Rubber Hand Illusion does not arise from comparisons with internal body model

(Alternative title: Multisensory integration model of Rubber Hand Illusion – v. 2.0)

1. Introduction

In the rubber hand illusion (RHI), participants experience a sense of ownership spread over a fake hand as a result of spatiotemporally congruent stimulation (Botvinick and Cohen, 1998). In a typical study design, participant's own hand is hidden from view and a rubber dummy is placed in front of her. 6-11 seconds after the onset of a spatiotemporally synchronized stimulation on both hands (Ehrsson, 2012), e.g., repeated brush strokes, participants start to experience the touch where they see it and, as a consequence, to feel as if the rubber hand was their own. After discovery of the RHI phenomenon, it has become a fruitful experimental paradigm, harnessed in the studies on the determinants and constraints of the sense of ownership (Tsakiris & Haggard, 2005; Constantini & Haggard, 2007; Lloyd, 2007; van Stralen et al., 2014; Constantini et al., 2016; Tsakiris, Tajadura-Jimenez & Constantini, 2011) and sense of agency (Kalckert & Ehrsson, 2012; 2014), both in healthy participants and patients with psychopathological or neuropsychological conditions (Thakkar, Nichols, McIntosh & Park, 2014; Peled, Pressman, Geva, Modai, 2003; Cascio et al., 2012; Ding et al., 2017).

Notwithstanding intensified research, comprehensive psychological and neurodynamical models of how exactly RHI arises (and in general, what psychological and neuronal mechanisms lie at the bottom of embodiment) have not yet been developed, although some attempts have already been made (Tsakiris, 2010; 2017; Apps & Tsakiris, 2014; Ehrsson, 2012; Blanke, Slater & Serino, 2015; Limanowski & Blankenburg, 2015). Although these models underline different processes and constraints for embodiment, there is a general consensus that it results from dynamic interactions between top-down and bottom-up processes (Azañón et al., 2016; Ratcliffe & Newport, 2017). According to the bottom-up approach, the sense of ownership should be mainly stimulus-driven and simply result from multisensory stimulation complying with the requirements of the laws of multisensory integration, e.g., spatiotemporal matching of the signals. Originally, RHI was described as a bottom-up phenomenon (Botvinick & Cohen, 1998; Armel & Ramachandran, 2003). However, some of the studies suggested that certain top-down processes, such as prior knowledge, expectations, pattern recognition, or contextual information, are involved in the process of incorporating external objects (Tsakiris 2010; Apps & Tsakiris, 2014). In this view, multisensory integration is necessary, but not sufficient, to elicit the illusion, since to-be-incorporated objects have to be highly probable to be taken as a part of one's body, for example because of a physical resemblance or anatomical plausibility.

Internal body model theory (Tsakiris 2010; Apps & Tsakiris, 2014), stressing the relevance of top-down modulations for multisensory integration processes, is an interesting attempt to provide a neurocognitive model for how the subjective sense of ownership comes up. Underlying the importance of appearance of to-be-incorporated dummy, it accounts for the attenuation or abolition of RHI for distorted hands (Ratcliffe & Newport, 2017), 2-D hand-like objects (Tsakiris et al., 2009) or non-hand-like objects (Limanowski & Blankenburg, 2016; Tsakiris & Haggard, 2005; Holmes, Snijders and Spence, 2008; Haans, Ijsselstein, de Kort, 2008; Guterstam, Gentile and Ehrsson, 2013); e.g. neither wooden sheets nor blocks can be incorporated. Consistently with the model, the illusion is also absent when a dummy is placed in an anatomically implausible posture (Ehrsson et al., 2004; Holle et al., 2011). However, "while these observations have been taken to support top-down approaches, they actually do not: dissimilarities between novel object and actual body part are likely to reduce the degree of intersensory matching (the key factor of bottom-up approaches), which renders this factor theoretically nondiagnostic" (Ma & Hommel, 2015a, p.76). In the present article, I will argue that there is no single piece of empirical evidence that unequivocally proves that top-down processes that *do not directly pertain to the properties of stimulation* (such as an internal body model, prior knowledge of anatomy or contextual information) are causally relevant for RHI.

Top-down approaches are widely accepted in the contemporary literature (Tsakiris, 2017; Ratcliffe & Newport, 2017; Blanke, Slater & Serino, 2015). In spite of relatively weak empirical evidence, doubts raised in the literature (e.g. Ma & Hommel, 2015a) are ignored. It may be caused by the fact that the explanation that internal body model theory provides is both simple and common-sense – cognitive systems test external objects for fit with visual representation of one's own body parts to prevent confusing mistakes. On the other hand, top-down explanations are not explanatorily efficient – e.g. introducing comparisons with internal body model as a prerequisite for embodiment (Tsakiris, 2010) makes the model less parsimonious, more challenging for empirical testing, raises doubts about the function of the posited additional component – what would we need it for, given that spatiotemporally congruent stimulations on our hands and other objects in peripersonal space are extremely rare? – and does not comply with existing empirical evidence (see section 3. for a detailed discussion of empirical reports incompatible with internal body model hypothesis).

As an alternative, I develop a multisensory integration model of RHI[1], which is a substantial extension of the model proposed by Ehrsson (2012). In this model, RHI arises from the optimal integration of multisensory cues and succumbs to the general laws of multisensory integration, such as Maximum Likelihood Estimation rule (Ernst & Banks, 2002; Ernst & Bülthoff, 2004). I will also describe the role of the predictive models encoding which signals in different sensory modalities tend to go together (Parise, 2016; van Dam, Parise & Ernst, 2014; Ernst, 2007). In this view, a new perspective on embodiment emerges – perception of one's own body is taken as a regular form of perception, based on the same principles as perception of external multisensory events (Ma & Hommel, 2015a). In

[1] For the sake of clarity, I will focus specifically on RHI, but the model can be generalized to other related phenomena (e.g. full body illusions) and passively induced sense of ownership in general.

particular, the sense of ownership is a natural consequence of a situation in which a cognitive system ascribes a common cause for signals when one of them comes from tactile modality. I will argue that this model – underlining the need for coherence of stimulation rather than resemblance of hand and to-be-embodied object – is more parsimonious and comprehensive.

The structure of the article is as follows. First (section 2.), I will describe the most influential contemporary models of embodiment, focusing on how they underline the importance of top-down processes; in particular, comparisons with an internal body model. In section 3., I will provide three arguments against the internal body model hypothesis. I will appeal to observations from experimental cognitive science that seem to be non-reconcilable with this approach and critically evaluate studies claimed to provide support for this theoretical approach. Then, I will proceed to a multisensory integration model, presenting laws of multisensory integration (section 4.) and proposing how RHI arises in accordance with these laws (section 5.). In the last chapter (section 6.), I will discuss future challenges and pre-register an experiment that would allow us to test precise predictions disambiguating between multisensory integration and internal body model models.

