Title: The free-energy self: A predictive coding account of self-recognition

Authors: Matthew A.J. Apps, Manos Tsakiris

PII: S0149-7634(13)00042-0
DOI: doi:10.1016/j.neubiorev.2013.01.029
Reference: NBR 1714

To appear in:

Received date: 21-11-2012
Revised date: 10-1-2013
Accepted date: 28-1-2013

Please cite this article as: Apps, M.A.J., Tsakiris, M., The free-energy self: A predictive coding account of self-recognition, *Neuroscience and Biobehavioral Reviews* (2013), doi:10.1016/j.neubiorev.2013.01.029

The free-energy self: A predictive coding account of self-recognition

Matthew A.J. Apps[1]& Manos Tsakiris[1]

[1]Laboratory of Action and Body, Department of Psychology, Royal Holloway, University of London.

Word Count: 10, 346

Number of Figures: 1

Number of Colour Figures: 1

Corresponding Author: Matthew Apps, PhD., Department of Psychology, Royal Holloway, University of London, Egham, Surrey, UK, tel: +44 (0) 1784 276551; Email: matthew.apps.2.2008@live.rhul.ac.uk or manos.tsakiris@rhul.ac.uk

Abstract

Recognising and representing one's self as distinct from others is a fundamental component of self-awareness. However, current theories of self-recognition are not embedded within global theories of cortical function and therefore fail to provide a compelling explanation of how the self is processed. We present a theoretical account of the neural and computational basis of self-recognition that is embedded within the free-energy account of cortical function. In this account one's body is processed in a Bayesian manner as the most likely to be "me". Such probabilistic representation arises through the integration of information from hierarchically organised unimodal systems in higher-level multimodal areas. This information takes the form of bottom-up "surprise" signals from unimodal sensory systems that are explained away by top-down processes that minimise the level of surprise across the brain. We present evidence that this theoretical perspective may account for the findings of psychological and neuroimaging investigations into self-recognition and particularly evidence that representations of the self are malleable, rather than fixed as previous accounts of self-recognition might suggest.

Keywords: Self-recognition, self-awareness, voice recognition, face recognition, body ownership, Bayesian, free energy, predictive coding, prediction error, rubber hand illusion, enfacement.

Highlights

- Self-recognition underpinned by Bayesian prediction and prediction error signals
- Illusory ownership of others' bodies underpinned by multisensory explaining away.
- Self-recognition is plastic and malleable during multisensory input
- Self processed by multimodal-unimodal interactions in non-self-specific regions

1. Introduction

The awareness of one's self and the concepts used to depict it are steeped in intellectual and scientific history. The ability to recognise one's own physical features in a mirror, or know that a voice is one's own is key for our self-awareness (Gallup, 1970), and also for our ability to communicate effectively with others (Bertenthal and Fischer, 1978). Such abilities are purportedly possessed by only a small selection of primate species, including humans (Reiss and Marino, 2001; Suarez and Gallup, 1981), and they are considered as behavioural markers of self-awareness. The question of what, if anything, makes the "self" special has led to a plethora of different research projects and hypotheses in psychological sciences and cognitive neuroscience (Devue and Bredart, 2011; Feinberg and Keenan, 2005; Gillihan and Farah, 2005; Legrand and Ruby, 2009). Despite extensive discourse in the literature, there has been a failure to reach a consensus across -or to large extent within- disciplines as to how the brain self-recognises. As a result there is also an absence of a theoretical framework which produces hypotheses which can be tested experimentally using neuroscientific methods. Despite the absence of a theoretical framework, attempts have been made to examine the neural mechanisms which underpin self-recognition (Legrand and Ruby, 2009). Such investigations have highlighted how many different areas, from primary unimodal sensory areas, to high-level multimodal association cortices are engaged when recognising one's self compared to the features of another (Devue and Bredart, 2011; Platek et al., 2008). However, recent reviews of this literature have concluded that the absence of a unifying theoretical framework has resulted in a largely incoherent picture of the circuits and mechanisms which are engaged during self-recognition.

Recently several reviews of the literature have noted the importance that efference copy (copies of multimodal sensori-motor commands which cause predictions across the brain about incoming sensory input) has in "self-processing", although not specifically in self-recognition (Legrand and Ruby, 2009). Specifically they argue that there are no self-specific networks in the brain, but that self-awareness and self-recognition result from the integration of motor efference (copies of the of

motor commands which generate predictions of the multisensory consequencex of an action) with reafference (the actual sensory consequences of an action). Alternative accounts have suggested that it is the integration of interoceptive efference and reafference that create the sense of a self or a self "presence" (Seth et al., 2011). Such accounts provide a useful insight into how self-specific information processing may arise in the brain, without the involvement of circuits that are specialised for processing "self-information". However, they do not deal with the more low-level and basic concept of how the brain processes an incoming visual, auditory, somatosensory, or interoceptive sensory input as "me" and how such input participates in the recognition of different aspects of one's physical self, such as one's face, body and its movement, or voice. In addition, many of the accounts of self-processing distinguish self information as special and therefore purportedly phenomenologically unique. As a result, it has been particularly difficult to embed theories of self-recognition and self-processing within theories of cortical function. Despite the aforementioned limitations, the salience of "self-processing" in human cognition and the wide network of areas that reported to be engaged during self-recognition, necessitates that theories of self-recognition are integrated within broad theories of cortical function.

In this paper we attempt to highlight how the free-energy principle, a recent attempt at a unifying theory of the brain, can explain many previous findings in self-recognition research (Friston, 2009). Within this framework we argue that self-recognition arises as a result of the brain's attempts to minimize the amount of free-energy (or 'surprise') in sensory systems in order to be in states where the environment is highly predictable. We outline how any aspect of the bodily self (e.g the physical features of a face or one's voice etc.), may be recognised as one's own through the optimisation of predictions about the sensory consequences of events occurring in the environment. Such optimisation occurs through the dynamic updating of Bayesian sensory predictions, when there is a discrepancy between a predicted sensory outcome and an actual sensory event (Clark, in press). Such discrepancies are referred to as prediction errors. Like previous accounts of self-awareness, we

place importance on the processing of discrepancies between predicted sensory states and actual sensory states (re-afference). However, by employing the free-energy principle as our conceptual and mathematical toolbox, we suggest that recognising one's physical form goes beyond integrating sensori-motor efference and reafference. Recognition of one's self will arise when predictions in the visual or auditory system about upcoming sensory input are congruent with other body related sensory information that includes, but is not exclusive to, predictions made based on corollary discharge (for a description see below). Recognition of one's self will therefore arise through the integration of sensory information creating multimodal representations of the self. Recently, it has been suggested that important metapsychological processes such as self-awareness can be explained within a free-energy framework (Fotopoulou, 2012). Here, we explain how this principle may also be able to account for empirical studies investigating self-recognition, which act as important behavioural markers of self-awareness.

The theory presented here is embedded within the Bayesian theoretical and mathematical framework of the free-energy principle. Within this article we will not provide a full treatment of the mathematics of free-energy, as eloquent and thorough accounts have been provided elsewhere (Friston, 2005; Friston, 2008a; Friston, 2009; Friston and Kiebel, 2009b). However, a description of the theory is pertinent for our aims and thus the earlier sections of this paper will provide an outline of the free-energy principle as a global theoretical account of cortical function. In later sections we will then outline what predictions this theory's many components make about how the brain might self-recognise. We will then discuss the extent to which this theory can account for the findings of Psychological and Neuroscientific investigations of self-recognition.

2. The Free-Energy Principle

The free-energy principle states that biological agents resist a natural tendency towards disorder in a constantly changing environment (Friston, 2005). The phenotype of an organism defines the extent of the physiological and sensory states that an agent can be in and therefore the boundaries of what

states that an organism can occupy. There is therefore a high probability that an agent (and its brain) will be in a small set of states and a low probability that it will be in a larger set of states. The often used example is that of a fish. A fish will have a very low probability of being on land, but a high probability of being in water. A fish on land is therefore in a very surprising and unlikely state. Mathematically speaking, the brain (as the organ within an agent that evaluates information about the external and internal milieu and resists disorder) must have a low level of *entropy* (Entropy being the surprise averaged over all events encountered) (Friston, 2005). To do this the brain only needs to minimize surprise associated with the current event by making predictions aboutwhat sensorial consequences will be evoked by events in the environment. Predictions are updated and optimised continuously over time in order that a low level of entropy is maintained across the brain. In the long-term, this means that the brain as a whole minimises the average of surprise in all sensory systems, learning how best to model and predict incoming sensory input. Additionally, it means that short-term phasic surprises ('prediction errors'), which are processed locally at each node of each sensory system, are avoided by actions that minimise surprise.

