions in the same way (note how once again the FEP is simple in its assumptions). Crucially, this means that the minimal self is the result of an ongoing, dynamic process, not a static representation. In this account, mineness is thus situated in a spatio*temporal* reference frame (see Metzinger, 2004a; Hohwy, 2007), where prediction introduces the temporal component of "being already familiar" with the predicted input (Hohwy, 2007; see Kiebel et al., 2008; Bubic et al., 2010).

Perhaps a good example for this construction of temporally extended phenomenal experience from predictive processes is the classical concept of a *body schema* (Head and Holmes, 1911–1912; Merleau-Ponty, 1962). The body schema describes the dynamic organization of sensorimotor processes subserving motor and postural functions in a form of "embodied memory" that ultimately presents the body for action (Gallagher, 2009). These processes are pre-reflective, operating "below the level of self-referential intentionality" (Gallagher and Cole, 1995), and thus the body schema is not a static representation (Gallagher, 2005a). But note that the body schema defines the range of possible actions that my body can perform, while being "charged" with what has happened before (see Gallagher, 2009, for a nice review). In the hierarchical generative self-model, the body schema might thus be pictured as encoded by a structure of predictions (e.g., of self-location and proprioception).

In conclusion, the following picture seems to emerge from the reviewed literature: the FEP is capable of describing the functional regularities of the brain's "ontology" (Gallese and Metzinger, 2003), such as the prediction and integration of intero- and exteroceptive signals (Hohwy, 2010; Seth et al., 2011; Apps and Tsakiris, 2013), the importance of action and agency (Gallagher, 2000; Hohwy, 2007; Friston, 2012a), and the mineness of experience (Hohwy, 2007, 2010). In agreement with the Good Regulator theorem (Conant and Ashby, 1970; Edelman, 2008; Friston et al., 2012), which states that every good regulator of a system will ultimately become a model of that system, both the FEP and the philosophical account of minimal selfhood agree that the agent *is* the current embodied model of the world (Metzinger, 2004a; Hohwy, 2007; Friston, 2011).

THE PERSPECTIVITY OF THE SELF-MODEL

In accordance with the FEP, the phenomenal self-model (PSM) theory views selves as processes, not objects. Accordingly, the self is perceived *because* systems with a PSM constantly assume, or model, their own existence as a coherent entity (Metzinger, 2004a; Blanke and Metzinger, 2009). However, to assume that there is a perceiver is a fallacy ("no such things as selves exist in the world", Metzinger, 2005). Rather, a conscious self is a result of the system's identification with its self-model ("you *are* the content of your PSM", Metzinger, 2005).

This self-identification is possible because the "attentional unavailability of earlier processing stages in the brain for intro-

spection" (Metzinger, 2003, 2005) leads to a gradually increasing *transparency* of higher-level phenomenal states. Transparency thus describes the fact that only the contents of phenomenal states, not their underlying mechanisms, are introspectively accessible to the subject of experience (Metzinger, 2003, 2004a). Interestingly, it has been proposed that the cognitive impenetrability of predictive coding mechanisms can be explained by the fact that hierarchically higher levels predict on longer timescales, and more abstractly than lower levels (Hohwy, 2007, 2010; Kiebel et al., 2008). Failures in these mechanisms may result in severe symptoms that seem to be related to a loss of global experiential selfhood, as demonstrated by certain disorders of "presence" such as depersonalization disorder (Seth et al., 2011). These phenomena might also be described by a loss of transparency ("if . . . the self-model of a conscious system would become fully opaque, then the phenomenal target property of experiential "selfhood" would disappear", Metzinger, 2004b).

Thus, the crucial implication of transparency is that the PSM "cannot be recognized as a model by the system using it" (Metzinger, 2004a), which greatly reduces computational load within the system by efficiently avoiding an infinite regression that would otherwise arise from the logical structure of self-modeling (Metzinger, 2004a, 2005): "I can never conceive of what it is like to be me, because that would require the number of recursions I can physically entertain, plus one" (Friston et al., 2012). Similarly, the FEP states that systems operating with a self-model will have an advantage because "a unified self-model is what best allows computation of the system's current state such that action can be undertaken" (Hohwy, 2010; see Friston et al., 2012, for a discussion).

Note how, by the transparent spatiotemporal centeredness of the model onto the self (Metzinger, 2003, 2004a; see also Hohwy, 2007; Friston, 2011, 2012b), the model takes on a 1PP (Vogeley and Fink, 2003). However, the centeredness of the model is *phenomenal*, and not just (but also) geometrical (a temporal centering on the subject happens through successful prediction, see previous section). This is well reflected by Blanke and Metzinger (2009), who distinguish between the phenomenally distinct *weak 1PP*, and *strong 1PP*: The weak 1PP means a purely geometric centering of the experiential space upon one's body, and thus corresponds most to the "egocentre" (Roelofs, 1959; Merker, 2007) or "cyclopean eye" (von Helmholtz, 1962), which can be traced back to Hering's (1942) projective geometry. Experimental work on extending the RHI paradigm has shown that the strength of illusory self-identification with a dummy or virtual body crucially depends on this kind of 1PP (Petkova and Ehrsson, 2008; Petkova et al., 2011b; Maselli and Slater, 2013), and that in addition to proprioceptive information, vestibular information is crucial for determining self-location in space (Schwabe and Blanke, 2008; Blanke, 2012).

As an attempt to summarize the reviewed accounts of the basic constituents of MPS, **Figure 1** shows a schematic depiction of a hierarchical generative model, predicting from the *minimal phenomenal self* to increasingly specific, unimodal lower levels on shorter timescales (Kiebel et al., 2008; Hohwy, 2010; Clark, 2013). For simplicity, we have only included one intermediate level in the hierarchy, consisting of the basic aspects of minimal selfhood as

discussed in the reviewed articles (see Figure caption for a detailed description).

In the generative self-model (**Figure 1**), the first-person perspective (1PP) node should be taken as a purely geometrical point of convergence of sensory information from a particular sensory modality (a "weak 1PP"), whereas the phenomenal centeredness of the model onto the experiencing subject would correspond to a "strong 1PP" (Blanke and Metzinger, 2009). Note that although the weak 1PP and self-location usually coincide, these two phenomena can be decoupled in neurological patients with autoscopic phenomena, while MPS still seems to be normal in these conditions (Blanke and Metzinger, 2009; Blanke, 2012). This seems to speak for a probabilistic processing of minimal selfhood, and also for a relative independence of 1PP and self-location (which are therefore also modeled as separate nodes on the intermediate level of the generative model in **Figure 1**).

In conclusion, the experienced 1PP presents itself as a key feature of "mineness", and thus as a basic constituent of, and a prerequisite for a minimal self (Gallagher, 2000; Vogeley and Fink, 2003; Metzinger, 2004a; Blanke and Metzinger, 2009). Some authors speak of a system's "ability" to take the 1PP, meaning the ability to integrate and represent experience, i.e., mental states, in a common egocentric reference frame centered upon the body (Vogeley and Fink, 2003). The FEP very comfortably complies with the assumption that a body model "defines a volume within a spatial frame of reference . . . within which the origin of the weak 1PP is localized" (Blanke and Metzinger, 2009; Friston, 2011, 2012b). In this light, we now review the explanatory power of the FEP for mechanisms of modeling other agents.