2. Contemporary models of Rubber Hand Illusion

In a neurocognitive model of three critical comparisons, proposed by Tsakiris (2010), multisensory integration processes are preceded by a comparison of visual representation of a to-be-embodied object and a template of a corresponding body-part, stored in an internal body model. The sense of ownership may be spread over objects only if they pass the test of the first critical comparison. The visual form congruency is crucial at this stage (Haans et al., 2008; Tsakiris et al., 2010), but some visual features (e.g. skin color) seem to be irrelevant in the context of the first critical comparison – therefore, internal body model should not be identified with a body image. In case of a match, a second critical comparison evaluating congruency between seen and felt body postures takes place. The illusion is absent for anatomically implausible positions of rubber dummies or discrepancies in seen and felt hand orientations (Ehrsson et al., 2004); however, small discrepancies may be tolerated as long as provided stimulation is congruent in hand-centred reference frame (Constantini & Haggard, 2007). Congruent postures lead to recalibration of visual and tactile coordinates, e.g., change in the location of visuotactile receptive fields to the fake arm (Blanke, Slater & Serino, 2015), as long as both hands are not separated by a large distance (<30cm; Lloyd, 2007). Third critical comparison pertains to congruence of visual and tactile information – seen and felt touches. Spatiotemporally congruent stimulation leads to a subjective sense of ownership.

The model of three comparisons has recently been nuanced within a predictive processing (PP) framework (Apps & Tsakiris, 2014). In the PP approach (Friston, 2005; Clark, 2013), cognitive systems have direct access only to activations in their perceptual subsystems. These sensory signals are sparked by external stimuli (e.g., light hitting the photoreceptors in the retina). To identify the external causes of activations (e.g., objects reflecting the light hitting the receptors), cognitive systems develop – through interaction with the environment, in search of dependencies between behavior and perceptual changes – and continuously test an internal, hierarchical and generative model of the world. The model instantiates predictions which flow in a top-down manner, originating from very general and abstract expectations operating at the slower timescales, constrain more detailed predictions on lower levels of the hierarchical model and determine low-level content operating at the timescale of perception (Seth et al., 2011). In the face of incongruent sensory evidence, discrepancies are propagated up the hierarchy until they are finally resolved, e.g. via adjustment of predictions or optimization of higher-level assumptions of the model. "The idea is that a brain operating this way will come to encode (in the form of predictive or generative models) a rich body of information about the sources of signals by which it is regularly perturbed" (Seth, 2015, p. 5), building increasingly comprehensive and accurate model of the world.

Most importantly, the content of perception is constantly negotiated between sensory evidence and predictions based on prior experience, and perception reflects internally generated hypotheses about the causes of the sensory signals. In case of multisensory experiences, a cognitive system must resolve a correspondence problem and determine whether sensory signals from different modalities share a common cause (Welch & Warren, 1980; Ernst & Bülthoff, 2004). To do so, it exploits both spatiotemporal cues – in particular, spatiotemporal correlations of the signals from different modalities (Parise, Spence & Ernst, 2012) and prior knowledge (van Dam, Parise & Ernst, 2014). According to Apps and Tsakiris (2014), RHI occurs when probability that a rubber hand is one's own hand exceeds the probability of one's own hand being one's own. Note that the former is equivalent in meaning with a situation in which a common cause is ascribed to visual and tactile signals. Given that discrepancies between visual and tactile signals are substantial (since loci of seen and felt touches are outlying), the solution to a correspondence problem largely depends on the prior probability. It is determined by the visual form of the to-be-embodied object and its orientation in space (both of which may be gathered under the term "body-related visual information"; Blanke, Slater & Serino, 2015). This body-related visual information is of particular importance in this context, since the cognitive system ascribes higher reliability to visual rather than tactile or proprioceptive signals, based on the history of their lower variability (Hohwy, 2012; Limanowski & Blankenburg, 2016). Therefore, for body-related visual information matching predictions generated under the hypothesis "that is my hand", the subjective sense of ownership comes up (Apps & Tsakiris, 2014). The PP based model has been recently nuanced by Tsakiris (2017) who stresses the importance of interoception for body ownership in the self-other context.

3. Internal body model hypothesis does not comply with empirical evidence

It may be slightly challenging to specify what kind of empirical data could directly support or count against the internal body model hypothesis, as precise scientifically tractable predictions and falsifiability conditions are rarely specified by its proponents. However, some phenomena which seem to be irreconcilable with internal body model hypotheses may be indicated. In this chapter, I will discuss studies that show that non-hand-like objects and virtual effectors can actually be incorporated and the illusion is not attenuated as compared to hand-shaped objects or virtual hand-like effectors (3.1). In the next subsection, I will show that the use of differently morphed objects or hands placed in anatomically implausible posture necessarily entails elevated sensory mismatch and, as such, cannot support the internal body model hypothesis (3.2; see also Ma & Hommel, 2015a). Finally (3.3), I will critically evaluate the alleged neuroscientific support for the model. In particular, I will show that neuroscientists, discussing their results in favor of internal body model hypothesis (e.g. Limanowski and Blankenburg, 2015; Zeller et al., 2016), succumb to a consistency fallacy (Cole and Klein, 2010; Coltheart, 2013).

3.1. Objects that would not match the representation stored in internal body model can actually be incorporated

Ma & Hommel (2015a) have shown that the sense of ownership may be spread over virtual 2D shapes – balloons and rectangles. In their experiments, using mediated-reality conditions, they displayed a virtual effector on a monitor. The participants were asked to freely perform two kinds of movements: opening/closing of the hand and changing the orientation of the hand. The virtual effector was changing in shape or color in synchrony with participants' movements, e.g. opening of the hand was making it bigger or greener. As a result, participants reported a sense of ownership over a disconnected (separated by a distance) and anatomically implausibly placed (the screen was oriented perpendicularly to the floor and the participant's hand) virtual balloon (exp.1); the strength of the illusion increased after horizontal presentation of a virtual rectangle on a monitor placed closer to the participant's body and connected with the use of a textile covering the space between the participant and the monitor (exp. 2). The strength of the illusion did not differ for virtual rectangles and hands. In their follow-up study, Ma and Hommel (2015b) have shown that active exploration of a mediated-reality environment coupled with sensory feedback on a virtual effector induces the sense of ownership both over rectangles and hands and, even though the illusion was stronger for hand-resembling effector, the visual form did not interact with the synchronicity of stimulation. This finding is in contradiction with the model proposed by Tsakiris (2010), as it suggests that visual resemblance did not further influence multisensory integration processes (the illusion was reported to be stronger for hand-like objects both in synchronous and asynchronous conditions).

On the basis of their findings, Ma and Hommel (2015a; 2015b) propose that connectedness, spatial proximity and multimodal correlations are crucial for the sense of ownership to arise. Active exploration of an environment significantly increases the amount of sensory information revealing multisensory contingencies, strengthening the sense of ownership. Appearance seems to be irrelevant; the very possibility of embodying a 2D effector is in direct contradiction with the model of critical comparisons (Tsakiris, 2010), as the visual form of this shape would have been extremely unlikely to pass a test-for-fit stage. It is also worth noting that the possibility of incorporation of detached and perpendicularly presented 2D virtual effector – even if the strength of the sense of ownership induced was relatively weak – questions the importance of anatomical plausibility as well.