What role does free-energy play and how can agents minimise and avoid surprise? Free-energy acts as the upper bound on the level of surprise, which necessitates that surprise is minimised in two ways (Friston, 2010b; Friston et al., 2012c). Firstly, agents can act upon the environment to alter the incoming sensory events, sampling the environment in a manner that minimises prediction errors. That is agents will perform actions with predictable consequences that are confirmatory of expectations across the sensory systems. In turn, this minimises surprise across the brain in the long-term, as actions with surprising sensory outcomes are avoided and consequently the prediction errors evoked in each node in each sensory system are low. Secondly, prediction errors can cause agents to update estimates about the causes of the sensory events in a Bayesian manner (Clark, in press; Friston, 2005), in order that more optimal inferences about the actual causes of sensory

events can be made. Prior to any event, expectations are made based on representations of the probability of a sensory event occurring. These predictions are represented as a probability distribution, which are coded for by the internal states of the brain prior to an event (i.e. the activity of neurons and the strength of synaptic connections). When there is a sensory event which is discrepant from the expected input, the prediction errors coded for by neurons in sensory systems cause an update of the prior expectations dynamically, to give posterior probabilistic representations. The updating of posterior probabilities or beliefs pertains not just to predicted states of the world but also to contingencies between sensory events that determine how states evolve. In neurobiologically plausible implementations of free energy minimisation (predictive coding) this leads to a distinction between perceptual inference - in which the activity of population of neurons that encode posterior beliefs about the states of the world minimise prediction error - while changes in connection strengths or synaptic efficacy change over a slower timescale, to learn associations and statistical relationships, which serve to minimise the average level of prediction error over time. This leads to a distinction between perceptual inference and perceptual learning that we will return to below

Thus, short-term inferences about the actual sensory causes of events, and therefore putatively the content of conscious perception, will be made based on these posterior probabilities (Friston, 2005). In turn, updating the estimates of the priors of the causes of a sensory event, modifies future expectations (i.e., perceptual learning), such that similar sensory events in the future are predicted (Friston, 2012a; Friston and Kiebel, 2009a; Friston, 2008b). The brain is therefore processing dynamically shifting generative models of what is causing incoming sensory events, based on probabilistic predictions about how likely something is to have happened and what the likely causes are. In essence this means that a surprising sensory event, causes short-term phasic prediction errors, which are avoided by actions with minimal predicted surprise, and by changes in the

representations (i.e., the probability distribution) of what was likely to have caused the sensory input. In addition, future incidences of similar sensory events become less surprising as they are represented as more probable and therefore more predictable as a result of the previous surprise they evoked. In summary, the free-energy principle states that the brain has the overarching functional property of minimising surprise by (i) optimising probabilistic representations at local nodes in a network as a result of prediction errors and (ii) performing actions that have predictable consequences in order to avoid prediction errors.

3. Hierarchical Predictive Codes.

An important aspect of the free energy principle is that it makes assumptions about the organisation of sensory systems and also about the flow of information in these systems (Friston, 2008a; Friston and Kiebel, 2009b). These assumptions can be summarised within a "predictive coding" model, a framework that can be used to explain the architecture of sensory processing (Clark, in press; Lee and Mumford, 2003; Rao and Ballard, 1999). Previous accounts have discussed how visual, auditory and interoceptive sensory signals may be explained by predictive codes (Gagnepain et al., 2012; Rao and Ballard, 1999; Rauss et al., 2011; Seth et al., 2011; Summerfield and Egner, 2009; Wacongne et al., 2012; Winkler et al., 2012), but none have related these directly to the experience of how self-stimuli are recognised.

Predictive coding argues for complimentary hierarchical top-down and bottom-up processes, which are distinguished by the nature of the information that they process. Bottom-up information flowing through the hierarchy reflects the impact of a sensory event, i.e. prediction errors. Top-down information flowing through the hierarchy is in the in the form of predictions about the sensory consequences of events. At the top of the hierarchies are multisensory areas that will process abstract, supramodal representations of sensory input. Predictive coding also argues that such information will be processed by two separate classes of neurons; representational units, which

process probabilistic representations (or predictions) about upcoming sensory input and error units, which code prediction errors when there is a divergence between expected and actual sensory events (Clark, in press; Friston, 2005). Within each level of the hierarchy, there is a considerable exchange of information between the representational and error units, such that surprising events elicit a large, early response and locally update the prior probabilistic representations (i.e., they create a *posterior* probabilistic representation). In addition to the local exchange, any unexplained surprise in the error units is projected up the hierarchy to the representational units in the next level. This causes surprising events to evoke prediction errors that flow up the hierarchy. However, representational units dynamically update prior predictions and project these down the hierarchy. As such, representational units "explain away" error in the immediately preceding level of the hierarchy. As this dynamic process is bounded by minimising free-energy, the system iteratively and rapidly minimises surprise (or prediction errors) in sensory systems by updating probability distributions in the generative model, until the most probable cause of a sensory event is inferred.

In summary, predictive coding suggests that probabilistic representations act as a top-down influence on expectations explaining away bottom-up prediction errors. The inferred cause of a sensory event will be the posterior probability distribution when error has been minimised. At this point it should be explicitly stated that the aim of this paper is not to discuss the validity or evidence supporting the free-energy principle and predictive coding as overarching, unifying theories of the brain. There is evidence in support of its claims (Brown and Friston, 2012; Friston, 2010a, b, 2012a; Friston et al., 2012a; Friston and Ao, 2012; Friston et al., 2012b; Friston et al., 2012c; Friston et al., 2010), although the theory is still in its infancy and therefore is neither largely supported or refuted as yet (Clark, in press). In addition, there is considerable debate within the literature about whether predictive coding models need to be bound by free-energy and even whether predictive coding models need to operate within Bayesian principles (Clark, in press; Friston, 2012b). However, the purpose of this article is to outline whether, and how, free-energy might explain self-recognition.

The aim of this discussion is therefore to determine how free-energy and predictive coding might provide a useful theoretical framework of self-recognition that can account for existing results but also generate testable empirical hypotheses.

4. Free-energy Self-recognition.

The assumptions of the free-energy principle have implications for the neural and psychological processes that might underpin self-recognition. In this section we wish to highlight how its assumptions lead to several predictions about the mechanisms that will underpin self-recognition. At this point we will provide an overview of how the model could explain self-recognition in an abstract manner and not directly discuss the self-processing literature. In later sections we will discuss the core components of the theory in relation to empirical studies of self-recognition.

Perhaps the most important aspect of the free-energy principle is that sensory information is processed probabilistically, with prior predictions and posterior inferences made based on Bayesian optimised probabilities (Friston, 2005). Specifically, the Free-energy principle is underpinned by an empirical Bayesian framework. In Bayes theorem, the level of evidence about the true state of the world is expressed in terms of the level of belief, or the probability, in the occurrence of an event. The level of belief is a function of the prior probability distribution (the probabilistic level of uncertainty in the prediction of a sensory event) and the likelihood (the probability that the event actually occurred given the evidence). In empirical Bayes, the posterior probability, which is the result of the updating that occurs following sampling (i.e. a sensory event), reflects the degree of belief in the current model of the world. In turn, the posterior probability becomes the prior distribution the next time an event is sampled. Thus, the belief an agent has about what caused a sensory event is a conditional, probabilistic estimation of what happened and what was predicted.

It follows that the mental representation of the physical properties of one's self are therefore also probabilistic. That is, one's own body is the one which has the highest probability of being "me" as other objects are probabilistically less likely to evoke the same sensory inputs. This information can be considered as highly abstract with respect to the low-level properties of the stimuli and can only be represented as "self" when different streams of multisensory information are integrated. That is, the self-face will only be recognised as "self" when a visual stimulus has been processed hierarchically for its low level visual properties, its configural features and then it's identity. The self-face will therefore be represented as an abstract, supramodal representation of visual input e.g. this is a face, that I have seen before, that I am familiar with, and that is associated with congruent corollary discharge, vestibular, somatosensory and interoceptive information when seen on a reflective surface.