MODELING OTHERS

In opposition to the 1PP, the third-person perspective (3PP) is the perspective of the observer, i.e., the perspective that is taken when states are ascribed to someone else (Vogeley and Fink, 2003; Blanke and Metzinger, 2009; Fuchs, 2012). This form of perspective taking is of essential importance, for how we make sense of ourselves in a social environment depends on the representation of, and distinction between, actions and states of the self and those of others (Decety and Sommerville, 2003; Frith, 2007; Bernhardt and Singer, 2012; Farmer and Tsakiris, 2012; Frith and Frith, 2012). Traditionally, at least two distinct mechanisms have been postulated to underlie our understanding of other's internal states: *experience sharing* and *mentalizing* (Brown and Brüne, 2012; Zaki and Ochsner, 2012). While experience sharing refers to a mere mirroring of others' action intentions, sensations, or emotions (Gallese and Sinigaglia, 2011), the term mentalizing describes explicitly reflecting others' internal states: in a recent review, Zaki and Ochsner (2012) define the mechanism behind mentalizing as "the ability to represent states outside of a perceiver's 'here and now'", thus having both a spatial 1PP and a temporal (present versus past and future) aspect. Crucially, this involves a representation of other agents as possessing a 1PP that differs from one's own (Farmer and Tsakiris, 2012). One can also describe these processes as simulating other PSMs (Metzinger, 2004a); in this way, a pre-reflective, phenomenally transparent self-model is necessary for the formation of higher-level cognitive
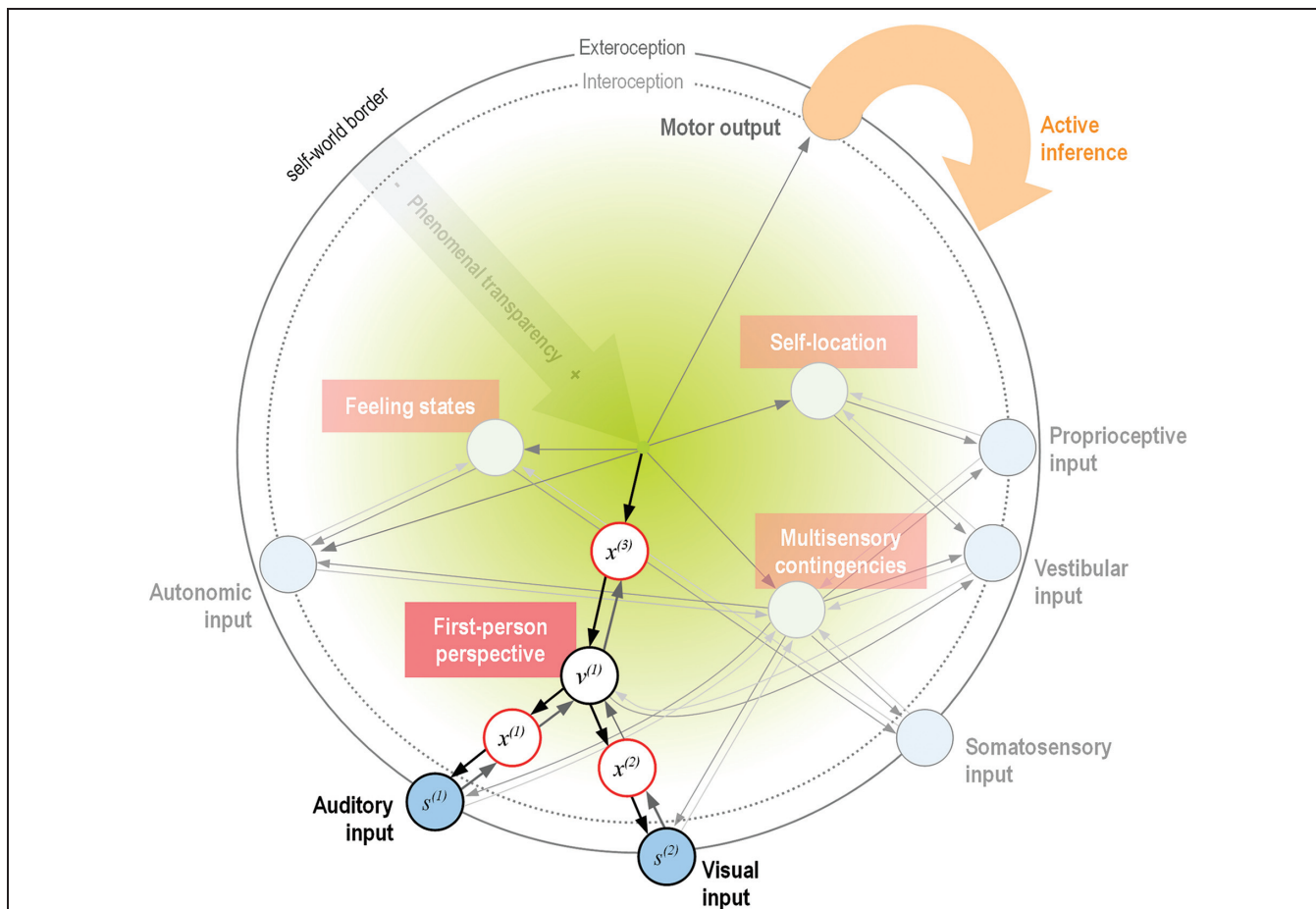
FIGURE 1 | Schematic proposal for a mapping of the *phenomenal self-model* onto a hierarchical generative model (format adapted from Bastos et al., 2012). Shown here is only the system's model of itself, where representational nodes at each level generate descending predictions to increasingly specialized lower levels (symbolized by darker arrows). In this special case, the single modeled cause of sensations is the *minimal phenomenal self* (Metzinger, 2004a), which generates predictions about the state of one or many sensory modalities (blue circles). The inversion of this generative model (a predictive coding scheme, lighter arrows) infers hidden causes—and thus ultimately, the self as the single cause—of sensory input via minimization of prediction error (Friston, 2011). For simplicity, only one intermediate level of nodes within the hierarchy is displayed, consisting of the basic properties of minimal selfhood as reviewed (white circles). As a (simplified) illustration of the hierarchical generative processing, the case of the 1PP is highlighted. Here, descending predictions of the unified self-model (black arrows) generate sensory data $s^{(i)}$ in the respective modalities (auditory and visual). This happens via a hierarchy of hidden states $x^{(i)}$ and hidden causes $v^{(i)}$ (the 1PP), which generate predictions about data in the level below. The green gradient symbolizes increasing transparency of the accompanying phenomenal states with ascending hierarchy, where the final cause (the self) is completely transparent. Note that at this (highest) level, there is no further representational node; this acknowledges the fact that the perception of a unified minimal self is the result of a temporally extended predictive process, not a static representation (Metzinger, 2004a; Hohwy, 2007). The experience of "mineness" of the self (and of perception and action in general, Hohwy, 2007) is a result of the model's successful predictions and thus implicitly symbolized by the arrows. Input into this system-model comes from intero- and exteroception (blue circles), while active inference is a means of changing predicted input in all modalities through interaction with the environment. As the model-evidence is evidence for the agent's existence (Friston, 2011, 2013b), the model will necessarily be a veridical model of the agent: if there was too much unexplained prediction error, the model would be abandoned in favor of a model with a higher evidence; the self in the present form would cease to exist (Hohwy, 2010; Friston, 2011, 2012b).

and social mental concepts (Metzinger, 2003, 2004a, 2005; Edelman, 2008; Blanke and Metzinger, 2009).

Humans display first instances of experience sharing almost from birth onwards (Tomasello et al., 2005), for example, human infants as young as one hour after birth can already imitate facial gestures (Meltzoff and Moore, 1983). It hence seems that an "experiential connection" between self and others is already present in newborn infants (Gallagher and Meltzoff, 1996; Fuchs, 2012). Another example for such a pre-reflective self-other con-nection is sensorimotor mirroring ("neural resonance", Zaki and Ochsner, 2012). Many studies have reported vicarious activations of the motor system by observing others' actions (Rizzolatti and Craighero, 2004), or likewise of the somatosensory system by the observation of touch (Keysers et al., 2010) or pain to others (Bernhardt and Singer, 2012). These findings suggest a very basic, automatic activation of one's representations to another person's action intentions, or experience (Keysers et al., 2010; Zaki and Ochsner, 2012). There have been arguments for a link between

sensory mirroring mechanisms and higher-level perspective taking abilities (see Preston and de Waal, 2002, for a discussion), suggesting that although such vicarious responses are activated automatically, they are not purely sensory-driven (Singer and Lamm, 2009).