Even a more striking phenomenon – an "invisible hand illusion" (Guterstam, Gentile and Ehrsson, 2013) – arises when congruent spatiotemporal patterns of stimulation are delivered to a participant's real hand and an empty space. As a result, tactile sensations are referred to a volume of empty space and a sense of ownership is induced. Since visual information pertaining to structural properties or anatomical plausibility is lacking, occurrence of this phenomenon seems to question the importance of body-related visual information (or to-be-embodied-object-related visual information) – it is neither a necessary factor (Tsakiris, 2010) nor a constraint (Blanke, Slater and Serino, 2015) for a sense of ownership to come up. However, it does not mean that visual information is irrelevant in general, but rather that it is visual information pertaining to spatiotemporal properties of stimulation that matters, as the invisible hand illusion obeys general rules of multisensory integration – it arises only when 1) spatiotemporal patterns of visual and tactile stimulations are carefully matched in tridimensional space (note that, during the experiment, "a trained experimenter moved the paintbrush (…) following the shape of the knuckles and angles of the finger phalanges, as if it were touching an identical invisible right hand", p. 1080) and in hand-centered reference frame – brushstroke directions should match the position of the real arm 2) "stimulation" of an empty space is confined to peripersonal space. Taken together, these results seem to be irreconcilable with internal body model hypothesis without additional assumptions (e.g. that participants were imagining the real hand in the empty space, which is impossible when the space is occupied by a dissimilar object; Aymerich-Franch and Ganesh, 2016).

Aymerich-Franch and colleagues (2017a), utilizing a virtual reality set-up, showed that the sense of ownership may be spread over robotic arms dissimilar to human hands in terms of anatomical properties: lacking fingers (exp.1) or ended with a metal gripper (exp. 2). In the experiment, participant's perspective was shifted to a human-sized robot point of view with the use of a head-mounted display receiving visual feedback from the camera mounted on the robot's head. After the careful matching of robotic and real arms' positions, synchronous visuotactile stimulation was delivered to both hands. The sense of ownership was successfully induced for both robotic arms and did not differ in strength from real-hand conditions. Importantly, stimulation was delivered in the knuckles area. In their other work, using very similar virtual-reality set-up along with a robotic arm identical with the one used in exp. 1 of the experiment discussed above, Aymerich-Franch and colleagues (2017b) demonstrated that around 60% of the participants experienced a haptic sensation when they observed – from the first person perspective – a robot touching a curtain, without any tactile feedback. The felt sensation was projected to the area around the knuckles, which "might indicate that participants identified the end of the robot hand with the area corresponding to the knuckles" (2017b, p. 224). Therefore, it seems that the tactile stimulation was delivered to the corresponding parts of real and robotic hands, resulting in stimulation that was spatiotemporally congruent. Taken together, this "study demonstrates that humans can embody robotic limbs which are drastically different from a human limb in terms of shape, color, material, and texture" (Aymerich-Franch, Petit, Ganesh and Kheddar, 2017a, p. 488).

Tsakiris (2010) directly states that "body-model should not be equated with conscious body image" (p. 707) and points out to the fact that some body/hand features are irrelevant in the context of first critical comparison. However, features like 1) hand-like shape 2) tridimensionality 3) solid state and occupation of a certain space 4) finger possession 5) skin-like external layer seem to be too fundamental to be excluded – if they are irrelevant, which features actually matter? Note that other properties, such as hand color, size of the hand and its fingers, quantity of limbs and hand gender should also be excluded from the putative internal body model – all of them were experimentally shown to be causally irrelevant for illusion: hands of different skin color (Holmes, Snijders & Spence, 2006; Farmer, Tajadura-Jiménez & Tsakiris, 2012), elongated arms (Kilteni, Normand, Sanchez-Vives & Slater, 2012), large hands (Pavani and Zampini, 2007), shrunken and elongated fingers (Perera, Newport and McKenzie, 2015), supernumerary limbs (Ehrsson, 2009; Guterstam, Petkova & Ehrsson, 2011; Chen, Huang, Lee & Liang, 2018) and hands of opposite gender (own unpublished observations) may be incorporated[2]. Most of these findings are generalizable to a global body level, as shown by experiments in body-swap and virtual reality paradigms (see Aymerich-Franch and Ganesh, 2016 for a review). Thus, top-down constraints stemming from an internal body model would have to evince an enormous plasticity or interindividually variable selectivity of relevant features. Positing a mechanism of critical comparison between visual representation and appearance of a to-be-embodied object, in a situation in which most fundamental body features do not enter the comparison, and not a single visual body property was hitherto unambiguously specified to do so, is explanatory redundant. Therefore, we should assume that converging empirical evidence unequivocally and directly contradicts the internal body model hypothesis.

Moreover, this unidentified set of relevant features should prompt us to question the function of such internal body models. According to Tsakiris (2010), the filter operates in a gradual rather than a bottleneck fashion: "the more the viewed object matches the structural appearance of the body-part's form, the stronger the experience of body-ownership will be" (p. 707). Consistently, a gradual reduction of the strength of the feeling of ownership is sometimes reported with the distortions of the appearance of the hand (e.g. Ratcliffe and Newport, 2017). However, the role of the critical filter would be even more mysterious if its function would not be to either riddle or let an object representation through. What would be the function of gradually operating, composed of interindividually variable property sets and extremely liberal body models, which are competitive and functionally distinct from other body representations, e.g. body image and body schema, given the scarcity of its everyday applications?

3.2. Experimental results supporting internal body model hypothesis are actually inconclusive

Some objects have been repeatedly shown to resist embodiment: particularly wooden blocks (Guterstam, Gentile & Ehrsson, 2013), sticks (Tsakiris & Haggard, 2005) and sheets (Tsakiris et al., 2009); the effect is driven by inconsistent shape rather than the texture of the surface (Haans, Ijsselstein & de Kort, 2008; Aymerich-Franch, Petit, Ganesh and Kheddar, 2017a; see Hohwy & Paton, 2010; Armel & Ramachandran, 2003 for reports of possibility of

[2] Note that predictive processing framework may predict exclusion of some body properties from body model; a continuously adapting and liberal model would be more functional in the case of constantly changing body properties (e.g. hand size changes when one puts on weight, skin color temporarily changes from bruises and sun exposition etc.). However, this applies only to a limited set of properties.

incorporation of non-hand-looking objects; however, carry-over may be responsible for these effects as "object" conditions always followed "rubber hand" conditions). Since these objects do not resemble real hands, such reports are cited as supporting "interaction of top-down and bottom-up processes" hypotheses in numerous recent empirical and theoretical contributions (Ratcliffe and Newport, 2017; Tsakiris, 2017; Azañón et al., 2016). In this line of reasoning, visual representations of these objects are rejected during the first critical comparison (Tsakiris, 2010); therefore, one can say that top-down knowledge on the appearance of one's hand precludes embodiment.

However, these reports cannot account specifically for the internal body model hypothesis since they do not disambiguate between the effects of distorted appearance and reduced intersensory matching and, as such, are inconclusive in the matter of fact (Ma & Hommel, 2015a). As opposed to the studies carried out by Guterstam and colleagues (2013) or Aymerich-Franch and colleagues (2017a), stimulation delivered to the object did not closely mimic the one delivered on participant's hand; in particular, stimulations were incongruent in tridimensional space. In most of the studies, control objects were flat (e.g. Gutestam, Gentile & Ehrsson, 2013; Tsakiris et al., 2009; Haans, Ijsselstein and de Kort, 2008). As a result, stimulation delivered on objects was ideally parallel to the underlying surface, whereas stimulation delivered on hands was more or less diagonal. It could be crucial in the context of multisensory integration processes, since the rubber hand illusion is very sensitive to discrepancies in stimulation orientations in hand-centred reference frame (Constantini and Haggard, 2007; Gutestam, Gentile & Ehrsson, 2013).