In predictive coding accounts, abstract information is encoded in terms of posterior beliefs at high levels of a hierarchical model; i.e., probability distributions of abstract supramodal events (Clark et al In Press). In hierarchical models, beliefs at intermediate levels of the hierarchy are referred to as empirical priors because they are constrained and (plastically) optimised by both top-down and bottom-up influences - in other words, they are prior beliefs that are sensitive to empirical sensory evidence. High level empirical priors are essentially the same as low level empirical priors but generally represent abstract multimodal beliefs about states of the world that change slowly over time. Such beliefs are learned through associations being formed between congruent, low-level sensory events from different systems, that over time result in one event having a high probability of predicting another sensory event (Ballard et al., 1997; Friston, 2008b). However, to produce these parallel (multimodal) predictions, there must be a high level representation (of self) that elaborates descending predictions to multiple unimodal systems.

This has implications for self-recognition processes, as it highlights how sensory events in one system can become associated with events in another and therefore how abstract representations

of one's body may be formed. To illustrate, visually observed touch on the skin that is temporally congruent with touch detected by the somatosensory system will become associated with each other, resulting in a prediction of a somatosensory event when contact to the skin is about to occur. In contrast, touch between two other non-corporeal objects will never evoke a somatosensory event, and thus the prior probability of a somatosensory event following touch on such objects is very low. So one's own body is probabilistically likely to become and be the object that touch is predicted to be experienced upon. The visual properties of different body parts will also be perceptually learned such that when any object approaches the body, a somatosensory event will be predicted. Thus, perceptual learning within the free-energy and predictive coding frameworks leads to generative models where aspects of one's body are processed as probabilistically the most likely object (or collection of objects) that when touched, moved, threatened, or acted upon in any way, evokes events in the other sensory systems that detect the state of the body. In short, the notion that there is a "self" is the most parsimonious and accurate explanation for sensory inputs. In mathematical terms, this parsimonious accuracy is exactly the quantity that is optimised when minimising free energy or prediction error

To illustrate, we use the example of recognising one's self in the mirror. At the ontogenetic level, self-recognition in the mirror poses two challenges. First comes the challenge of matching the sensorimotor experience of the body with the sensorimotor behaviour of the reflected image. The second challenge relates to how a mental representation of facial and bodily appearance is acquired in the first place. Given that the infant cannot have a priori knowledge of their appearance, the infant encountering a mirror for the first time must succeed in matching their sensorimotor experience with the observed sensorimotor behavior of the object seen inside the mirror. This matching between felt and observed sensorimotor signals will lead to the formation of a mental representation of visual appearance (i.e., "that is my body reflected in the mirror; therefore that is what I look like"). This process of self-identification allows successful performance in the classic

'rouge' task of mirror self-recognition, in which infants are exposed to their mirror reflection and their response to a spot of rouge covertly applied to their nose is registered (e.g., they might respond by touching their own nose; see (Brooksgunn and Lewis, 1984). When looking in a mirror there are several surprising features that need to be explained away. First is the spatially surprising nature of reflective surfaces, as the agent perceives the visual (sensory) consequences of bodily movements in an allocentric frame of reference. Second is the temporally surprising nature of the event as there is an object (i.e. a body) which moves in a temporally congruent manner to the corollary discharge of the agent. In this setting, corollary discharge is no more, or less, than any other descending prediction other than that it produces movement by containing some proprioceptive and kinaesthetic components. Third, is the surprise that the body seen in the mirror has a specific visual form. How can these surprises be explained away? The surprises evoked during mirror exposure will be explained away in multisensory areas that integrate visual information with corollary discharge, updating the probability that actions will result in movement of that body in the mirror. This will explain away the visual surprise. In turn, perceptual learning will lead to the visual features of one's body being processed as a highly likely input when one looks in a reflective surface. Thus, the viewing of an agent's own actions in a mirror (including arm movements, facial expressions etc.) will lead to optimised high level empirical priors about one's body, which will in turn modulate expectations in the visual system about the expected visual consequences of one's own actions. The agent will also therefore begin to recognise her face as "me" because it is typically the face that is processed when looking in a mirror, and *that* face rarely violates expectations instantiated by the agent's actions. One's body is therefore represented as the most probable to be "me" when seen in a mirror due to it being the most likely visual input when viewing a reflective surface.

At this point it is pertinent to point out the distinctions between this theoretical framework and other accounts of self-recognition. Previous accounts have highlighted the importance of the congruency of motor (or interoceptive) efference, and sensory input for self-awareness and for self-

recognition processes. It is notable that in the account we have given thus far and in the example of the mirror, that the congruency of predictions driven by corollary discharge and incoming sensory input is also important for driving self-recognition in our account. However, unlike previous accounts, the free-energy framework provides flexibility, with fewer constraints on what types of information can drive self-recognition.

Within a predictive coding framework, top-down predictive information processed in multisensory association cortices, plays an important role in altering the perception of sensory input. Prior beliefs will therefore modulate how self-stimuli are recognised. In addition, predictive coding argues that surprise in one system can be minimised by the top-down effects of multisensory nodes (Lee and Mumford, 2003). This suggests that surprise in any system could be explained away by probabilistic representations which are derived from information in any other system, if this is the optimal manner in which free-energy can be minimised. The free-energy account discussed here is therefore distinct from others in highlighting how information from any system can be used to explain away information in any other system (Mitchell, 1993). This distinguishes our theoretical perspective from previous accounts of self-recognition, which have argued that self-processing is tied to processing in one "self" network (Northoff et al., 2006), or arises as a result of congruencies between sensory input and motor efference alone (Legrand and Ruby, 2009). Our suggestion is that self-recognition is more complex, with information from each and every sensory system potentially able to modulate self-recognition. This is particularly important, given the evidence to suggest that the continuity of the self may be underpinned by many different types of information, the integration of which leads to a coherent sense of one's body (Blanke, 2012; Tsakiris, 2010; Tsakiris et al., 2008).

Furthermore, the free-energy principle reframes the nature of signals from the motor system, further distinguishing our theoretical perspective from others. We have outlined how action can be construed as minimising prediction error, as only actions that have predictable sensory outcomes are performed. This is known as active inference and rests on minimising prediction error relating to

proprioceptive expectations, through the use of classical motor reflexes. As such, the motor efference is relegated to the motor commands arising in the spinal cord and cranial nerve nuclei (Friston and Ao, 2012; Friston et al., 2011). In other words, cortical signals that drive movement are descending predictions about the proprioceptive consequences of movement and are therefore better considered as corollary discharge (the predictions of the sensory consequences of movement) as opposed to efference copy (copies of the motor commands used to form corollary discharge). In this sense, corollary discharge is just like any other descending prediction apart from the fact that one or more predictions will elicit movement and are therefore proprioceptive or kinesthetic in nature.

5. The Psychological Self

As stated above, an important aspect of the free-energy principle is that the brain can minimise surprise by updating probabilistic representations (Friston, 2005). Therefore, at the core of this theory is the notion that probabilistic representations are plastic and updated when new information reveals a discrepancy between a predicted sensory state and the actual sensory state. Self-recognition should therefore also be plastic, such that surprising sensory events may be explained away by changes in how sensory inputs that are "self" or "other" are processed.

5.1 Self Plasticity

Traditionally, self-recognition is measured using self-other detection tasks, self-other morphing tasks, (where participants stop a video morph between self and other when it looks more like "me"), or masked priming tasks where reaction times are compared between self and other related primes (Bredart, 2004; Devue and Bredart, 2008; Devue et al., 2009; Frassinetti et al., 2008; Heinisch et al., 2011; Keenan et al., 2000; Keenan et al., 1999; Kircher et al., 2001; Pannese and Hirsch, 2010, 2011; Rotshtein et al., 2005; Tsakiris, 2008). All of these methods converge on the notion of a self-bias,

with self-stimuli being more salient and processed faster. However, these methodologies are inherently examining unimodal representations of self-stimuli and therefore may violate the normal conditions in which one's body is experienced. This has rendered most self-recognition studies unable to examine the multisensory nature of self-processing.