The FEP emphasizes models of the behavior and intentions of others as a crucial determinant of our own behavior (Frith, 2007; Friston, 2012a). It has accordingly been proposed that mechanisms of social cognition are based on predictive coding as well (Baker et al., 2011; Brown and Brüne, 2012; Frith and Frith, 2012), where perspective taking can be described as forming "second order representations" (Friston, 2013b). In other words, as agents, we also have to predict the behavior of other agents, by not only generating a model of the physical world (and our body) but also of the mental world-models of our conspecifics based on their behavior (Frith, 2007; Frith and Frith, 2012). Crucially, we have to continually update our models of others' mental states via prediction errors, because these states are not stable but vary over time (Frith and Frith, 2012). This task is far from trivial, and involves many levels of differential self-other modeling ranging from a purely spatial differentiation (other agents occupy different positions in the world) to the abstract modeling of other minds like in Theory of Mind (Vogeley and Fink, 2003; Baker et al., 2011).

Several recent accounts have proposed that associative learning updated through prediction errors is a common computational mechanism underlying both reward learning and social learning (Behrens et al., 2008; Hampton et al., 2008; Frith and Frith, 2012). Experimental evidence from these studies suggests that prediction errors code for false predictions about others' mental states (Behrens et al., 2008; Hampton et al., 2008), and even for discrepancies between predictions of others and actual outcome of their choice (Apps et al., 2013). Interestingly, it seems that even low-level predictions can also be updated interactively. For example, dyads of individuals with similar perceptual sensitivity may benefit from interactive decision-making, as shown by an increased performance in a collective perceptual decision task during which levels of confidence were communicated (Bahrami et al., 2010). As mentioned before, if these basic predictive mechanisms fail, pathological behavior can emerge (Fletcher and Frith, 2008; Brown et al., 2013). For example, perspective taking abilities seem to be often impaired in individuals suffering from Autism Spectrum Disorder (ASD; Oberman and Ramachandran, 2007; but cf. Hamilton et al., 2007), while there is also evidence for impaired predictive coding mechanisms in ASD (Friston, 2012a).

An intriguing question is whether the brain uses the same models to generate predictions about own and other behavior. In a predictive coding account of action understanding, Kilner and colleagues (Kilner et al., 2007; Friston et al., 2011) have argued that the *mirror neuron system* is part of a generative model predicting the sensory consequences of actions, and that indeed, it seems that the brain applies the same model to predict one's own, and others' actions. Actions are thereby modeled on four hierarchical levels (Hamilton and Grafton, 2008): intentions, goals, kinematics, and muscles. By inversion of the model, the brain can thus infer the causes of own and others' actions, via explaining away prediction error across these four levels. Thus the mirror neuron system is active during action observation because the "own" generative model is inverted to infer the intention underlying the observed action. A similar argument is made by Gallese and Sinigaglia (2011) (see also Goldman and de Vignemont, 2009) to explain embodied simulation in general by the fact that representations of states of the self and others' states have the same bodily format, and thus the same constraints. Correspondingly, there is evidence that the same neuronal structures may be involved in predicting own and others' internal states (Bernhardt and Singer, 2012), for example, in predicting how pain will feel for others (Singer et al., 2004). In sum, there is strong evidence that others' mental states are inferred via internal models. It seems that the use of generative models by the brain can explain many of these basic, as well as more elaborated social mechanisms. Thereby (at least partially) common predictive mechanisms for self and others strongly support the notion of perspective taking as an "embodied cognitive process" (Kessler and Thomson, 2010). This is a relatively young, but promising field of research; it is up to future studies to evaluate the explanatory power of the FEP in this domain.

CONCLUSION

In this review, we have summarized proposals from different authors, all emphasizing the concept of hierarchical generative models to explain processes underlying the bodily foundations of MPS, including its fundamental constituents such as multisensory integration, the sense of agency, the experience of mineness, perspectivity, and its phenomenal transparency. We have reviewed these free energy accounts of key aspects of minimal selfhood in the light of the premise that the self is the result of a generative process of self-modeling (Metzinger, 2004a; Hohwy, 2007). The approaches reviewed here show that the FEP complies with the claim that minimal selfhood emerges from physiological processes (Gallagher, 1986, 2000; Zahavi, 1999; Legrand, 2006; Blanke and Metzinger, 2009), and acknowledges both the phenomenal and spatiotemporal centeredness of the generative self-model as a key for minimal self-awareness. Albeit still schematic, these accounts demonstrate that the predictive coding account can inform theoretical and experimental approaches towards the normal and pathological self. The FEP is increasingly gaining influence as a "deeply unified account of perception, cognition, and action" (Friston, 2010; Hohwy, 2010; Apps and Tsakiris, 2013; Clark, 2013), up to recent accounts proposing it as a general mechanism underlying evolution and the "emergence of life" itself (Friston, 2013c). A particular strength of the approach seems to be that it makes relatively few conceptual assumptions (Hohwy, 2007, 2010; Friston, 2008; Friston and Kiebel, 2009; Friston et al., 2012), thus being capable of formalizing both spatial and social aspects of self-models. Of course, there are many outstanding issues, and the free energy formulation will have to withstand thorough empirical testing (for discussions, Friston et al., 2012; Apps and Tsakiris, 2013; see Clark, 2013). While it is well-established in the domains of action and perception, future work will have to show whether the FEP can be similarly influential in cognitive and social domains. Particularly, the social domain lacks models (Frith and Frith, 2012), and currently the FEP seems one of the most promising candidate theories to formally describing the mechanisms

underlying the experience of being a "self in relation to others" (Frith, 2007; Friston, 2012a). The FEP may thus provide a framework to address philosophical debates about self-modeling (Gallagher, 2005b; cf. Metzinger, 2006), and perhaps help to bridge gaps between neuroscientific and philosophical approaches to the self.

REFERENCES

Anderson, M. L., and Perlis, D. R. (2005). The roots of self-awareness. *Phenomenol. Cogn. Sci.* 4, 297–333. doi: 10.1007/s11097-005-4068-0

Apps, M. A., and Tsakiris, M. (2013). The free-energy self: a predictive coding account of self-recognition. *Neurosci. Biobehav. Rev.* doi: 10.1016/j.neubiorev.2013.01.029. [Epub ahead of print].

Apps, M. A., Green, R., and Ramnani, N. (2013). Reinforcement learning signals in the anterior cingulate cortex code for others' false beliefs. *Neuroimage* 64, 1–9. doi: 10.1016/j.neuroimage.2012.09.010

Armel, K. C., and Ramachandran, V. S. (2003). Projecting sensations to external objects: evidence from skin conductance response. *Proc. Biol. Sci.* 270, 1499–1506. doi: 10.1098/rspb.2003.2364

Aspell, J. E., Lenggenhager, B., and Blanke, O. (2009). Keeping in touch with one's self: multisensory mechanisms of self-consciousness. *PLoS One* 4:e6488. doi: 10.1371/journal.pone.0006488

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., and Frith, C. D. (2010). Optimally interacting minds. *Science* 329, 1081–1085. doi: 10.1126/science.1185718

Baker, C. L., Saxe, R. R., and Tenenbaum, J. B. (2011). "Bayesian theory of mind: modeling joint belief-desire attribution," in *Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society*, Boston, 2469–2474.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

Bechara, A., Damasio, H., and Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cereb. Cortex* 10, 295–307. doi: 10.1093/cercor/10.3.295

Behrens, T. E., Hunt, L. T., Woolrich, M. W., and Rushworth, M. F. (2008). Associative learning of social value. *Nature* 456, 245–249. doi: 10.1038/nature07538

Bermúdez, J. L., Marcel, A. J., and Eilan, N. (Eds.). (1998). *The Body and the Self.* Cambridge: MIT Press.