In some cases, there are actually good reasons to ascribe abolition of the illusion to elevated sensory mismatch rather than to distorted appearance. For example, Tsakiris, Carpenter, James and Fotopoulou (2009) employed five different objects: a wooden sheet (object 1.) that was gradually transformed to a flat hand-like shape with all fingers (object 4.) via addition of a thumb-like feature (object 2.) and wrist (object 3.). 3D real-sized prosthetic hand was the fifth object. Despite gradual likening to the hand, the illusion was absent for all flat objects (objects 1.-4.) and could only be elicited for realistic prosthetic hand. These results seem to contradict the internal body model hypothesis as presented by Tsakiris (2010, p. 707): "the more the viewed object matches the structural appearance of the body-part's form, the stronger the experience of body-ownership will be". It seems that increased intersensory matching present in the fifth condition was an actual turning point (note that intersensory matching was increased in other sensory domains as well, e.g. expected weight of the prosthetic hand is consistent with a felt weight of one's hand as opposed to thin wooden sheets). Surprisingly, the authors interpret their results as coherent with internal body model hypothesis: "viewed object must fit with a reference model of the body that contains important structural information about body parts" (Tsakiris, Carpenter, James & Fotopoulou, 2009, p. 343).

3.3 There is no direct neuroscientific evidence for internal body model hypothesis

Apps and Tsakiris (2014) list a set of brain areas engaged in self-attribution processes. Distinct functional properties are ascribed to each of the structures, with "temporoparietal junction (TPJ) processing the confluence of visual information and bodily related information, the anterior insula (AI) processing the confluence of emotional, interoceptive and motor information about the body, the intraparietal sulcus (IPS) processing visuo-spatial information about somatosensory input to the body and the interior frontal gyrus processing the mappings between abstract rules and the body (IFG)" (Apps & Tsakiris, 2014, p. ?). These structures have been repeatedly shown to be activated during rubber hand illusion (Limanowski & Blankenburg, 2014). The role of anterior insula should perhaps be underlined (Tsakiris, 2017). A metaanalysis conducted by Grivaz, Blanke and Serino (2017) has shown that anterior insula is selectively activated during body ownership, but not when multisensory cues are simply presented within peripersonal space. Moreover, right insular cortex lesion prevents the integration of body-related exteroceptive and interoceptive signals into united self in cardio-visual stimulation conditions (Ronchi et al., 2015), which occurs in healthy individuals (Aspell et al., 2013; Suzuki et al., 2013). However, neuronal evidence for the internal body model is mainly circumstantial as no direct support has been found.

The strongest neuroscientific evidence for internal body model hypothesis comes from the study carried out by Limanowski and Blankenburg (2015) who proposed a neurodynamic PP-based model. Using dynamic causal modeling (Friston, Harrison & Penny, 2003), they have shown the strengthening of effective connectivity from lower-level perceptual areas, such as lateral occipital cortex (LOC) and secondary somatosensory cortex (SII), to higher-level integrative multisensory hub (intraparietal sulcus; IPS) when spatially congruent stimulation was applied (as opposed to incongruent stimulation). This bottom-up model outperformed bidirectional and top-down models. Counterintuitively, the bottom-up model was interpreted as lending empirical support for PP-based interpretation. According to the authors, spatiotemporal congruence of seen and felt touches leads to their association and ascription of a common cause. Since locations of seen and felt touches are discrepant, this results in increased prediction error propagated up the hierarchy from LOC to IPS which counters this mismatch via recalibration of somatosensory reference frame coordinates onto visual reference frame. This leads to an error suppression in LOC, but elevates prediction error in somatosensory areas since changed somatosensory coordinates do not match skin-based and proprioceptive information about the location of the hand.

This line of reasoning is highly speculative. Firstly, such interpretation rests on the assumption that enhanced neuronal activity reflects the spreading prediction error. Secondly, one would expect effective top-down modulations to come forward, since, according to the PP model presented by Apps and Tsakiris (2014, p. ?), "surprise in one system can be minimised by the top-down effects of multisensory nodes". In particular, modulations from IPS to LOC silencing prediction errors via recalibration of somatosensory reference frame coordinates should be present in spatially congruent stimulation (and, perhaps, error-related effective connectivity from LOC to IPS should not be present as these errors would have to be resolved for the illusion to come up). However, enhanced connections from IPS to LOC were found independently from congruency and were interpreted as top-down attention to visual processing resulting in increased weighting of visual signals in multisensory integration processes. Then again, this interpretation may be challenged. Intrinsic connectivity in both LOC and SII was attenuated regardless of the experimental context (Fig. 6, p. 2297). Lowered intrinsic connectivity in primary somatosensory cortex during RHI was also found – using dynamic causal modeling - and interpreted by Zeller, Friston and Classen (2016) as reduced precision weighting. If this is true, reduced intrinsic connectivity could not result from top-down attention in the PP framework. Moreover, Limanowski and Blankenburg (2015) refer to the finding that, during performance of a visuotactile task, connection weights between LOC/SII and IPS change accordingly with the reliability of the corresponding modality (e.g. for reliable visual information connectivity between LOC and IPS is enhanced; Beauchamp, Pasalar and Ro, 2010). However, Beauchamp and colleagues (2010) are agnostic about the causal direction (and they seem to think of it as of a bottom-up rather than top-down connection; e.g. see description of fig. 5). They also explicitly write that their "data is incompatible with a simple effect of top-down visual attention, and consistent with behavioral studies showing that reliability weighting is independent of attention" (p. 8).

Interestingly, Limanowski and Blankenburg (2015) focus on the PP-based explanation, which is very thoroughly analyzed, despite their own claim that these effects "may also be interpreted as reflecting processes of multisensory integration that produce the coherent ownership experience" (p. 2301). In this simpler interpretation, signals from LOC and SII would evoke multisensory integration processes in IPS only in the case of congruent information. However, this path is not explored and the latter explanation is waved as consistent with PP account as well. Thereby, Limanowski and Blankenburg commit a consistency fallacy (Cole & Klein, 2010; Coltheart, 2013) – they claim that their "results comply with the idea that the brain's inference mechanisms rely on the hierarchical propagation of prediction error" (p. 2284) even though these results are not inconsistent with a competing theory. As such, the study is theoretically nondiagnostic since the results do not specifically account for any theory. That said, the study is cited in a contemporary literature as providing *empirical support* for internal body model hypothesis (Tsakiris, 2017; author's emphasis).

Finally, Limanowski and Blankenburg (2015), performing dynamic causal modeling, did not define models that would count against PP theory. Instead, they elaborated PP-based post-hoc explanation – and the plausibility of such explanations may depend on rhetoric capabilities rather than data. Falsifiability conditions should be pre-defined prior to the experiment – since the authors did not argue what kind of data would be incompatible with the theory, obtaining such data was impossible. According to Coltheart (2013) this is a form of consistency fallacy, since an experiment is planned in such a way that it cannot provide results inconsistent with a tested theory. This criticism may also apply to other PP-inspired studies employing dynamic causal modeling to study how RHI arises (e.g. Zeller, Friston & Classen, 2016).

4. Multisensory integration

According to Ehrsson (2012, p. 797), "the natural constraints of the rubber hand illusion fit nicely with the multisensory integration hypothesis". It is constrained by the peripersonal space (Lloyd, 2007) and arises only when synchronous stimulation is applied to both hands (Botvinick and Cohen, 1998); therefore, it obeys basic rules of multisensory integration saying that stimuli originating from similar spatial locations and presented at the same time are more strongly integrated (Holmes & Spence, 2005). Stimulation patterns misaligned in the hand-centred reference frame, even when aligned in external space, result in the reduction of the illusion strength (Constantini and Haggard, 2007). RHI is also abolished in the case of large anatomical implausibility of dummy's position (e.g. when it is rotated by 90°; Tsakiris and Haggard, 2005; however, incompatible body postures may be taken as a top-down factor when interpreted as a body-related visual information; e.g. Blanke, Slater & Serino, 2015; Apps & Tsakiris, 2014) which underlines the importance of certain visuoproprioceptive coherence (Erro, Marotta, Tinazzi, Frera & Fiorio, 2018; however, small discrepancies in hand orientations may be tolerated; Constantini and Haggard, 2007).