A number of studies have highlighted the plasticity of the bodily self by showing how multisensory stimulation can modulate how one's own body, face and voice are recognised (Blanke, 2012; Botvinick and Cohen, 1998; Ehrsson et al., 2005; Ehrsson et al., 2004; Tsakiris, 2008; Zheng et al., 2011) (for reviews see(Blanke, 2012; Tsakiris, 2010). Perhaps the best known illustration of such plasticity is the "rubber hand illusion" (RHI) (Botvinick and Cohen, 1998). In the typical formulation of this illusion a rubber hand is placed in front of a participant and their own hand is placed out of view. The participant then receives tactile stimulation on their hand, whilst observing tactile stimulation on the rubber hand. When the tactile stimulation is delivered in temporal synchronicity, on congruent specular locations on the two hands, participants come to experience a sense of ownership over the rubber hand (Tsakiris and Haggard, 2005). In addition, participants' perception of the location of their own hand shifts to a spatial location closer to the rubber hand than its actual location (Tsakiris et al., 2006). Thus, a simultaneous multisensory experience can update the representation of the rubber hand as "not me", such that the probability that the rubber hand is "me" increases.

Similar effects of visuo-tactile synchrony have been shown to induce ownership for whole bodies. Ehrsson (2007) used synchronous or asynchronous visuo-tactile stimulation while participants were looking at their back with the perspective of a person sitting behind them with stereoscopic vision. Synchronous but not asynchronous visuo-tactile stimulation induced a shift in the 1st person perspective such that participants experienced being located at some distance behind the visual image of their own body as if they were looking at someone else. In the study by Leggenhanger et al (2007), participants viewed the backs of their bodies filmed from a distance of 2m and projected

onto a three-dimensional (3D)–video head-mounted display. The participants' backs were stroked either synchronously or asynchronously with respect to the virtually seen body. Questionnaire and behavioural measures showed that only after synchronous stimulation, participants felt as if the virtual body was their body. These manipulations demonstrate the efficiency of current multisensory input in determining the experience of a minimal 1st person-perspective (Ehrsson, 2007), self-location (Leggenhanger et al,(2007) and self-identification (Petkova and Ehrsson, 2008) three conditions that are critical for the experience of selfhood (Blanke and Metzinger, 2009).

More recently multisensory stimulation was used to show the plasticity of self-face recognition in the "enfacement illusion" (Mazzurega et al., 2011; Sforza et al., 2010; Tajadura-Jimenez et al., 2012; Tsakiris, 2008). In the enfacement illusion tactile stimulation is applied to the participants face whilst they observe the face of another being touched in a video. This experience is highly unusual, as it simulates the situation where one's own face is viewed in a mirror, however, in this case the observed face is that of another person. The result of synchronous stimulation is that participants begin to respond on self-other recognition tasks as if the others face was more like their own face (Sforza et al., 2010; Tsakiris, 2008). Similarly, self-report measures indicate changes in the phenomenological experience of one's own and the other's face, such that the face is experienced as more like "me" than before (Tajadura-Jimenez et al., 2012).

A similar "rubber voice" illusion has also been reported in the auditory system (Zheng et al., 2011). In this illusion participants talked into a microphone, whilst receiving auditory feedback that was either their own vocalisations or the voice of another in temporal synchrony. The result was that the stranger's voice is reported as a distorted version of one's voice, not as the voice of another and also modulated the pitch of their own speech.

Each of these illusions illustrates how multisensory stimulation can lead to changes in how one's face, voice and body are processed, leading to an update of what is recognized as "me". How can the

free-energy account explain changes in self-recognition that are driven by multisensory stimulation? The free-energy principle highlights how surprising events in one sensory system can be explained away by more parsimonious information in another by the convergence of information at multimodal nodes in the cortex. In each of these three illusions there is considerable bottom up sensory surprise evoked in one system. The somatosensory experience of touch on one's hand that is temporally congruent with the vision of touch on the rubber hand is surprising, as prior to stimulation participants cannot see the touch on their own hand and would not predict that touch on the rubber hand would evoke a sensation of touch. Similarly, there is surprise in the somatosensory system during the enfacement illusion and surprise in the auditory system during the rubber voice illusion. This surprise will be explained away by top-down effects from multisensory areas. In turn, perceptual learning processes will update representations of one's appearance or voice, such that the probabilistic representation of one's body and voice is different after synchronous multisensory stimulation. For example, it has been shown that only subjectively perceived physical similarity between the participant's hand and the rubber hand is influenced by the experience of the illusion (Longo et al., 2009). Participants who experienced the RHI perceived their hand and the rubber hand as significantly more similar, than participants who did not experience the illusion, suggesting that changes in ownership leads to changes in more abstract body image representations of one's appearance.

These illusions highlight how representations of one's body are malleable and can be updated when expectations about multisensory events are violated. We argue that such effects can be accounted for by top-down explaining away of the bottom-up surprise evoked by an unexpected event. This treatment of illusions - in the context of self recognition - is entirely consistent with current understanding that illusory phenomena are a result of Bayes optimal inference. In other words, almost universally, illusions can be explained as an unusual set of sensory circumstances being

interpreted under prior beliefs about their causes in a Bayes optimal fashion. In our examples, these prior beliefs reflect the fact that most of our sensations are caused by ourselves.

5.2 Probabilistic, predictive codes of the self.

An important point to note so far is that our account of the rubber hand illusion and also the enfacement illusion is distinct from any previous account (there has to date only been one paper that has reported the rubber voice illusion and so there has been no theoretical account). Previous accounts suggest that multisensory stimulation leads to changes in "representations of the self" through visual capture, and updates to bodily reference frames following the visual capture (Makin et al., 2008; Petkova and Ehrsson, 2008; Tsakiris, 2010). However, there has not previously been a theoretical perspective which can account for all three illusions, which are clearly driven by similar multisensory processes. In our account we argue that all three illusions can be explained through the principles of free-energy and predictive coding. In doing so, we argue that the driving effect for each illusion is an increase in the probability that the other object (a face, voice or body part) will be represented as part of the body and a decrease in the probability that one's actual body will be represented as "self".

In the free-energy framework, a change in the likelihood that a stimulus is represented as "me" is reflected by an updating of the high level empirical prior probability distribution representing a face, voice, or body being one's own that is updated. In turn, the high level empirical prior probability that one's actual face, voice or body are one's own is decreased. Interestingly, recent studies have reported physiological changes to the real hand, including a reduction in temperature (Hohwy and Paton, 2010; Moseley et al., 2008) and also an increase in histamine level (Barnsley et al., 2011) during the illusory experience. These findings are consistent with the notion that the real limb is being partially rejected and therefore there is a decreased likelihood that it is "self". Thus, the

likelihood that one's own body is "me" decreases, but the likelihood that a rubber hand is "me" increases during the illusion.

If there are probabilities of my actual hand being "mine" and another object being "me, how are these two representations resolved in order that one's self is recognised? The answer to this lies within predictive coding accounts of how competing models (i.e., explanations of the sensory input) are selected among. In predictive coding, more than one model of an event is processed at a time, with models with less evidence not being selected for and inferences based on the information they code are not made (Clark, in press). Pools of neurons that process evidence in favour of a particular model suppress the activity of neurons that process alternative models. As such, perceptual experiences are a function of the relative fit of models of the environment to the actual sensory input. To illustrate Friston et al., outlined how predictive coding accounts can explain the well-known binocular rivalry phenomenon (Blake and Logothetis, 2002; Leopold and Logothetis, 1996; Lumer et al., 1998; Tong et al., 1998). In binocular rivalry studies two different images are presented simultaneously to each eye. For example, a face is presented to one eye and a house to another. The resulting perceptual experience is a continuous flip from one percept to another. The free-energy account suggests that as one generative model of the image presented to one eye (e.g. the house) it becomes increasingly likely through the minimisation of surprise., the percept that is propagated by its top-down probabilistic representations is experienced, i.e., I see a house and not a face. However, when this percept is experienced, there is surprise evoked by the fact that the image presented to the other eye (e.g the face) does not fit with the generative model of what could cause the evoked sensory input. The probability that the cause of the sensory input is an image of a face therefore declines as a result of the prediction error signals. The alternative generative model ("face") is updated and the likelihood that the cause of the sensory input is "face" increases. As such, the flip between the two percepts is a function of the balance between the two generative models of the visual input (Hohwy et al., 2008). Each of these generative models will be processed by competing

pools of neurons within regions, across competing regions within systems or across competing systems. The probability that one inference will be made is a function of the relative difference in the fit by one explanation of the world compared to another.