Bernhardt, B. C., and Singer, T. (2012). The neural basis of empathy. *Annu. Rev. Neurosci.* 35, 1–23. doi: 10.1146/annurev-neuro-062111-150536

Blakemore, S. J., Wolpert, D. M., and Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends Cogn. Sci.* 6, 237–242. doi: 10.1016/S1364-6613(02)01907-1

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nat. Rev. Neurosci.* 13, 556–571. doi: 10.1038/nrn3292

Blanke, O., and Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends Cogn. Sci.* 13, 7–13. doi: 10.1016/j.tics.2008.10.003

Botvinick, M. (2004). Probing the neural basis of body ownership. *Science* 305, 782–783. doi: 10.1126/science.1101836

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

Brown, E. C., and Brüne, M. (2012). The role of prediction in social neuroscience. *Front. Hum. Neurosci.* 6:147. doi: 10.3389/fnhum.2012.00147

Brown, H., Adams, R. A., Parees, I., Edwards, M., and Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cogn. Process.* doi: 10.1007/s10339-013-0571-3. [Epub ahead of print].

Brown, H., and Friston, K. J. (2012). Free-energy and illusions: the corn-sweet effect. *Front. Psychol.* 3:43. doi: 10.3389/fpsyg.2012.00043

Bubic, A., Von Cramon, D. Y., and Schubotz, R. I. (2010). Prediction, cognition and the brain. *Front. Hum. Neurosci.* 4:25. doi: 10.3389/fnhum.2010.00025

Clark, A. (1999). An embodied cognitive science? *Trends Cogn. Sci.* 3, 345–351. doi: 10.1016/s1364-6613(99)01361-3

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/s0140525x12000477

Conant, R. C., and Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *Int. J. Syst. Sci.* 1, 89–97. doi: 10.1080/00207727008920220

Costantini, M., and Haggard, P. (2007). The rubber hand illusion: sensitivity and reference frame for body ownership. *Conscious. Cogn.* 16, 229–240. doi: 10.1016/j.concog.2007.01.001

Critchley, H., and Seth, A. (2012). Will studies of macaque insula reveal the neural mechanisms of self-awareness? *Neuron* 74, 423–426. doi: 10.1016/j.neuron.2012.04.012

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., and Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nat. Neurosci.* 7, 189–195. doi: 10.1038/nn1176

Damasio, A. (1999). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness.* London: Vintage Books.

de Preester, H., and Tsakiris, M. (2009). Body-extension versus body-incorporation: is there a need for a body-model? *Phenomenol. Cogn. Sci.* 8, 307–319. doi: 10.1007/s11097-009-9121-y

de Vignemont, F. (2007). How many representations of the body? *Behav. Brain Sci.* 30, 204–205. doi: 10.1017/s0140525x07001434

de Vignemont, F., Tsakiris, M., and Haggard, P. (2005). "Body mereology," in *Human Body Perception from the Inside Out*, eds G. Knoblich, I. M. Thornton, M. Grosjean and M. Shiffrar (New York: Oxford University Press), 147–170.

Decety, J., and Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends Cogn. Sci.* 7, 527–533. doi: 10.1016/j.tics.2003.10.004

Edelman, S. (2008). *Computing the Mind*. New York: Oxford University Press.

Edwards, M. J., Adams, R. A., Brown, H., Pareés, I., and Friston, K. J. (2012). A Bayesian account of 'hysteria'. *Brain* 135, 3495–3512. doi: 10.1093/brain/aws129

Ehrsson, H. H., Spence, C., and Passingham, R. E. (2004). That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science* 305, 875–877. doi: 10.1126/science.1097011

Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science* 317, 1048. doi: 10.1126/science.1142175

Ehrsson, H. H., Holmes, N. P., and Passingham, R. E. (2005). Touching a rubber hand: feeling of body ownership is associated with activity in multisensory brain areas. *J. Neurosci.* 25, 10564–10573. doi: 10.1523/jneurosci.0800-05.2005

Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J., and Passingham, R. E. (2007). Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. *Proc. Natl. Acad. Sci. U S A* 104, 9828–9833. doi: 10.1073/pnas.0610011104

Farmer, H., and Tsakiris, M. (2012). The bodily social self: a link between phenomenal and narrative selfhood. *Rev. Philos. Psychol.* 3, 125–144. doi: 10.1007/s13164-012-0092-5

Fletcher, P. C., and Frith, C. D. (2008). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. doi: 10.1038/nrn2536

Fotopoulou, A. (2012). "Towards psychodynamic neuroscience," in *From the Couch to the Lab: Trends in Psychodynamic Neuroscience*, eds A. Fotopoulou, M. Conway and D. Pfaff (New York: Oxford University Press), 25–48.

Friston, K. (2005a). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622

Friston, K. J. (2005b). Hallucinations and perceptual inference. *Behav. Brain Sci.* 28, 764–766. doi: 10.1017/s0140525x05290131

Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol.* 4:e1000211. doi: 10.1371/journal.pcbi.1000211

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301. doi: 10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

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.

Friston, K. (2012a). Prediction, perception and agency. *Int. J. Psychophysiol.* 83, 248–252. doi: 10.1016/j.ijpsycho.2011.11.014

Friston, K. (2012b). Embodied inference and spatial cognition. *Cogn. Process.* 13(Suppl. 1), S171–S177. doi: 10.1007/s10339-012-0519-z

Friston, K. (2013a). Active inference and free energy. *Behav. Brain Sci.* 36, 212–213. doi: 10.1017/S0140525X12002142

Friston, K. (2013b). The fantastic organ. *Brain* 136, 1328–1332. doi: 10.1093/brain/awt038

Friston, K. (2013c). Life as we know it. *J. R. Soc. Interface* 10, 1–12. doi: 10.1098/rsif.2013.0475

Friston, K. J., and Stephan, K. E. (2007). Free-energy and the brain. *Synthese* 159, 417–458. doi: 10.1007/s11229-007-9237-y

Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300

Friston, K., Kilner, J., and Harrison, L. (2006). A free energy principle for the brain. *J. Physiol. Paris* 100, 70–87. doi: 10.1016/j.jphysparis.2006.10.001

Friston, K., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biol. Cybern.* 104, 137–160. doi: 10.1007/s00422-011-0424-z

Friston, K., Thornton, C., and Clark, A. (2012). Free-energy minimization and the dark-room problem. *Front. Psychol.* 3:130. doi: 10.3389/fpsyg.2012.00130

Frith, C. (2007). *Making Up the Mind: How the Brain Creates Our Mental World.* Oxford, UK: Blackwell.

Frith, C. (2012). Explaining delusions of control: the comparator model 20 years on. *Conscious. Cogn.* 21, 52–54. doi: 10.1016/j.concog.2011.06.010

Frith, C. D., and Frith, U. (2012). Mechanisms of social cognition. *Annu. Rev. Psychol.* 63, 287–313. doi: 10.1146/annurev-psych-120710-100449

Fuchs, T. (2012). The phenomenology and development of social perspectives. *Phenomenol. Cogn. Sci.* 1–29. doi: 10.1007/s11097-012-9267-x

Gallagher, S. (1986). Lived body and environment. *Res. Phenomenol.* 16, 139–170. doi: 10.1163/156916486x00103

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* 4, 14–21. doi: 10.1016/s1364-6613(99)01417-5

Gallagher, S. (2005a). *How the Body Shapes the Mind.* Oxford: Clarendon Press.

Gallagher, S. (2005b). Metzinger's matrix: living the virtual life with a real body. *Psyche* 11, 1–9.

Gallagher, S. (2009). "Philosophical antecedents to situated cognition," in *Cambridge Handbook of Situated Cognition*, eds P. Robbins and M. Aydede (Cambridge: Cambridge University Press), 35–51.

Gallagher, S., and Cole, J. (1995). Body schema and body image in a deafferented subject. *J. Mind Behav.* 16, 369–390.

Gallagher, S., and Meltzoff, A. N. (1996). The earliest sense of self and others: Merleau-Ponty and recent developmental studies. *Philos. Psychol.* 9, 211–233. doi: 10.1080/09515089608573181

Gallese, V., and Metzinger, T. (2003). Motor ontology: the representational reality of goals, actions and selves. *Philos. Psychol.* 16, 365–388. doi: 10.1080/0951508032000121760

Gallese, V., and Sinigaglia, C. (2011). What is so special about embodied simulation? *Trends Cogn. Sci.* 15, 512–519. doi: 10.1016/j.tics.2011.09.003

Gibson, J. J. (1977). "The theory of affordances," in *Perceiving, Acting and Knowing*, eds R. Shaw, and J. Bransford (Hillsdale: Erlbaum), 67–82.