In the present article, multisensory integration theory, as compared to the one presented by Ehrsson (2012), will be developed and radicalized: RHI actually follows much more complex multisensory integration rules and evinces features of a regular multisensory integration process – e.g. reliance on correlation of temporal structures rather than mere temporal coincidence (van Dam, Parise & Ernst, 2014; Parise, Harrar, Ernst & Spence, 2013), optimal integration rule (Ernst & Bülthoff, 2004) and crossmodal correspondences (Parise, 2016) between visual and tactile signals. As such, it may be modeled as a multisensory integration process. In particular, I propose that RHI does not involve any dedicated neurocognitive mechanism of self-recognition or embodiment. In this reductionist take, RHI occurs when seen and felt touches are falsely interpreted as caused by the same external event. Since tactile modality defines real-time boundaries of the body (informing about current body-world touchpoints), any ascription of the common cause to visuotactile signals necessarily results in recognition of an object as one's body part – and it may be an external object in the case of actually distinct origins of spatiotemporally synchronized touch patterns. However, basic concepts of multisensory integration theory should be introduced before the presentation of the developed multisensory integration model of RHI.

4.1. Maximum Likelihood Estimation

None of the sensory modalities can provide reliable information about the multidimensional structure of the world in all circumstances (Ernst & Bülthoff, 2004). Unimodal sensory estimates may be 1) noisy, due to changing environmental conditions and spontaneous neural activity 2) specialized – signal reliabilities vary depending on the nature of the perceptual task. For example, visual modality is appropriate for the localization task because of its high spatial resolution. On the other side, auditory modality tends to dominate over vision in temporal judgments (Shams et al., 2000; Burr, Banks & Morrone, 2009) because of higher sampling rate of auditory signals 3) biased – unimodal estimates may be invariant yet repetitively inaccurate and 4) ambiguous. Multisensory integration of unimodal signals can alleviate these problems (van Dam, Parise & Ernst, 2014). Sensory information is integrated according to the Maximum Likelihood Estimation (MLE) rule (Ernst & Bülthoff, 2004). Given that consecutive sensory samples yield slightly varying estimates, environmental properties may be represented as likelihood functions with varying degrees of uncertainty (width of the distribution). On that basis, assuming that noises in different modalities are normally distributed and independent from each other, the reliability of each signal may be quantified as its inverse variance. Then, weights inversely proportional to a signal's variance are ascribed to each of the signals, yielding the optimal estimate – weighted average of unimodal estimates.

Note that, in contrast with the concept of precision weighting in PP (Hohwy, 2012), ascription of weights may take place in a bottom-up fashion – based on the signal's variance in a short time bracket directly preceding the estimate (quasi bottom-up; Ernst & Bülthoff, 2004) or the size of receptive fields of neurons providing the estimate (Parise & Ernst, 2016). In the latter, the "reliability of a signal's estimate is the emergent property of neuronal tuning of the particular stimulus" (Parise & Ernst, 2016, p. 6) and is inversely proportional to the size of receptive fields of activated neurons. Let's take V1 neurons as an example – they are highly specialized (e.g. react only to particular, well-defined orientations) and have small receptive fields. Therefore, their activations are highly specific. In the case of activations of neurons sensitive to particular orientation and the concurrent lack of or weak activation of neurons with overlapping receptive fields, but sensitive to other orientations, the distribution of responses has a well-defined peak (Ernst & Banks, 2002) and the signal is highly precise.

The optimal integration model has a substantial empirical support and many perceptual phenomena may be modeled in this way (van Dam, Parise & Ernst, 2014). Alais and Burr (2002) have shown that "visual capture" in audiovisual spatial localization task (present in a well-known "ventriloquist effect) may be reversed after adding noise to a visual signal. Even more importantly in the context of this paper, Ernst and Banks (2002) obtained analogous results for a visuo-haptic task in which participants had to determine which one of two consecutively presented ridges is taller. For unimodal discriminations, vision proved to be more reliable than touch when either no or small (67%) noise was added, equally reliable for moderate noise (133%) and less reliable for intense noise (200%). Using unimodal data, MLE-based model was developed to predict weights ascribed to particular modalities in a crossmodal task in which visual and tactile signals were slightly discrepant for the second ridge. Height judgments followed the MLE rule: they relied on visual signals for low noise conditions and on tactile signals when high noise was added to a visual signal; thus, weights were inversely proportional to signals' variances. Van Dam, Parise and Ernst (2014) provide a comprehensive review of a wide variety of crossmodal and within-modality effects that were experimentally shown to obey the Maximum Likelihood Integration rule.

4.2. The Correspondence Problem and probabilistic models of multisensory integration

Multisensory integration improves the precision of the estimate of a given property of interest as compared to unimodal estimates (van Dam, Parise & Ernst, 2014) and may improve it even if weights ascribed to particular signals are suboptimal (Ernst & Bülthoff, 2004). However, the benefit is present only if integrated signals are actually caused by the same external event – otherwise, there is a risk that inaccurate combined estimate biased by irrelevant information will be yielded. Therefore, the cognitive system has to solve the so-called correspondence problem and determine whether various signals have the same underlying external cause. To perform this task, perceptual systems use various sources of information, e.g. pertaining to spatiotemporal proximity of signals (the closer in space and time they occur, the more likely they are to share a common cause; Holmes & Spence, 2005) and temporal cross-correlation (van Dam, Parise & Ernst, 2014; Parise & Ernst, 2016). The latter seems to be more important than mere temporal coincidence; unimodal signals are integrated if they co-vary across time and have closely correlated complex temporal structures (Parise, Harrar, Ernst & Spence 2013; Parise, Spence & Ernst, 2012).

In addition to bottom-up factors discussed above, cognitive systems use knowledge on natural mappings between sensory cues from different modalities - crossmodal correspondences – as a top-down factor determining whether sensory fusion will take place. Parise (2016) discusses three categories of cue pairings: redundant cues (both modalities provide information about the same environmental property – e.g. stimulus location), related cues (when cues from different modalities pertain to seemingly non-related sensory features, but are reciprocally predictable to a certain extent, e.g. auditory pitch and object’s size) and unrelated cues. Cues may be associated on the basis of statistical intersensory dependencies revealed in the process of the continuous interaction with the environment. For example, high sensory pitch and small size may be associated since they frequently co-occur. In this manner, “sensory systems become fine-tuned to the natural mapping across cues” (Parise, 2016, p. 13), developing predictive models encoding which signals tend to go together and how strongly they are related. These predictive models seem to be very flexible and experience-informed. Studies implementing perceptual learning paradigms show that new crossmodal mappings may be learned in laboratory conditions for initially unrelated sensory cues (Ernst, 2007) and existing intersensory associations may be reversed after the repeated exposure to inverted mapping between cues (Flanagan, Bittner & Johansson, 2008).