The competition between alternative generative models of a stimulus as "self" offers an explanation of the processes of self-plasticity that occur in the multisensory illusion outlined above (see figure 1). For instance, when one's own face is presented in a mirror, the likelihood that this will be processed as "me" will be a function of the comparisons between two generative models of which face is "me". The difference in the likelihood that the self-face and the other face are actually one's own is reduced, leading to changes in the degree to which one's own face is recognised as "self". This is consistent with the evidence that self-face recognition performance shows a difference before and after stimulation in the enfacement illusion, as people perceive the other's face as being more similar to their own face after stimulation (Mazzurega et al., 2011; Sforza et al., 2010; Tajadura-Jimenez et al., 2012; Tsakiris, 2008). Behaviour on self-recognition tasks will also be dependent on the relative fit of the two generative models which code the probability of the other being me and the probability my own body being me. Thus, the predictive coding account illustrates how changes in the probability that another's face is "mine" can lead to changes in the likelihood that one's own actual face is "me".

In summary, in the last two sections we have highlighted how the free-energy principle suggests that one's own body, face and voice are probabilistically represented as "me". In addition, the nature of the processing that leads to a representation of a body part as "me" is tapped into in each of the multisensory illusions. As a result, the probability that another's body part is "me" increases and as a result the ability to distinguish between self and other is diminished. Such an effect is driven by changes in the high level prior probability that another or a different body is "me". Thus, in these sections we argue that these illusions tap into the mechanisms that operate to create a sense that one's body is one's own and that my body, face and voice are "mine".

5.3 Self priors and the self in context.

In the previous sections we highlighted how the physical features of one's self are processed as probabilistically the most likely to be "self" and particularly how incoming sensory input in self-related illusions causes others' bodies faces and voices to be more likely to be one's own. However, another important aspect of the free-energy principle is that top-down probabilistic representations can influence information processing prior to incoming sensory input and as a result influence the likelihood that an object will be recognised as "me". The context within which sensory stimulation is perceived will therefore influence priors and high level priors, resulting in self-other distinctions being dependent on expectations prior to the presentation of a self or other stimulus.

Contextual effects have been known for some time to significantly influence self-recognition. Indeed, self-related primes in one domain can decrease reaction times in self-other recognition tasks in another (Pannese and Hirsch, 2010; Platek et al., 2004). This effect is still apparent when the self-related priming stimuli are masked and therefore not consciously perceived (Pannese and Hirsch, 2011). Priming with a self-stimulus can therefore influence phasic expectations about the prior probability of further self-stimuli. As such, a stimulus is processed within the context of stimuli immediately preceding it.

There is also evidence of more long-term contextual influences on self-recognition related priors, highlighted by the role that cultural and societal effects have on self-other decision-making. For instance, self-other face recognition has been shown to be different across cultures (Liew et al., 2011; Sui et al., 2009). Indeed, Westerners show a greater self-bias on such tasks than East Asian individuals. In fact, a recent study has suggested that Chinese individuals show a reversal of the self-bias in the presence of a supervisor, with reaction times becoming faster for the face of the social superior than the self-face (Liew et al., 2011). Western individuals, however, maintain a self-bias in identical circumstances. Thus, self-face recognition is manipulated by cultural effects. Another

illustration of the impact of social context upon self-face recognition has been shown by the fact that religious individuals do not have as strong a self-other bias as atheists (Ma and Han, 2012). Cultural and societal norms can therefore create differences in prior beliefs.

Contextual effects are also found in the illusions outlined in the previous section. In the rubber hand illusion, if the object placed in front of the participant is non-corporeal (Tsakiris et al., 2010a), is a rubber hand of a different skin tone (Farmer et al., 2012) or is placed in a spatially incongruent location (Bekrater-Bodmann et al., 2012; Cadieux et al., 2011; Costantini and Haggard, 2007; Folegatti et al., 2012; Hohwy and Paton, 2010; Holle et al., 2011; Makin et al., 2008), the sense of ownership over the object is modulated or not present at all. So it is clear that the context within which stimuli are perceived modulates multisensory expectations about stimuli, leading to variability in the likelihood that another body will be processed as "me".

How can self-recognition be modified by contextual effects within the free-energy account? In the free energy principle, emphasis is placed on the importance of top-down probabilistic priors processed before a sensory event. In each of the contextual effects that is outlined above, it is information prior to the presence of a self-stimulus or to visuo-tactile stimulation that is modulating the sense of ownership and recognition of a stimulus once it is perceived. Thus, objects or faces that are placed within peripersonal space that are congruent in terms of their physical properties with the learned probabilistic representation of bodies or faces, are more likely to be recognised and processed as self. When the object violates contexts that are a necessity for objects to be processed as "me" e.g. when the hand is spatially incongruent to the body, this results in a low probability that the objects will be labelled as "self" even when there is congruent synchronous tactile stimulation. This is consistent with the view that contextual priors significantly influence whether a novel stimulus, such as a rubber hand or another's face, whether an object will be adopted into the model of the body.

To summarise, the free-energy account argues that information prior to an event will nuance predictions about the likely sensory input, and when sensory input is received, the prior information biases the probabilistic inferences that are made causes of an event. Here, we have suggested that contextual effects can modulate the processing of stimuli, by influencing expectations before a sensory event is perceived and also modulate the inferences that are made following the perception of a self-stimulus.

6. The anatomy of the self

So far we have argued that self-stimuli are recognised as "me" when surprise in one sensory system is explained away at a multimodal node which processes information from a system in which there is minimal surprise. The information processed at multimodal nodes will therefore be highly abstract, i.e., they will process high level prior information about self-stimuli and explain away surprise in unimodal sensory systems by labelling a stimulus as "me". This view is therefore predicated on three assumptions about functional anatomy that will be recruited during the processing of self-stimuli. First, self-stimuli in any one domain will activate portions of a sensory system which are also engaged when processing non-self stimuli. Second, self-stimuli presented in one modality will engage multisensory areas that are involved in processing high-level empirical prior information about the stimulus being "self" i.e. multisensory areas will be activated during self-recognition. Third, when a self-stimulus leads to predictable sensory input, activity in response to the predicted stimulus will be suppressed due to the absence of prediction error.

It should be noted that this is not the first discussion of Free-energy in the context of self-other information processing. Previously, Kilner and colleagues (Friston et al., 2010; Kilner, 2011; Kilner et al., 2007a, b) have offered a thorough account of how the free-energy framework can provide a compelling explanation of how the intentions and actions of others are learned within the mirror

neuron system, through the predictive simulation of the others' corollary discharge, i.e. predicting the kinematics of others' actions. Here, we extend beyond their work which discussed how we process the corollary discharge of others, and discuss how any self-stimuli may come to be labelled "self" and therefore not "other". Thus, whilst we make similar neural predictions, our discussion relates to how self-stimuli are processed and not how other-stimuli are processed. Furthermore, we make claims that more broadly define how stimuli are labelled as "self" in situations where actions are not a driving factor for the minimisation of free-energy.

6.1 Bottom-up non-self specific processing

Our first prediction was that no unimodal areas will exclusively be engaged by self-stimuli. In brief, if the emergence of self depends upon amodal high level empirical priors that send descending, parallel, and divergent predictions down the hierarchy, then the representation of self must be hierarchically distributed and recruit in all unimodal systems that register the consequences of self made acts. This precludes the possibility that a unimodal system will be exclusively engaged by self made stimuli.

A large body of neuroimaging research has investigated the functional anatomy of self-recognition. Such studies suggest that a broad range of unimodal areas are engaged during self-recognition. Self-face, voice and body recognition has been shown to activate regions within the core face processing network including: the inferior occipital gyrus, the fusiform gyrus, as well as face selective areas in the superior temporal sulcus (Kaplan et al., 2008; Platek et al., 2006; Platek et al., 2008; Sugiura et al., 2008; Uddin et al., 2005; Verosky and Todorov, 2010), portions of the auditory system in the superior temporal gyrus (Kaplan et al., 2008) and portions of extrastriate cortex that process body parts (Sugiura et al., 2006; Vocks et al., 2010) . However, it is well established that these areas process information about all faces, bodies or voices, even if the profile of their response is different between self and other (Barraclough and Perrett, 2011; Belin and Zatorre, 2003; Formisano et al.,

2008; Grill-Spector et al., 2004; Kanwisher and Barton, 2011; Myers and Sowden, 2008; Perrett et al., 1992; Perrett et al., 1982; Pitcher et al., 2009; Pitcher et al., 2011; Pitcher et al., 2007; Vocks et al., 2010; von Kriegstein et al., 2005). This is consistent with the view that these regions are not processing anything that is specific to the self. The increased activity in unimodal areas when processing self-stimuli may therefore reflect the surprise evoked by self-stimuli that is passed up sensory hierarchies to multisensory areas i.e. self-stimuli may evoke more surprise than non-self-stimuli that needs to be explained away by multimodal top-down effects