Goldman, A., and de Vignemont, F. (2009). Is social cognition embodied? *Trends Cogn. Sci.* 13, 154–159. doi: 10.1016/j.tics.2009.01.007

Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Ann. N Y Acad. Sci.* 1156, 97–117. doi: 10.1111/j.1749-6632.2009.04425.x

Hamilton, A. F. D. C., Brindley, R. M., and Frith, U. (2007). Imitation and action understanding in autistic spectrum disorders: how valid is the hypothesis of a deficit in the mirror neuron system? *Neuropsychologia* 45, 1859–1868. doi: 10.1016/j.neuropsychologia.2006.11.022

Hamilton, A. F., and Grafton, S. T. (2008). "The motor hierarchy: from kinematics to goals and intentions," in *Sensorimotor Foundations of Higher Cognition: Attention and Performance xxii*, eds Y. Rosetti, M. Kawato and P. Haggard (New York: Oxford University Press), 381–408.

Hampton, A. N., Bossaerts, P., and O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. U S A* 105, 6741–6746. doi: 10.1073/pnas.0711099105

Head, H., and Holmes, G. (1911–1912). Sensory disturbances from cerebral lesions. *Brain* 34, 102–245. doi: 10.1093/brain/34.2-3.102

Hering, E. (1942). *Spatial Sense and Movements of the Eye* (A. Raddle, Trans.). Baltimore, MD: American Academy of Optometry (original work published 1879).

Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche (Stuttg)* 13, 1–20.

Hohwy, J. (2010). "The hypothesis testing brain: some philosophical applications," in *ASCS09: Proceedings of the 9th Conference of the Australasian Society for Cognitive Science*, Sydney, Australia, 135–144.

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Front. Psychol.* 3:96. doi: 10.3389/fpsyg.2012.00096

Hohwy, J. (2013). Delusions, illusions and inference under uncertainty. *Mind Lang.* 28, 57–71. doi: 10.1111/mila.12008

Hohwy, J., and Paton, B. (2010). Explaining away the body: experiences of supernaturally caused touch and touch on non-hand objects within the rubber hand illusion. *PLoS One* 5:e9416. doi: 10.1371/journal.pone.0009416

Holmes, N. P., and Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cogn. Process.* 5, 94–105. doi: 10.1007/s10339-004-0013-3

IJsselsteijn, A. W., de Kort, Y. A. W., and Haans, A. (2006). Is this my hand i see before me? The rubber hand illusion in reality, virtual reality, and mixed reality. *Presence: Teleop. Virt.* 15, 455–464. doi: 10.1162/pres.15.4.455

James, W. (1890). *The Principles of Psychology.* New York: Dover.

Kessler, K., and Thomson, L. A. (2010). The embodied nature of spatial perspective taking: embodied transformation versus sensorimotor interference. *Cognition* 114, 72–88. doi: 10.1016/j.cognition.2009.08.015

Keysers, C., Kaas, J. H., and Gazzola, V. (2010). Somatosensation in social perception. *Nat. Rev. Neurosci.* 11, 417–428. doi: 10.1038/nrn2833

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4:e1000209. doi: 10.1371/journal.pcbi.1000209

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cogn. Process.* 8, 159–166. doi: 10.1007/s10339-007-0170-2

Knoblich, G., Elsner, B., Aschersleben, G., and Metzinger, T. (2003). Grounding the self in action. *Conscious. Cogn.* 12, 487–494. doi: 10.1016/j.concog.2003.08.009

Legrand, D. (2006). The bodily self: the sensori-motor roots of pre-reflective self-consciousness. *Phenomenol. Cogn. Sci.* 5, 89–118. doi: 10.1007/s11097-005-9015-6

Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 24, 1096–1099. doi: 10.1126/science.1143439

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 Cogn.* 64, 104–109. doi:10.1016/j.bandc.2006.09.013

Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., and Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition* 107, 978–998. doi: 10.1016/j.cognition.2007.12.004

Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., and Haggard, P. (2009). Self awareness and the body image. *Acta Psychol. (Amst)* 132, 166–172. doi: 10.1016/j.actpsy.2009.02.003

Makin, T. R., Holmes, N. P., and Ehrsson, H. H. (2008). On the other hand: dummy hands and peripersonal space. *Behav. Brain Res.* 191, 1–10. doi: 10.1016/j.bbr.2008.02.041

Maselli, A., and Slater, M. (2013). The building blocks of the full body ownership illusion. *Front. Hum. Neurosci.* 7:83. doi: 10.3389/fnhum.2013.00083

Meltzoff, A. N., and Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Dev.* 54, 702–709. doi: 10.2307/1130058

Merker, B. (2007). Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav. Brain Sci.* 30, 63–81. doi: 10.1017/s0140525x07000891

Merleau-Ponty, M. (1962). *Phenomenology of Perception* (C. Smith, Trans.). London: Routledge and Kegan Paul (original work published 1945).

Metzinger, T. (2003). Phenomenal transparency and cognitive self-reference. *Phenomenol. Cogn. Sci.* 2, 353–393. doi: 10.1023/b:phen.0000007366.42918.eb

Metzinger, T. (2004a). *Being No One: The Self-Model Theory of Subjectivity.* Cambridge: MIT Press.

Metzinger, T. (2004b). The subjectivity of subjective experience: a representationalist analysis of the

first-person perspective. *Networks* 3–4, 33–64.

Metzinger, T. (2005). Précis: being no one. *Psyche* 11, 1–35.

Metzinger, T. (2006). Reply to Gallagher: different conceptions of embodiment. *Psyche* 12, 4.

Moseley, G. L., Olthof, N., Venema, A., Don, S., Wijers, M., Gallace, A., et al. (2008). Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. *Proc. Natl. Acad. Sci. U S A* 105, 13169–13173. doi: 10.1073/pnas.0803768105

Mumford, D. (1992). On the computational architecture of the neocortex. *Biol. Cybern.* 66, 241–251. doi: 10.1007/BF00202389

Oberman, L. M., and Ramachandran, V. S. (2007). The simulating social mind: the role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. *Psychol. Bull.* 133, 310. doi: 10.1037/0033-2909.133.2.310

Pennebaker, J. W., and Skelton, J. A. (1981). Selective monitoring of bodily sensations. *J. Pers. Soc. Psychol.* 41, 213–223.

Petkova, V. I., and Ehrsson, H. H. (2008). If i were you: perceptual illusion of body swapping. *PLoS One* 3:e3832. doi: 10.1371/journal.pone.0003832

Petkova, V. I., Björnsdotter, M., Gentile, G., Jonsson, T., Li, T. Q., and Ehrsson, H. H. (2011a). From part-to whole-body ownership in the multisensory brain. *Curr. Biol.* 21, 1118–1122. doi: 10.1016/j.cub.2011.05.022

Petkova, V. I., Khoshnevis, M., and Ehrsson, H. H. (2011b). The perspective matters! Multisensory integration in ego-centric reference frames determines full-body ownership. *Front. Psychol.* 2:35. doi: 10.3389/fpsyg.2011.00035

Preston, S. D., and de Waal, F. (2002). Empathy: its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1–20. doi: 10.1017/s0140525x02000018

Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230

Roelofs, C. O. (1959). Considerations on the visual egocentre. *Acta Psychol. (Amst)* 16, 226–234. doi: 10.1016/0001-6918(59)90096-4

Schachter, S., and Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychol. Rev.* 69, 379–399. doi: 10.1037/h0046234

Schwabe, L., and Blanke, O. (2008). The vestibular component in out-of-body experiences: a computational approach. *Front. Hum. Neurosci.* 2:17. doi: 10.3389/neuro.09.017.2008

Seth, A. (2009). Explanatory correlates of consciousness: theoretical and computational challenges. *Cognit. Comput.* 1, 50–63. doi: 10.1007/s12559-009-9007-x

Seth, A. K., Suzuki, K., and Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Front. Psychol.* 2:395. doi: 10.3389/fpsyg.2011.00395

Singer, T., and Lamm, C. (2009). The social neuroscience of empathy. *Ann. N Y Acad. Sci.* 1156, 81–96. doi: 10.1111/j.1749-6632.2009.04418.x

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162. doi: 10.1126/science.1093535

Slater, M., Perez-Mercos, D., Ehrsson, H. H., and Sanchez-Vives, M. V. (2009). Inducing illusory ownership of a virtual body. *Front. Neurosci.* 3:29. doi: 10.3389/neuro.01.029.2009

Srinivasan, M. V., Laughlin, S. B., and Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond. B Biol. Sci.* 216, 427–459. doi: 10.1098/rspb.1982.0085

Suzuki, K., Garfinkel, S., Critchley, H., and Seth, A. K. (2013). Multisensory integration across interoceptive and exteroceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia* doi: 10.1016/j.neuropsychologia.2013.08.014. (in press).