Within a Bayesian framework, the input of these predictive models may be operationalized as Bayesian coupling priors representing beliefs that two signals were caused by the same external event (van Dam, Parise & Ernst, 2014). The distribution of a coupling prior is determined both by spatiotemporal properties of signals (spatiotemporal proximity, correlation of a temporal structure) and their learned relatedness. Unimodal estimates are multiplied by a coupling prior to determine whether combined estimate or separate estimates are more likely to reflect external cause(s) of the sensory signals. When particular sensory cues are considered to be independent, the distribution of a coupling prior is flat (with infinite variance and no clear peak) and posterior is determined by separate unimodal estimates. For narrow distributions with well-defined peaks and minimal variance, standing for stronger association between the senses, sensory fusion takes place and the combined estimate is provided - since redundant cues follow a one-to-one mapping, entire variance of one of the unimodal estimates may be explained by its relation to another (hence their “redundancy”). Importantly, for slight intermodal discrepancies and partial relatedness, “partial fusion will take place, and there will be perceptual benefit for estimating the property of interest” (van Dam, Parise & Ernst, 2014, p. 219). Therefore, multisensory integration should be envisioned as a gradually operating rather than a “go/no-go” process.

5. Multisensory integration model of RHI – v. 2.0

Equipped with a multisensory integration theory lexicon, we can use its concepts to develop a parsimonious and comprehensive model of RHI. During the multisensory stimulation set to induce RHI, the cognitive system observes simultaneous visual and tactile signals that originate in PPS and are spatially congruent in hand-centred reference frame, yet discrepant[3] in external space. These are redundant cues: 1) they pertain to the same environmental property (area of space from where tactile sensations emerge) 2) as such, they have been learned to assuredly go together: in everyday interactions, spatiotemporally congruent visuotactile signals unambiguously attest that particular spot on one's body is being touched 3) the temporal structure of brushstrokes is complex and highly correlated (note that RHI is stronger for irregular than regular synchronized stimulation patterns; Guterstam, Petkova & Ehrsson, 2013; Rohde, di Luca & Ernst, 2011). As a result, Bayesian coupling priors have a distribution with a well-defined peak, promoting sensory fusion of visual and tactile signals. A common cause to both signals is ascribed and the touch is now referred to a rubber dummy (the only one object touched in one's visual field). Since tactile modality defines body boundaries, the sense of ownership is spread over a rubber hand.

In the proposed model, coupling prior distributions waive the need for internal body models (Tsakiris, 2010, 2017; Apps & Tsakiris, 2014) as the relevant visual information is stimulation-related rather than body- or object-related. To-be-embodied objects are reconceived of as "carriers of sensory signals"; their properties that are irrelevant in the context of the stimulation (e.g. hand color, general appearance or hand-like character) remain irrelevant for embodiment as well. It is the congruency of visual and tactile signals that steepens the distribution of a coupling prior, promoting sensory fusion and the occurrence of the illusion. In particular, visual and tactile signals following parallel spatial curvatures in the tridimensional space (even in the absence of the object; Guterstam, Gentile & Ehrsson, 2013) and hand-centred reference frame (Constantini & Haggard, 2007), as well as complex, irregular temporal patterns (Guterstam, Petkova & Ehrsson, 2011) lend weight to the hypothesis that there is a single external cause underlying distinct unimodal estimates. When stimulation patterns diverge (e.g. brushstrokes are delivered perpendicularly to the underlying surface on a block of wood and diagonally on a hand), coupling prior flattens, since these signals are very unlikely to go together. RHI sensitivity to postural incongruencies underlines also the importance of visuoproprioceptive coherence; however, the full list of relevant sensory cues is yet to be elaborated. Other factors, such as 1) contact area between the underlying surface and the dummy 2) dummy's expected weight 3) inclination and positioning of participant's fingers may play an important role (these factors may be challenging to control in experiments, given the large individual differences in weight, size, skin conformity or finger shape and positioning among the participants).

[3] However, both hands have to be placed within a peripersonal space (Blanke, Slater and Serino, 2015). Peripersonal space has repeatedly been shown to have sharp, well-defined boundaries and visuo-tactile or audio-tactile multisensory facilitation effects occur only within these individually defined ranges (Serino et al., 2017). Therefore, visual and tactile signals in RHI, even though discrepant in terms of physical space, are "peripersonally congruent", since they both emerge from spatial locations lying within an interface of potential interactions with external objects.

Analogously, if everyday sense of ownership is driven by bottom-up processes, disownership of real hand should occur as a result of breakdown of integration of visual, tactile and proprioceptive signals. This has been shown by Newport and Gilpin (2011) who described a "disappearing hand trick" exemplifying the relevance of multisensory congruence for the sense of ownership over one's own limb. In their experiment, they used mediated-reality conditions so the participant could view the live video stream of her own hand. Initially, displayed and real locations overlapped. Then, the real hand was displaced using the sensorimotor adaptation procedure – participants had to maintain the hand within the boundaries of the display. Since the displayed area was very slowly shifting towards one side, participants were subliminally displacing their own hands in accordance with the direction of the shift. After sensorimotor adaptation procedure, the displaced hand could not be found with the contralateral hand reaching to the perceived location of the real hand. This resulted in an immediate loss of ownership over the displaced hand, evincing as an inability to assess the real position of the hand, self-reported disownership and lack of physiological arousal in a situation threatening the hand in both perceived and real locations. In the authors' words (p. 805), "the lack of hand awareness (and associated lack of a skin conductance response) in the disappearing hand condition indicates a failure to resolve disintegrated vision (removed), proprioception (realigned) and touch (absent) in these key neural networks, resulting in a lack of ownership for the real hand".

Moreover, the multisensory integration model predicts interindividual differences in the distribution of coupling priors, accounting for interindividual differences in the proneness to the illusion. In particular, it operationalizes these differences as resulting from idiosyncrasies in perceptual processing rather than liberal or conservative internal body models. Interindividual differences are a well-known aspect of RHI, with 66-80% of subjects experiencing the illusion (Durgin, Evans, Dunphy, Klostermann and Simmons, 2007; Capelari, Uribe & Brasil-Neto, 2009; note that in the original work by Botvinick and Cohen (1998), prevalence was reported to be as low as 42% and, therefore, is very likely to rely on the quality of the stimulation delivered) and mean ownership ratings spanning below maximal ratings (e.g. Siedlecka, Klimza, Łukowska & Wierzchoń, 2014, Capelari, Uribe & Brasil-Neto, 2009; however, to the author's best knowledge, a comprehensive study dedicated to ownership ratings have not yet been carried out). It is foreseeable from the multisensory integration theory perspective: because of the existing spatial discrepancy between visual and tactile signals, the distributions of coupling priors should have non-zero variances. Therefore, incomplete fusion should take place, with stronger illusions occurring for prior distributions with better-defined peaks and weaker or non-existing illusions for widespread, flat prior distributions. Individual prior distributions may differ due to various factors; for example, scope of temporal binding window – a maximal tolerable asynchrony between signals for a cognitive system to judge them as occurring simultaneously. Synchronous and asynchronous stimulation in RHI can be redefined as occurring, respectively, within and outside individual temporal binding windows which have been shown to vary among people (Constantini et al., 2016). Interestingly, with the increase of delay between visual and tactile signals, subjective strength of the illusion tends to diminish but variability of reported illusion strengths increases (Shimada, Fukuda and Hiraki, 2009), which suggests that even for large delays between visual and tactile signals they may fall within a liberal temporal binding window (Constantini et al., 2016). Asynchronous stroking is well-known to entail some diminished form of the illusion (e.g. Guterstam, Petkova & Ehrsson, 2011). Slight temporal discrepancies during RHI elicitation are also expected, particularly in manual stroking procedures; however, whether automated procedures yield stronger illusions remains unclear, with various results being reported (e.g. Rohde, di Luca & Ernst, 2011; Rohde, Wold, Karnath & Ernst, 2013). Individual temporal binding windows may also be crucial for illusion strength indices based on differences between synchronous and asynchronous conditions (e.g. Tsakiris & Haggard, 2005; Tsakiris, Prabhu & Haggard, 2006).