6.2 The multisensory self

Our second prediction was that self-recognition of a stimulus in one modaility and self-recognition during multisensory stimulation will engage multimodal areas of the brain. Neuroimaging studies suggest that unimodal self face, voice and body recognition activates a broad range of multimodal areas. The list of areas includes: the posterior cingulate gyrus, the anterior cingulate gyrus (ACC), medial portions of the superior frontal gyrus / paracingulate cortex, the temporo-parietal junction (TPJ), the superior temporal sulcus (STS), the temporal poles, the hippocampus, the anterior insula (AI), mid-portions of the inferior frontal gyrus (IFG), the middle frontal gyrus (MFG), the intraparietal sulcus (IPS) and the inferior parietal lobule (Apps et al., 2012b; Devue and Bredart, 2011; Devue et al., 2007; Heinisch et al., 2011; Kaplan et al., 2008; Morita et al., 2008; Pannese and Hirsch, 2011; Platek and Kemp, 2009; Platek et al., 2009; Platek et al., 2006; Platek et al., 2008; Ramasubbu et al., 2011; Sugiura et al., 2008; Sugiura et al., 2006; Sui et al., 2004; Taylor et al., 2009; Uddin et al., 2006; Verosky and Todorov, 2010). Some have therefore argued that information processed in some of these regions is self-specific and it is processing in these areas that leads to a self-concept (Northoff et al., 2006; Platek et al., 2008). However, there is little evidence that these regions are engaged exclusively by self-recognition processes and indeed, each of these regions is not found to be activated exclusively during the processing of self-stimuli (Legrand and Ruby, 2009)

Interestingly, several of these regions, including the TPJ, IPS, AI and the IFG are also activated when participants experience the rubber-hand illusion (Ehrsson et al., 2005; Ehrsson et al., 2004; Tsakiris et al., 2008; Tsakiris et al., 2007; Tsakiris et al., 2010b), and when participants experience similar multisensory illusions where the perceived spatial location of the whole body is manipulated (Ionta et al., 2011; Petkova et al., 2011). This evidence would seem to suggest that there is a core set of regions that are crucial for recognising different aspects of one's self. There is also evidence that posterior portions of the STS and adjacent portions of the supramarginal gyrus around the TPJ, both have strong connections to the IFG, IPS and to the AI (Mars et al., 2012; Petrides and Pandya, 2009; Seltzer and Pandya, 1989). There are also known to be connections between portions of the AI and the IFG (Mesulam and Mufson, 1982; Mufson and Mesulam, 1982; Petrides and Pandya, 2006), suggesting that these three regions may comprise a core circuit which is engaged when recognising one's self and creates a sense of ownership over one's body.

However, the connectional fingerprints of these regions and also the neuroimaging research examining their functional properties, suggest that these regions are not specialised for processing self-stimuli. The TPJ has afferent and efferent connections from the inferior temporal gyrus, caudal portions of the superior temporal gyrus, ventral portions of the premotor cortex and the anterior cingulate gyrus (Fletcher et al., 1995; Mars et al., 2012; Petrides and Pandya, 2009; Vogt and Pandya, 1987). Respectively, these regions are engaged by visual (Li et al., 1993), aural (Friederici, 2002), motor (Dum and Strick, 1991; Gallese et al., 1996) and social (Apps et al., 2012a; Apps et al., in press) information. The AI is sometimes referred to as primary interoceptive cortex due to its receipt of interoceptive signals from the body. This area also receives projections from the amygdala (Mesulam and Mufson, 1982; Mufson and Mesulam, 1982), the anterior cingulate gyrus (Augustine, 1996; Vogt and Pandya, 1987), primary and secondary somatosensory cortices (Augustine, 1996) and portions of the orbitofrontal cortex (Haber et al., 1995; Morecraft et al., 1992). The AI therefore receives information about the internal state of the body (Craig, 2009; Seth et al., 2011), emotions (Adolphs,

2002; Phillips et al., 2003), social information (Apps et al., 2012a), tactile stimulation (Avillac et al., 2005) and rewards (Schultz et al., 2000). The IPS has connections to primary and secondary somatosensory areas , to anterior and midportions of the IFG and the MFG, to posterior parietal cortex and to the inferior and superior temporal sulcus (Pandya et al., 1981; Petrides and Pandya, 1999; Schmahmann and Pandya, 1992; Seltzer and Pandya, 1980, 1984, 1986, 1994). Thus, the IPS has connections to areas involved in processing vestibular, somatosneosry , visuo-spatial (see Blanke, 2012 for a review) and abstract motor information (Passingham et al., 2002). Finally the IFG has connections to several areas of the motor system, including the Supplementary motor areas (SMA), the Cingulate Motor Areas (CMAs) and primary motor cortex (Petrides and Pandya, 1999; Picard and Strick, 1996), as well as portions of the frontal lobe and the cerebellum engaged in abstract mappings from cognitive rules to motor plans (Kelly and Strick, 2003; Ramnani, 2006). However, it also receives projections from portions of the parietal lobe engaged by vestibular and tactile information (Avillac et al., 2005; Bremmer et al., 2002).

These connectional fingerprints are indicative of distinct functional properties, with the TPJ processing the confluence of visual information and bodily related information, the AI processing the confluence of emotional, interoceptive and motor information about the body, the IPS processing visuo-spatial information about somatosensory input to the body and the IFG processing the mappings between abstract rules and the body. However, despite their functional discrepancies, the three regions are unified by the fact that none of them can be considered us processing unimodal sensory information and also, that their functional processing extends beyond the processing of self-related stimuli. We therefore argue that these regions are important for the process of recognising a stimulus as self and integrating multisensory information, in the form of abstract empirical priors, in order to explain away unimodal sensory surprise as “me”. However, processing in these areas is not specific to the self, rather self-stimuli evoke surprise that must be explained away by processing in these regions.

6.3 Self-predictive activity

Our third prediction was that that there will be a suppression of activity when a self-stimulus is predicted or when a self-stimulus leads to the expectation of a sensory event. Evidence is provided of such a notion by research examining self-touch. A seminal paper by Blakemore et al.,(Blakemore et al., 2000; Blakemore et al., 1998) found that participants cannot experience a tickling sensation when they apply tactile stimulation to their own skin, only when tactile stimulation is externally delivered. Functional imaging research has highlighted that this effect is purportedly driven by reduced activity in primary somatosensory cortex (SI), a unimodal area, when receiving self-delivered tactile stimulation (Blakemore et al., 1999a; Blakemore et al., 2000; Blakemore et al., 1999b). Other studies have also shown that self-generated actions result in attenuation in sensory systems, as compared to exogenously cued events that result in identical sensory events to the body. Similarly it has been shown that looking at one's own face whilst experiencing touch reduces activity in the (unimodal) somtosensory cortex and the multisensory Inferior Frontal Gyrus (IFG) , as compared to the activity evoked by looking at another's face and experiencing touch (Cardini et al., 2011). This is consistent with the view that a self-stimulus can act as a prediction of another sensory event in another modality. A multisensory node, in this case the IFG, instantiates a prediction of the upcoming stimulus reducing activation when that stimulus is presented. These studies point to how activity in unimodal and multisensory areas can become suppressed when a self-related stimulus results in a predictable sensory input

7. Future Directions and Caveats

The aim of this article was to provide a new theoretical perspective on the cortical mechanisms that underlie self-recognition, in order to account for previous findings and provide a novel framework for future research. In doing so, we have largely looked for and reported evidence that supports some of the principles of free-energy and elsewhere have simply stated what this theory would

assume, with limited direct evidence. In this section, we raise some caveats related to this discussion. In conjunction, we also discuss how future research can begin to test the assumptions that we have made.

Perhaps the most important caveat is that there is, as yet, little direct evidence that predictions and prediction errors related to self-stimuli are processed anywhere in the brain. This may, at first, seem a somewhat damming appraisal of a core component of our theory. However, whilst there is little direct evidence that supports the claims, there is also none that refute it. At this point, to the best of our knowledge, there has not been a paper that has tested whether the brain processes information about self-stimuli in a manner that conforms to the principles of predictive coding. The lack of evidence is in part due to the novelty of this theory, but also due to the fact that the theory is hard to falsify using standard neuroscientific methods.