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–690. doi: 10.1017/s0140525x05000129

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

Tsakiris, M., and Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 80–91. doi: 10.1037/0096-1523.31.1.80

Tsakiris, M., Hesse, M. D., Boy, C., Haggard, P., and Fink, G. R. (2007). Neural signatures of body ownership: a sensory network for bodily self-consciousness. *Cereb. Cortex* 17, 2235–2244. doi: 10.1093/cercor/bhl131

Varela, F., Thompson, E., and Rosch, E. (1994). *The Embodied Mind.* Cambridge, MA: MIT Press.

Verschure, P. F., Voegtlin, T., and Douglas, R. J. (2003). Environmentally mediated synergy between perception and behaviour in mobile robots. *Nature* 425, 620–624. doi: 10.1038/nature02024

Vogeley, K., and Fink, G. R. (2003). Neural correlates of the first-person-perspective. *Trends Cogn. Sci.* 7, 38–42. doi: 10.1016/s1364-6613(02)00003-7

von Helmholtz, H. (1962). *Helmholtz's Treatise on Physiological Optics* (J. P. C. Southall, Trans. and Ed.). New York: Dover Publication (Original work published 1866).

Zahavi, D. (1999). *Self-Awareness and Alterity: A Phenomenological Investigation.* Evanston: Northwestern Univ Press.

Zaki, J., and Ochsner, K. N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nat. Neurosci.* 15, 675–680. doi: 10.1038/nn.3085

Paper 5

Limanowski, J. (2014). What can body ownership illusions tell us about minimal phenomenal

selfhood? *Frontiers in Human Neuroscience, 8,* 946. doi: 10.3389/fnhum.2014.00946

What can body ownership illusions tell us about minimal phenomenal selfhood?

Jakub Limanowski[1,2]*

[1] *Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany*
[2] *Neurocomputation and Neuroimaging Unit, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany*
*Correspondence: jakub.limanowski@fu-berlin.de

Illusions have become an invaluable tool for investigating how the sense of a body as *one's own* is constructed and maintained: During the *rubber hand illusion* (RHI, Botvinick and Cohen, 1998), congruent touch to one's hidden hand and a fake counterpart produces an illusion of feeling the touch on the fake hand and, more strikingly, an experience of the fake hand as part of one's body (Ehrsson et al., 2004; Tsakiris, 2010). The principles of the RHI paradigm have been extended to various body parts (Petkova et al., 2011), including even the face (Tsakiris, 2008; Apps et al., 2013); most notably, the RHI has been induced for the entire body *(full body illusion, FBI)*, producing similar behavioral and neural responses (anxiety responses, ownership of the fake body, misperception of one's physical location; Ehrsson, 2007; Lenggenhager et al., 2007; Maselli and Slater, 2013). Such *body ownership illusions (BOIs)* have generated a substantial amount of research and invaluable insights into the mechanisms of body ownership (Tsakiris, 2010; Blanke, 2012; Moseley et al., 2012). The special importance of these illusions lies in the fact that what they manipulate— the sense of having a body—is one of the enabling conditions of *minimal phenomenal selfhood* (MPS, Gallagher, 2000; Blanke and Metzinger, 2009; Metzinger, 2013a). MPS is defined as the most basic possible kind of self-consciousness or self-awareness (Blanke and Metzinger, 2009; Gallagher and Zahavi, 2010), and investigating its enabling conditions may help us understand what it takes for an organism to have the experience of being a self.

Nevertheless, in this paper I will argue that it is still unclear what exactly the mechanisms revealed by BOIs tell us about MPS, and that this needs to be clarified via a joint effort of phenomenological analysis and formal accounts of self-modeling.

BOIs rest on the induction of some crossmodal conflict (e.g., touch seen on a fake hand but felt on one's real hand), which violates the predictions of one's body-model about the unity of one's body (Hohwy, 2007). This conflict is resolved by the brain by remapping modality-specific body part-centered reference frames onto each other (e.g., proprioception onto vision), so that the multimodal representation of the body and the space surrounding it remains coherent (Holmes and Spence, 2004; Makin et al., 2008; Tsakiris, 2010; Blanke, 2012). Thereby the spatio-temporal and anatomical constraints of BOIs (touch needs to occur simultaneously at corresponding locations and on a congruent body part in an anatomically plausible posture) suggest that multisensory input has to be compatible with a prior representation of the body (Tsakiris and Haggard, 2005; De Preester and Tsakiris, 2009; Tsakiris, 2010; Moseley et al., 2012). The brain seems to make a probabilistic either-or decision based on current sensory input under a prior body model, which during BOIs results in the replacement of the real body part by the "owned" fake body part (Longo et al., 2008; Moseley et al., 2012). And indeed, the neural mechanisms integrating multisensory information during the RHI may be similarly employed for one's real body

parts (Gentile et al., 2013). When the rubber hand is threatened or injured, participants show behavioral, and neural anxiety responses similar as for one's real body part (Ehrsson et al., 2007). BOIs may even affect the regulation of one's physiological states: During illusory ownership, one's real limb's temperature may be downregulated (Moseley et al., 2008), and even the immune system may decrease "protection" of the own limb (Barnsley et al., 2011; Costantini, 2014). In sum, there is compelling evidence that BOIs interfere with the representation of one's body. Upon closer inspection, however, the fact that BOIs isolate "the various components that converge in the holistic experience of our bodies" (Maselli and Slater, 2013) may be a fundamental limitation when it comes to relating them to MPS.

MANIPULATING (FEATURES OF) THE SELF: UNDERSTANDING THE IMPLICATIONS OF BODY OWNERSHIP ILLUSIONS FOR MINIMAL PHENOMENAL SELFHOOD

Phenomenological analysis has emphasized the paradox role of the body in our experience (Legrand, 2010) as both an "objective" body that is a physical thing and thus part of the world, and a "subjective" or "lived" body that is our means of experiencing and interacting with the world (this can be traced back to Husserl, Merleau-Ponty, and arguably Sartre, see Gallagher, 1986; Gallagher and Zahavi, 2010). Crucially, during interaction with the world, we usually do not experience the lived body as a thing enabling this interaction, and it may be this very

"experiential absence" of the body that gives us the feeling of "being there" in the world (Gallagher, 1986; Metzinger, 2004). Such a phenomenological distinction may be of particular relevance for BOIs, which isolate and manipulate, and direct attention to perceptual features of oneself such as visual appearance or physical location. In William James' words, BOIs may interfere with the "me" (i.e., the features that one ascribes to oneself) but not the "I" (the subject of experience) (James, 1890; Christoff et al., 2011). Thus, the paradox role of the body may be part of what makes the interpretation of BOIs and their relation to MPS so difficult: When we speak of body "ownership" in these paradigms, we may only refer to self-identification with the perceived body.