Speculatively, other interindividual differences may come from differences in PPS size (and related personality traits, e.g. anxiety; Sambo & Iannetti, 2013) or in individualized patterns of perceptual processing of various kinds of sensory information, resulting in distinct weight ascription patterns. Spatial tactile acuity, defined as the ability to discriminate the spatial structure of surfaces (e.g. orientation of embossments) coming in contact with one's skin (Peters, Hackeman & Goldreich, 2009; van Boven & Johnson, 1994), may be of particular importance here. It is very likely to be related to sensitivity to discrepancies between stimulation orientations, which have been shown to play an important role in RHI (Constantini & Haggard, 2007). Spatial curvatures of stimulation patterns on both hands may not overlap exactly, particularly in manual stroking procedures (due to the inexact placement of both brushes, slight discrepancies in stimulation orientations and morphological differences between hands and dummies) – and such mismatches may be detected or overlooked depending on participant's tactile acuity. Tentatively, increased weighting of tactile signals should also be observed in people with high tactile discrimination skills, resulting in diminished proneness to RHI. Importantly, large individual differences in this ability have been observed, both between same-aged subjects and between younger and older adults, as spatial tactile acuity tends to decrease with age (Vega-Bermudez & Johnson, 2004). It may be related to morphological features, such as finger size (Peters, Hackeman & Goldreich, 2009) or skin conformance (Vega-Bermudez & Johnson, 2004). Taken together, spatial tactile acuity is very likely to underlie individual differences observed in RHI experiments.

5.1. Rubber Hand Illusion, Maximum Likelihood Integration and probabilistic multisensory integration

The current models of self-recognition acknowledge the importance of multisensory integration. For example, Apps and Tsakiris (2014, p. 3) write that "recognising one's self is a process of associating the unimodal properties of the body (i.e., the visual properties of one's hand), with other information about the body from any sensory system". However, they stress body-relatedness of multisensory information which is deemed unimportant in the presented account. In particular, body-related visual information is replaced by a visuotactile congruence and knowledge on the structural properties or anatomical plausibility is reconceived of as visuoproprioceptive congruence. Coupling priors pertain to the properties of stimulation (which sensory signals tend to go together and what is the strength of their relationship) and are agnostic about the nature of the object. On the other hand, internal body model hypothesis predict a large set of fundamental morphological constraints – including material presence, human-like appearance and shape of the hand, presence of the fingers, coherent laterality or gender and possibility of embodying only one arm at the time – all of which have been shown to be irrelevant for embodiment. When top-down morphological constraints stemming from the putative knowledge on the body are banished, curious plasticity of human self-recognition system may be explained with the sole reference to multisensory integration rules.

We can point out to "visual drift" and "additional limb" phenomena to show how multisensory integration theory may account for phenomena inexplicable from internal body model hypothesis perspective. According to Tsakiris (2010), as a result of sustained congruent stimulation, visual capture of tactile signals takes place: somatosensory coordinates are shifted onto visual reference frame. It may be conceived of as a "winner-takes-all" mechanism, where visual signals simply dominate tactile signals. This view is thought to be justified by a presence of a reproducible effect of "proprioceptive drift" (Botvinick & Cohen, 1998; Tsakiris & Haggard, 2005). After RHI elicitation, estimations of one's real hand position are skewed towards the rubber hand. However, proprioceptive drift is only partial as the real hand is localized in-between two hands rather than in the location of a rubber hand; as such, it is believed to be a causally unrelated correlate of the illusion (Abdulkarim & Ehrsson, 2016). "Winner-takes-all" view may also be challenged with the reference to recent reports showing that visual representation of a rubber hand is shifted towards real hand's position as well (Erro, Marotta, Tinazzi, Frera & Fiorio, 2018; Fuchs, Riemer, Diers, Flor & Trojan, 2016). When asked to localize the position of a *rubber* hand (either with the use of a verbal report or movement of a contralateral arm), participants tend to slightly mislocalize it in the direction of a real hand. This is consistent with Maximum Likelihood Estimation rule: in the process of multisensory integration, a weighted estimate, resulting in a unified percept, is obtained, with combined spatial representation being localized closer to the rubber hand because of higher weights ascribed to visual signals. Note, however, that spatial representations of two hands converge towards each other but do not completely overlap. This may be caused by the fact that, due to spatial discrepancies between visual and tactile signals, the fusion is incomplete and the cognitive system retains access to unimodal estimates (and uses this information while performing the task; van Dam, Parise & Ernst, 2014). Erro and colleagues (2018) speculate that incomplete convergence may come from a cognitive bias: top-down knowledge on placement of both hands prior to experimental manipulation.

Probabilistic multisensory integration framework may also be used to model a "supernumerary limb" phenomenon. Employing double paintbrush setup, Ehrsson (2009) has shown that tactile sensations may bifurcate: when two visible right rubber arms were stimulated in synchrony with an occluded real hand, participants reported two distinct yet simultaneous feelings of being touched on both hands. Guterstam, Petkova & Ehrsson (2011) replicated this finding in a slightly modified experimental setup: they placed an additional limb directly beside the visible real hand and confirmed the additional limb phenomenon with the use of self-report measurements as well as physiological recordings. Importantly, duplication of touch referral results in attenuation of the sense of ownership spread over particular hands, as if the total amount of sense of ownership was divided rather than doubled. It may be explained by a bimodal coupling prior distribution – in the face of ambiguous sensory evidence, signals are equiprobable to emerge from two spatial locations, which results in the sense of ownership being split into both hands (Ehrsson, 2009; Guterstam, Petkova & Ehrsson, 2011). This is why both fake hands have to be at the same distance from one's real hand (or both rubber and real arms should be placed at the same distance from one's shoulder) for the "additional limb illusion" to occur (Folegatti, Farnè, Saleme & de Vignemont, 2012).

6. Future challenges and empirical validation

Multisensory integration model v. 2.0 offers a parsimonious explanation to how RHI arises and proposes that RHI is not only constrained by basic multisensory integration rules (such as spatiotemporal proximity of visuotactile signals), but that it can be modeled using probabilistic multisensory integration model and more complex laws and constructs, such as Maximum Likelihood Estimation, coupling priors and structural temporal correlation. However, two major challenges for the theory may be indicated. The first one pertains to the exact role of proprioceptive signals in RHI which until now remains unclear. The second one concerns the relation between active (embodiment as a result of active exploration with coherent sensory feedback) and passive (embodiment as a result of passively received spatiotemporally congruent stimulation) condition in the context of multisensory integration.