Why is the theory hard to falsify? The difficulty arises due to the fact that the BOLD signal measured using 3Tesla fMRI scanners, which are those most commonly employed in research, is likely to be a function of both prediction and prediction error neurons. Whilst each type of neuron is, according to the theory (Friston and Kiebel, 2009b) and also recent evidence examining connectivity in cortical microcircuits (Bastos et al., 2012), found in separate cortical layers, the BOLD signal measured in these commonly used MRI scanners will likely be a function of both sets of neuronal types. As a result, a highly surprising event and a highly predicted event may, although not necessarily (Alink et al., 2010), both evoke large responses in populations of neurons that typically cannot be resolved, using a 3T fMRI scanner. Typical fMRI scanning may therefore suffer from an inability to detect self-recognition related free-energy responses. However, recently, studies using carefully considered designs in which error-related and probabilistic information are orthogonal (Egner et al., 2010), or using advanced analysis techniques such as multivoxel pattern analysis, have been able to identify activity related to these two distinct processes (de Gardelle et al., in press).Future research could use such techniques, in conjunction with high-field fMRI which affords the spatial resolution

necessary to distinguish between the BOLD signal evoked by neuronal responses in different layers of the cortex. This could further elucidate whether regions such as the TPJ, receive error signals from unimodal sensory areas of the cortex and whether information in the representational units is then projected back down the hierarchy during self-recognition.

Is the self-probabilistic? At the core of the free-energy is the concept that information is processed in the form of probability distributions which form the basis of prior beliefs before a sensory event and on which perceptual inferences are made after sensory stimulation. In several sections we have provided the case in support of the notion that one's own face voice and body are probabilistically the most likely to be recognised as one's own, but also that other features can be recognised as one's own. However, there has yet not been a direct test of whether the likelihood of stimuli being accepted as a part of one's body is probabilistic in nature, whether one's own body is represented probabilistically as "self" and not deterministically, and whether recognition and ownership of others' body parts are acquired during illusions in a manner that conforms to our predictive coding account of plasticity in the representation of another as "me". To tackle such issues, future research could use behavioural tasks in conjunction with predictive coding derived computational models (Egner et al., 2010; Huys et al., 2011; Rushworth et al., 2009). Such an approach affords the opportunity of testing whether self-recognition is truly probabilistic and Bayesian as we have suggested here.

Finally, we have suggested that self-recognition may be accounted for by the interaction between bottom-up and top-down processes interacting in a predictive coding framework. Whilst there is anatomical evidence to support the notion that multimodal areas may play an important role in integrating information across systems to explain away bottom-up unimodal surprise, there is little functional evidence. How might such functional evidence be provided? As stated before, the use of computational models derived from predictive coding, in conjunction with self-other designs, may provide a way forward to identify the functional anatomy. Applying novel functional imaging

techniques, such as Dynamic Causal Modelling (Friston et al., 2003), to examine the effective connectivity in these regions, may allow for the examination of whether self-stimuli induced error signals are projected up the hierarchy and whether prior probabilistic predictions about self-stimuli are projected down the hierarchy.

8. Summary

In this article, we have attempted to illuminate how the free-energy principle may account for self-recognition. To conform to the principles of predictive coding and free-energy, we have suggested 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. Such associations will be probabilistic such that one's own body is the most likely to be one's own. Representations of one's body are also plastic, such that other objects can become, probabilistically, more likely to be a part of one's self. Whether an object will be processed as self will also be based on contextual information that modulates prior beliefs about the likelihood that an object could be "me". Such processing will be processed hierarchically with multimodal areas processing the confluence of "self" information from different sensory systems and explaining away the surprising incoming sensory information from unimodal areas. However, within the free-energy framework that such processing will not be tied specifically to any particular circuit of brain. Rather, depending on the context and task within which the body is processed, any processing in any sensory system may ultimately be able to modulate how one's face is perceptually experienced.

In this article we have provided evidence in support of our theory, although we note that to date, empirical data neither largely supports nor refutes our account of self-recognition. However, this work does provide a broad range and extensive set of prediction about the nature of self-recognition

that can be tested empirically. We hope that such empirical investigation will generate important and novel findings that elucidate more about the neural and psychological basis of self-processing.

Acknowledgements

This study was funded by the European Research Council Starting Investigator Grant (ERC-2010-StG-262853) to MT. The authors would like to thank Dr. Lara Maister for useful discussions during the preparation of the manuscript and the two reviewers for helping us nuance our arguements.

References

Adolphs, R., 2002. Neural systems for recognizing emotion. Current Opinion in Neurobiology 12, 169-177.
Alink, A., Schwiedrzik, C.M., Kohler, A., Singer, W., Muckli, L., 2010. Stimulus Predictability Reduces Responses in Primary Visual Cortex. Journal of Neuroscience 30, 2960-2966.
Apps, M.A.J., Balsters, J.H., Ramnani, N., 2012a. The anterior cingulate cortex: Monitoring the outcomes of others' decisions. Social neuroscience 7, 424-435.
Apps, M.A.J., Green, R., Ramnani, N., in press. Reinforcement learning signals in the anterior cingulate cortex code for others' false beliefs. Neuroimage.
Apps, M.A.J., Tajadura-Jimenez, A., Turley, G., Tsakiris, M., 2012b. The different faces of one's self: An fMRI study into the recognition of current and past self-facial appearances. Neuroimage 63, 1720-1729.
Augustine, J.R., 1996. Circuitry and functional aspects of the insular lobe in primates including humans. Brain Research Reviews 22, 229-244.
Avillac, M., Deneve, S., Olivier, E., Pouget, A., Duhamel, J.R., 2005. Reference frames for representing visual and tactile locations in parietal cortex. Nature Neuroscience 8, 941-949.
Ballard, D.H., Hayhoe, M.M., Pook, P.K., Rao, R.P.N., 1997. Deictic codes for the embodiment of cognition. Behavioral and Brain Sciences 20.
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.
Barraclough, N.E., Perrett, D.I., 2011. From single cells to social perception. Philosophical Transactions of the Royal Society B-Biological Sciences 366, 1739-1752.
Bekrater-Bodmann, R., Foell, J., Diers, M., Flor, H., 2012. The perceptual and neuronal stability of the rubber hand illusion across contexts and over time. Brain Research 1452.
Belin, P., Zatorre, R.J., 2003. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14.
Bertenthal, B.I., Fischer, K.W., 1978. Development of self-recognition in infant. Developmental Psychology 14, 44-50.
Blake, R., Logothetis, N.K., 2002. Visual competition. Nature Reviews Neuroscience 3.
Blakemore, S.J., Frith, C.D., Wolpert, D.M., 1999a. Spatio-temporal prediction modulates the perception of self-produced stimuli. Journal of Cognitive Neuroscience 11.
Blakemore, S.J., Wolpert, D., Frith, C., 2000. Why can't you tickle yourself? Neuroreport 11, R11-R16.
Blakemore, S.J., Wolpert, D.M., Frith, C.D., 1998. Central cancellation of self-produced tickle sensation. Nature Neuroscience 1, 635-640.
Blakemore, S.J., Wolpert, D.M., Frith, C.D., 1999b. The cerebellum contributes to somatosensory cortical activity during self-produced tactile stimulation. Neuroimage 10.
Blanke, O., 2012. Multisensory brain mechanisms of bodily self-consciousness. Nature Reviews Neuroscience 13.
Blanke, O., Metzinger, T., 2009. Full-body illusions and minimal phenomenal selfhood. Trends in Cognitive Sciences 13.
Botvinick, M., Cohen, J., 1998. Rubber hands 'feel' touch that eyes see. Nature 391.
Bredart, S., 2004. Cross-modal facilitation is not specific to self-face recognition. Consciousness and Cognition 13, 610-612.
Bremmer, F., Klam, F., Duhamel, J.R., Ben Hamed, S., Graf, W., 2002. Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). European Journal of Neuroscience 16, 1569-1586.
Brooksgunn, J., Lewis, M., 1984. The development of early visual self-recognition. Developmental Review 4, 215-239.