Although MPS can be broken down into distinct features (e.g., self-identification, self-location in space and time, or a first-person perspective), it is still one phenomenal property that "does not have proper parts that could themselves count as a kind or type of self-consciousness" (Metzinger, 2013a, 2004). MPS thus acknowledges the "global and unitary character of self-consciousness" (Blanke and Metzinger, 2009). This unified character of MPS—characterized by the fact that we cannot introspectively access its underlying mechanisms and components—is what lends our experience its phenomenal "realness" (Metzinger, 2004, 2013a). As we have recently argued (Limanowski and Blankenburg, 2013), this MPS conceptualization is elegantly compatible with a theory of cortical information processing based on predictive coding (Friston, 2010): The *free energy principle (FEP)* postulates that the brain implements a hierarchical generative model of the world (including the organism itself) that constantly predicts the causes of its sensory input, and that may be updated if its predictions fail. Due to the model's constant bidirectional and hierarchical information flow with the aim of reducing overall prediction error, this model is spatially, temporally, and phenomenally centered onto the organism itself (Hohwy, 2007, 2013; Friston, 2011; Limanowski and Blankenburg, 2013; Apps and Tsakiris, 2014). The FEP account, albeit mechanistic in nature, thus acknowledges that one's

self-representation is plastic (i.e., probabilistic), hierarchical, and yet unified. It builds upon existing neuropsychological models of body ownership (e.g., Makin et al., 2008; Tsakiris, 2010), and shares key assumptions about being a self with the MPS conceptualization—most importantly, that the phenomenal self is a result of probabilistic self-modeling.

Thereby the MPS and FEP accounts both emphasize that such probabilistic self-modeling is a "risky business" for the phenomenal self: When the current generative model is abandoned in favor of another model that fits the sensory data better, the agent in its present form—encoded as the model evidence of the self-centered world-model—ceases to exist (Friston, 2011). Luckily, this does not need to happen: If prediction error can be explained away at lower levels, there is no need to adjust higher-level predictions, let alone to abandon the current world-model. In the case of BOIs this means that if the probabilistic representation of one's physical features can be updated to eliminate the surprise originating from the ambiguous sensory input, there is no need to abandon the actual self-model (Hohwy, 2013; Limanowski and Blankenburg, 2013; Apps and Tsakiris, 2014). Such an updating of self-representations is demonstrated, for example, by the increased perceived similarity of the "owned" dummy hand (Longo et al., 2009), and even of another face (Tsakiris, 2008; Apps et al., 2013) to oneself, or by the fact that the mere expectation of an upcoming touch on the dummy hand may evoke a BOI (Ferri et al., 2013).

In this light, BOIs can be conceived of as targeting specific inferential mechanism employed by the organism, thus directly confirming the functional architecture suggested by probabilistic models of MPS—but this also opens up some questions. It seems that BOIs primarily affect mechanisms operating at lower levels of self-modeling, because the induced conflict can readily be resolved without inducing panic or pathological conditions, and because the recalibration of one modality onto another relies on relatively basic multisensory mechanisms (Tsakiris, 2010; Blanke, 2012; Gentile et al., 2013). On the other hand, the adjustment of physiological responses (Ehrsson

et al., 2007; Moseley et al., 2008; Barnsley et al., 2011) might imply that BOIs also affect higher levels of the self-model (Seth, 2013). To understand the implications of BOIs for MPS, it is crucial to identify which levels of MPS they affect: When do these illusions merely alter perceptual features of oneself, and when—if at all—do they in fact affect MPS *per se*?

Metzinger (2013a,b) has recently clarified the MPS conceptualization by introducing the term *phenomenal unit of identification* (UI), "the phenomenal property with which we currently *identify*, exactly the form of currently active conscious content that generates the subjective experience of 'I *am* this'" (2013b). As I take it, the UI concept can be used in two ways: It specifies the content with which the self-model identifies—this self-identification leads to the experience of MPS (Metzinger, 2004). Formally, we may think of the UI as the evidence or the "origin" of the current generative model (the "region of maximal invariance," Metzinger, 2013a; see also Friston, 2011; Limanowski and Blankenburg, 2013).

The important point is that the UI concept may help to answer our question: BOIs *do* affect MPS, just not in its entirety. The resolution of BOI-evoked conflicts by changing the representation of certain features of oneself, despite attacking fundamental conditions of MPS, may not be sufficient to change the UI. Even the extension of BOIs to the fully body—although surely a step in the right direction—does not necessarily imply a change in the UI: For example, since FBIs seem to employ the same multisensory mechanisms as the RHI (Petkova et al., 2011; Maselli and Slater, 2013), it could be that what becomes subject to the illusion is actually not the whole body, but just the stimulated part of the body (e.g., the torso in the FBI, Smith, 2010; Metzinger, 2013a; see Tsakiris et al., 2006, for such evidence). More importantly, FBIs may also manipulate only individual features of one's self-representation such as one's visual appearance or physical location.

Interestingly, however, Metzinger (2013a,b) proposes that the UI may change during mind wandering episodes, bodiless dreams, and some forms of out-of-body experiences. The resulting claim that an explicit body representation may actually

not be required for MPS (Metzinger, ibd.) is in line with the proposal that a core self is present even when disconnected from exteroceptive sensory input (e.g., during sleep, Park and Tallon-Baudry, 2014). Such proposals are inspiring, but they also show that much work still needs to be done to understand the exact relation between the body, MPS, and its UI. Probabilistic models following the FEP offer a promising way of formally describing the underlying self-modeling mechanisms, but in addition, new experimental approaches addressing specific levels of MPS are needed. To avoid confusion, these approaches should acknowledge the dual role of the body as both subject and object of experience.

CONCLUSION

BOIs have proven an important tool for understanding MPS. But one should keep in mind that BOIs manipulate individual features of one's self-representation, and that this manipulation of certain dimensions of MPS not necessarily affects the UI. Nevertheless, if situated within a phenomenologically inspired probabilistic model, BOIs can contribute to our understanding of MPS by clarifying its hierarchy, and to further developing the UI concept and its relation to the body. Whether we can develop paradigms that manipulate the UI to actually create the illusion of *being someone else* is a different question, but there are promising new developments that encourage to pursue it.

REFERENCES

Apps, M. A., Tajadura-Jiménez, A., Sereno, M., Blanke, O., and Tsakiris, M. (2013). Plasticity in unimodal and multimodal brain areas reflects multisensory changes in self-face identification. *Cereb. Cortex*. doi: 10.1093/cercor/bht199. [Epub ahead of print].

Apps, M. A., and Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neurosci. Biobehav. Rev.* 41, 85–97. doi: 10.1016/j.neubiorev.2013.01.029

Barnsley, N., McAuley, J. H., Mohan, R., Dey, A., Thomas, P., and Moseley, G. L. (2011). The rubber hand illusion increases histamine reactivity in the real arm. *Curr. Biol.* 21, R945–R946. doi: 10.1016/j.cub.2011.10.039

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nat. Rev. Neurosci.* 13, 556–571. doi: 10.1038/nrn3292

Blanke, O., and Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends Cogn. Sci.* 13, 7–13. doi: 10.1016/j.tics.2008.10.003

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

Christoff, K., Cosmelli, D., Legrand, D., and Thompson, E. (2011). Specifying the self for cognitive neuroscience. *Trends Cogn. Sci.* 15, 104–112. doi: 10.1016/j.tics.2011.01.001

Costantini, M. (2014). Bodily self and immune self: is there a link? *Front. Hum. Neurosci.* 8:138. doi: 10.3389/fnhum.2014.00138

De Preester, H., and Tsakiris, M. (2009). Body-extension versus body-incorporation: is there a need for a body-model? *Phenomenol. Cogn. Sci.* 8, 307–319. doi: 10.1007/s11097-009-9121-y

Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science* 317:1048. doi: 10.1126/science.1142175

Ehrsson, H. H., Spence, C., and Passingham, R. E. (2004). That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science* 305, 875–877. doi: 10.1126/science.1097011

Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J., and Passingham, R. E. (2007). Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9828–9833. doi: 10.1073/pnas.0610011104

Ferri, F., Chiarelli, A. M., Merla, A., Gallese, V., and Costantini, M. (2013). The body beyond the body: expectation of a sensory event is enough to induce ownership over a fake hand. *Proc. R. Soc. B* 280, 20131140. doi: 10.1098/rspb.2013.1140

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K. (2011). "Embodied inference: or 'I think therefore I am, if I am what I think,'" in *The Implications of Embodiment (Cognition and Communication)*, eds W. Tschacher and C. Bergomi (Exeter: Imprint Academic), 89–125.