6.1. The role of proprioception

Proprioception is frequently assumed to be a constraint of RHI. Illusion occurs only if the distance separating the hands is shorter than 30cm (Lloyd, 2007) and the rubber hand must be placed within peripersonal space for the illusion to arise (Blanke, Slater & Serino, 2015). However, the extent to which online proprioceptive signals contribute to the illusion may be smaller than previously thought. Admittedly, certain visuoproprioceptive coherence is necessary (since anatomically implausible rubber hand positions and large hand orientation mismatches abolish the illusion; Ehrsson, 2012). On the other hand, Abdulkarim and Ehrsson (2016) have shown that mechanical displacement of participant's hand during illusion elicitation (either towards or contrariwise to the rubber hand) does not influence the strength of the illusion. In their interpretation, the causal role of proprioception is limited since the illusion is not dependent on shifting proprioceptive representations. Moreover, the onset of the illusion in PPS is abrupt and the strength of the illusion is not related to distance separating the hands as long as they are both placed within PPS (Lloyd, 2007; Samad, Chung & Shams, 2015). In our lab (Motyka & Litwin, in preparation), we confirmed these observations as we observed that RHI strength did not differ for small (8 cm) and large (24 cm) discrepancies between locations of the hands. More importantly, proprioceptive accuracy – operationalized as a mean absolute difference between initial and reproduced positions in a task requiring repeated active reproduction of one's arm position (Lubiatowski et al., 2013) – was not a significant predictor of the illusion strength, both in "close" and "far" conditions.

Bayesian Factor analyses confirmed that our results reflected genuine null effects rather than experimental insensitivity. Taken together, it seems that weighting of proprioceptive signals (which should be higher for participants evincing high proprioceptive accuracy) does not influence the illusion strength and, therefore, the relevance of proprioception in multisensory integration processes underlying RHI onset is minimal. Tentatively, it is the tactile input rather than proprioceptive input that provides relevant information (e.g. about coherence of stimulation orientation in hand-centred reference frame). However, this remains to be a subject for future empirical investigations.

6.2. Active vs passive condition

The multisensory integration model, as outlined above, pertains to passively induced sense of ownership (like in the classic RHI setup) and is largely agnostic about the so-called "active" conditions of illusion elicitation. Active conditions further diminish the importance of appearance[4], since the sense of ownership results from sensory feedback matching predicted input rather than spatiotemporal coherence of stimulation. Therefore, passive induction method constraints (e.g. hand shape) are irrelevant in active conditions for which factors such as coherence of spatiotemporal properties of movement or tactile input resulting from contact with another object seem to play the major role. As a result, sense of ownership may be spread over objects strikingly distinctly morphed [Ma & Hommel, 2015a; b] – the sense of ownership over 2D rectangle would be very unlikely to arise from visuotactile stimulation.

The relationship between possibility to act, the sense of agency and the sense of ownership is a matter of debate in the contemporary literature (Tsakiris, Prabhu & Haggard, 2006; Kalckert & Ehrsson, 2012; 2014). Aymerich-Franch and Ganesh (2016) write that "agency is a much stronger modus for inducing embodiment than multi-sensory stimulations" (p. 34). Indeed, active movements were shown to result in a stronger illusion (Dummer, Picot-Annand, Neal & Moore, 2009). On the other hand, Kalckert and Ehrsson (2014) failed to find any differences in the sense of ownership over a rubber hand for different induction methods (active movements, passive movements and visuotactile stimulation), but, in their experiment, participants could only perform tapping movements with their index fingers. Visuomotor elicitation of the illusion results in illusion strengths spanning below maximal ratings, similarly to classic RHI induction methods (Sanchez-Vives, Spanlang, Frisoli, Bergamasco & Slater, 2010, tab. 2). Whether embodiment of strikingly dissimilar objects (rectangles) results in the same strength of the sense of ownership remains unclear, with both similar and weaker ratings of sense of ownership over rectangles being reported (Ma & Hommel, 2015a; b).

Nonetheless, Aymerich-Franch and Ganesh (2016) propose that internal body models refer to functional structural properties of the body. According to their Gibson-inspired functional body model hypothesis, external objects may be embodied if their physical

[4] See Ratcliffe & Newport (2017) for a report of attenuated illusion for distorted hand-like objects in mediated-related conditions. However, in this study, the virtual environment was not explored [as opposed to e.g. Ma & Hommel, 2015a] as participants' activity was restricted to the tapping with an index finger.

properties are sufficient to afford actions that the brain ascribes to the bodily counterpart of a to-be-embodied-object. Interestingly, it is sufficiency rather than correspondence that matters: objects that allow new actions are welcome. Therefore, people may embody larger hands but tend to reject the smaller ones (Pavani and Zampini, 2007); smaller hands significantly constrain action possibilities, whereas larger hands do not (see also experiments on robotic arms performed by the researchers, e.g. Aymerich-Franch, Petit, Ganesh and Kheddar, 2017a). This view corresponds with the classic accounts of body schema extension (Merleau-Ponty, 1962), e.g. embodiment of navigation-affording rods by blind people. Note that, however, some of the reports may be problematic to explain from this perspective (e.g. 2D rectangles or balloons significantly constrain action possibilities and do not allow new ones; Ma & Hommel, 2015a). Reconciliation of functional body hypothesis with multisensory integration theory – so they could collectively, comprehensively account for the experimental results obtained both in active and passive paradigms – remains to be an interesting future challenge.

6.3. Experimental predictions

While contemporary approaches underline the importance of appearance of a to-be-embodied-object (Tsakiris, 2010; Apps & Tsakiris, 2014; Azañón et al., 2016; Ratcliffe & Newport, 2017; Blanke, Slater & Serino, 2015), in the proposed model it is considered to be irrelevant. Past experiments on the effect of appearance failed in preventing concurrent reduction of the degree of intersensory matching (Ma & Hommel, 2015a) and, as such, they cannot be interpreted as providing support for internal body model hypothesis. In this section, I would like to pre-register an experiment allowing manipulation of an object's appearance without simultaneous influence on any of its stimulation-related properties.

The experiment would employ the gradual likening paradigm used by Tsakiris, Carpenter, James and Fotopoulou (2009). However, unlike in their study, participant's hand and all of the objects would be matched in terms of weight, volume, underlying surface area, layer ("skin") conformity, orientation and shape – to let the trajectory of brushstrokes be parallel. Note that such experimental paradigm would require separate preparation of a set of objects for any individual participant. Objects would be gradually likened via progressive addition of hand-like features in the finger area – for example, a fingerless hand-shaped mass (object 1.), would be likened to a hand through chiseling of channels imitating fingers closed together (object 2.) and addition of other finger features (e.g. fingernails, joints dividing fingers into distal, middle and proximal phalanxes; object 3.). Importantly, subjects would be asked to keep their fingers together. Then, spatiotemporally congruent stimulation would be delivered to the area over the knuckles and the corresponding area on the objects. In this setup, intersensory matching does not decrease in the process of likening as all of the objects are designed to "feel like one's own hand". For all objects, spatiotemporal stimulation patterns would be kept unchanged and any tactile and proprioceptive signals expected would be coherent with actual signals from the hand. As such, the distribution of the coupling prior should not change in the process of likening. Such experimental paradigm would let us adjudicate between competing models, as any differences in the sense of ownership between conditions would count against multisensory integration model as outlined above. On the other hand, lack of significant differences would count against internal body model hypothesis (Tsakiris, 2010; 2017; Apps & Tsakiris, 2014). Optimally, the experiment should be carried out in between-subject design to prevent any carry-over effects or response bias (since participants would be very likely to feel obliged to report that the sense of ownership is stronger in the case of more hand-like object).

7. Concluding remarks