Gallagher, S. (1986). Lived body and environment. *Res. Phenomenol.* 16, 139–170. doi: 10.1163/156916486X00103

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

Gallagher, S., and Zahavi, D. (2010). "Phenomenological approaches to self-consciousness," in *The Stanford Encyclopedia of Philosophy, Winter 2010 Edn.*, ed E. N. Zalta. Available online at: http://plato.stanford.edu/archives/win2010/entries/self-consciousness-phenomenological

Gentile, G., Guterstam, A., Brozzoli, C., and Ehrsson, H. H. (2013). Disintegration of multisensory signals from the real hand reduces default limb self-attribution: an fMRI study. *J. Neurosci.* 33, 13350–13366. doi: 10.1523/JNEUROSCI.1363-13.2013

Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche* 13, 1–20.

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199682737.001.0001

Holmes, N. P., and Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cogn. Process.* 5, 94–105. doi: 10.1007/s10339-004-0013-3

James, W. (1890). *The Principles of Psychology*. New York, NY: Dover. doi: 10.1037/11059-000

Legrand, D. (2010). "Phenomenological dimensions of bodily self-consciousness," in *Oxford Handbook of the Self*, ed S. Gallagher (Oxford: Oxford University Press), 204–227.

Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 24, 1096–1099. doi: 10.1126/science.1143439

Limanowski, J., and Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Front. Hum. Neurosci.* 7:547. doi: 10.3389/fnhum.2013.00547

Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., and Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition* 107, 978–998. doi: 10.1016/j.cognition.2007.12.004

Longo, M. R., Schüür, F., Kammers, M. P., Tsakiris, M., and Haggard, P. (2009). Self awareness and the body image. *Acta Psychol.* 132, 166–172. doi: 10.1016/j.actpsy.2009.02.003

Makin, T. R., Holmes, N. P., and Ehrsson, H. H. (2008). On the other hand: dummy hands and peripersonal space. *Behav. Brain Res.* 191, 1–10. doi: 10.1016/j.bbr.2008.02.041

Maselli, A., and Slater, M. (2013). The building blocks of the full body ownership illusion. *Front. Hum. Neurosci.* 7:83. doi: 10.3389/fnhum.2013.00083

Metzinger, T. (2004). *Being No One: the Self-Model Theory of Subjectivity*. Cambridge, MA: MIT Press.

Metzinger, T. (2013a). Why are dreams interesting for philosophers? The example of minimal phenomenal selfhood, plus an agenda for future research. *Front. Psychol.* 4:746. doi: 10.3389/fpsyg.2013.00746

Metzinger, T. (2013b). The myth of cognitive agency: subpersonal thinking as a cyclically recurring loss of mental autonomy. *Front. Psychol.* 4:931. doi: 10.3389/fpsyg.2013.00931

Moseley, G. L., Gallace, A., and Spence, C. (2012). Bodily illusions in health and disease: physiological and clinical perspectives and the concept of a cortical "body matrix." *Neurosci. Biobehav. Rev.* 36, 34–46. doi: 10.1016/j.neubiorev.2011.03.013

Moseley, G. L., Olthof, N., Venema, A., Don, S., Wijers, M., Gallace, A., et al. (2008). Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. *Proc. Natl. Acad. Sci.* 105, 13169–13173. doi: 10.1073/pnas.0803768105

Park, H. D., and Tallon-Baudry, C. (2014). The neural subjective frame: from bodily signals to perceptual consciousness. *Philos. Trans. R. Soc. B* 369, 20130208. doi: 10.1098/rstb.2013.0208

Petkova, V. I., Björnsdotter, M., Gentile, G., Jonsson, T., Li, T. Q., and Ehrsson, H. H. (2011). From part-to whole-body ownership in the multisensory brain. *Curr. Biol.* 21, 1118–1122. doi: 10.1016/j.cub.2011.05.022

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* 17, 565–573. doi: 10.1016/j.tics.2013.09.007

Smith, A. J. T. (2010). "Comment: minimal conditions for the simplest form of self-consciousness," in *The Embodied Self: Dimensions, Coherence and Disorders*, eds T. Fuchs, H. Sattel, and P. Henningsen (Stuttgart: Schattauer Verlag), 35–41.

Tsakiris, M. (2008). Looking for myself: current multisensory input alters self-face recognition. *PLoS ONE* 3:e4040. doi: 10.1371/journal.pone.0004040

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

Tsakiris, M., and Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *J. Exp. Psychol. Hum.* *Percept. Perform.* 31, 80–91. doi: 10.1037/0096-1523.31.1.80

Tsakiris, M., Prabhu, G., and Haggard, P. (2006). Having a body versus moving your body: how agency structures body-ownership. *Conscious. Cogn.* 15, 423–432. doi: 10.1016/j.concog.2005.09.004

Supplements

A   Curriculum vitae

For reasons of data protection, the curriculum vitae is not included in the online version.

For reasons of data protection, the curriculum vitae is not included in the online version.

B   Publications

Peer-reviewed publications

Limanowski, J., Blankenburg, F. (2015). Network activity underlying the illusory self-attribution of a dummy arm. *Human Brain Mapping (Epub ahead of print)*.

Limanowski, J. (2014). What can body ownership illusions tell us about minimal phenomenal selfhood? *Frontiers in Human Neuroscience, 8*, 946.

Limanowski, J., Lutti, A., & Blankenburg, F. (2014). The extrastriate body area is involved in illusory limb ownership. *NeuroImage, 86*, 514-524.

Wold, A., **Limanowski, J.**, Walter, H., & Blankenburg, F. (2014). Proprioceptive drift in the rubber hand illusion is intensified following 1 Hz TMS of the left EBA. *Frontiers in Human Neuroscience, 8*, 390.

Limanowski, J., & Blankenburg, F. (2013). Minimal self-models and the free energy principle. *Frontiers in human neuroscience, 7*.

Limanowski, J., & Hecht, H. (2011). Where do we stand on locating the self? *Psychology, 2*(4), 312-317.

Manuscripts under review

Limanowski, J., & Blankenburg, F. That's not quite me! Limb ownership encoding in the human brain. Under review at *Social Cognitive and Affective Neuroscience.*

Conference presentations

Oral presentations

Probabilistic modeling of oneself and others (2014). *Aegina Summer School "Shared Experiences: the boundaries of the social brain", Aegina, Greece.*

Multisensory representation of one's limb in extrastriate cortex (2014). *2nd Mind, Brain & Body-Symposium, Mind-Brain Institute, Berlin, Germany.*

Body ownership: Reasoning within the brain (2013). *2nd Neuro-Philosophical Workshop, Venice International University, San Servolo, Italy.*

Poster presentations

Probabilistic modeling of oneself and others (2014). *Aegina Summer School "Shared Experiences: the boundaries of the social brain", Aegina, Greece.*

Multisensory limb representation in the extrastriate body area (2014). *20th Annual Meeting of the Organization for Human Brain Mapping, Hamburg, Germany.*

Multisensory representation of one's limb in extrastriate cortex (2014). *2nd Mind, Brain & Body-Symposium, Mind-Brain Institute, Berlin, Germany.*

Activity in extrastriate body area reflects the Rubber Hand Illusion in an automated fMRI setup (2013). *Poster presentation. 19th Annual Meeting of the Organization for Human Brain Mapping, Seattle, WA, USA.*

Activity in extrastriate body area reflects limb ownership (2013). *14th UCL Queen Square Symposium, Intitute of Neurology, University College London, London, UK.*

Statement of authorship

I hereby certify that this dissertation has been composed by me and is based on my own work, unless stated otherwise. Ideas and thoughts cited directly or indirectly from other work have been cited accordingly.

_____   _____

Date signed   Signature