Brown, H., Friston, K.J., 2012. Free-energy and illusions: the cornsweet effect. Frontiers in psychology 3.
Cadieux, M.L., Whitworth, K., Shore, D.I., 2011. Rubber hands do not cross the midline. Neuroscience Letters 504.
Cardini, F., Costantini, M., Galati, G., Romani, G.L., Ladavas, E., Serino, A., 2011. Viewing One's Own Face Being Touched Modulates Tactile Perception: An fMRI Study. Journal of Cognitive Neuroscience 23.
Clark, A., in press. Whatver next? Predictive brains, situated agents and the future of cognitive science. Behavioural Brain Sciences.
Costantini, M., Haggard, P., 2007. The rubber hand illusion: Sensitivity and reference frame for body ownership. Consciousness and Cognition 16.
Craig, A.D., 2009. How do you feel - now? The anterior insula and human awareness. Nature Reviews Neuroscience 10, 59-70.
de Gardelle, V., Waszscuk, M., Egner, T., Summerfield, D., in press. Concurrent Repetition Enhancement and Suppression Responses in Extrastriate Visual Cortex. Cerebral Cortex.
Devue, C., Bredart, S., 2008. Attention to self-referential stimuli: Can I ignore my own face? Acta Psychologica 128.
Devue, C., Bredart, S., 2011. The neural correlates of visual self-recognition. Consciousness and Cognition 20, 40-51.
Devue, C., Collette, F., Balteau, E., Dequeldre, C., Luxen, A., Maquet, P., Bredart, S., 2007. Here I am: The cortical correlates of visual self-recognition. Brain Research 1143, 169-182.
Devue, C., Van der Stigchel, S., Bredart, S., Theeuwes, J., 2009. You do not find your own face faster; you just look at it longer. Cognition 111, 114-122.
Dum, R.P., Strick, P.L., 1991. The origin of corticospinal projections from the premotor areas in the frontal-lobe. Journal of Neuroscience 11, 667-689.
Egner, T., Monti, J.M., Summerfield, C., 2010. Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. Journal of Neuroscience 30, 16601-16608.
Ehrsson, H.H., 2007. The experimental induction of out-of-body experiences. Science 317.
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. Journal of Neuroscience 25.
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.
Farmer, H., Tajadura-Jimenez, A., Tsakiris, M., 2012. Beyond the colour of my skin: How skin colour affects the sense of body-ownership. Consciousness and cognition 21.
Feinberg, T.E., Keenan, J.P., 2005. Where in the brain is the self? Consciousness and Cognition 14, 661-678.
Fletcher, P.C., Happe, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S.J., Frith, C.D., 1995. Other minds in the brain - a functional imaging study of theory of mind in story comprehension. Cognition 57, 109-128.
Folegatti, A., Farne, A., Salemme, R., de Vignemont, F., 2012. The Rubber Hand Illusion: Two's a company, but three's a crowd. Consciousness and Cognition 21, 799-812.
Formisano, E., De Martino, F., Bonte, M., Goebel, R., 2008. "Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech. Science 322.
Fotopoulou A (2012) Towards a psychodymamic neuroscience. In Fotopoulou A, Pfaff D & Conway MA (Eds) From the Couch to the Lab: trends in psychodynamic neuroscience. Oxford University Press, pp.25-48
Frassinetti, F., Mainil, M., Romualdil, S., Galante, E., Avanzi, S., 2008. Is it mine? Hemispheric asymmetries in corporeal self-recognition. Journal of Cognitive Neuroscience 20.
Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. Trends in Cognitive Sciences 6, 78-84.

Friston, K., 2005. A theory of cortical responses. Philosophical Transactions of the Royal Society B-Biological Sciences 360.
Friston, K., 2008a. Hierarchical Models in the Brain. Plos Computational Biology 4.
Friston, K., 2009. The free-energy principle: a rough guide to the brain? Trends in Cognitive Sciences 13, 293-301.
Friston, K., 2010a. Is the free-energy principle neurocentric? Nature Reviews Neuroscience 11.
Friston, K., 2010b. The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 11.
Friston, K., 2012a. Prediction, perception and agency. International Journal of Psychophysiology 83.
Friston, K., 2012b. The history of the future of the Bayesian brain. Neuroimage 62.
Friston, K., Adams, R.A., Perrinet, L., Breakspear, M., 2012a. Perceptions as hypotheses: saccades as experiments. Frontiers in psychology 3.
Friston, K., Ao, P., 2012. Free Energy, Value, and Attractors. Computational and Mathematical Methods in Medicine.
Friston, K., Breakspear, M., Deco, G., 2012b. Perception and self-organized instability. Frontiers in Computational Neuroscience 6.
Friston, K., Kiebel, S., 2009a. Cortical circuits for perceptual inference. Neural Networks 22.
Friston, K., Kiebel, S., 2009b. Predictive coding under the free-energy principle. Philosophical Transactions of the Royal Society B-Biological Sciences 364.
Friston, K., Mattout, J., Kilner, J., 2011. Action understanding and active inference. Biological Cybernetics 104.
Friston, K., Thornton, C., Clark, A., 2012c. Free-energy minimization and the dark-room problem. Frontiers in psychology 3.
Friston, K.J., 2008b. Perception, attention and memory: A free-energy formulation. International Journal of Psychology 43.
Friston, K.J., Daunizeau, J., Kilner, J., Kiebel, S.J., 2010. Action and behavior: a free-energy formulation. Biological Cybernetics 102.
Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. Neuroimage 19, 1273-1302.
Gagnepain, P., Henson, R.N., Davis, M.H., 2012. Temporal Predictive Codes for Spoken Words in Auditory Cortex. Current Biology 22.
Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. Brain 119, 593-609.
Gallup, G.G., 1970. Chimpanzees . Self-recognition. Science 167, 86-&.
Gillihan, S.J., Farah, M.J., 2005. Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. Psychological Bulletin 131, 76-97.
Grill-Spector, K., Knouf, N., Kanwisher, N., 2004. The fusiform face area subserves face perception, not generic within-category identification. Nature Neuroscience 7.
Haber, S.N., Kunishio, K., Mizobuchi, M., Lyndbalta, E., 1995. The orbital and medial prefrontal circuit through the primate basal ganglia. Journal of Neuroscience 15, 4851-4867.
Heinisch, C., Dinse, H.R., Tegenthoff, M., Juckel, G., Bruene, M., 2011. An rTMS study into self-face recognition using video-morphing technique. Social Cognitive and Affective Neuroscience 6, 442-449.
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.
Hohwy, J., Roepstorff, A., Friston, K., 2008. Predictive coding explains binocular rivalry: An epistemological review. Cognition 108, 687-701.
Holle, H., McLatchie, N., Maurer, S., Ward, J., 2011. Proprioceptive drift without illusions of ownership for rotated hands in the "rubber hand illusion" paradigm. Cognitive Neuroscience 2.
Huys, Q.J.M., Moutoussis, M., Williams, J., 2011. Are computational models of any use to psychiatry? Neural Networks 24, 544-551.

Ionta, S., Heydrich, L., Lenggenhager, B., Mouthon, M., Fornari, E., Chapuis, D., Gassert, R., Blanke, O., 2011. Multisensory Mechanisms in Temporo-Parietal Cortex Support Self-Location and First-Person Perspective. Neuron 70.
Kanwisher, N., Barton, J., 2011. The Functional Architecture of the Face System: Integrating Evidence from fMRI and Patient Studies. In: Haxby, J., Johnson, M., Rhodes, G., Calder, A. (Eds.), Handbook of Face Perception. Oxford University Press, Oxford, pp. 111-130.
Kaplan, J.T., Aziz-Zadeh, L., Uddin, L.Q., Iacoboni, M., 2008. The self across the senses: an fMRI study of self-face and self-voice recognition. Social Cognitive and Affective Neuroscience 3, 218-223.
Keenan, J.P., Freund, S., Hamilton, R.H., Ganis, G., Pascual-Leone, A., 2000. Hand response differences in a self-face identification task. Neuropsychologia 38, 1047-1053.
Keenan, J.P., McCutcheon, B., Freund, S., Gallup, G.G., Sanders, G., Pascual-Leone, A., 1999. Left hand advantage in a self-face recognition task. Neuropsychologia 37, 1421-1425.
Kelly, R.M., Strick, P.L., 2003. Cerebellar loops with motor cortex and prefrontal cortex of a nonhuman primate. Journal of Neuroscience 23, 8432-8444.
Kilner, J.M., 2011. More than one pathway to action understanding. Trends in Cognitive Sciences 15.
Kilner, J.M., Friston, K.J., Frith, C.D., 2007a. Predictive coding: an account of the mirror neuron system. Cognitive processing 8.
Kilner, J.M., Friston, K.J., Frith, C.D., 2007b. The mirror-neuron system: a Baynesian perspective. Neuroreport 18.
Kircher, T.T.J., Senior, C., Phillips, M.L., Rabe-Hesketh, S., Benson, P.J., Bullmore, E.T., Brammer, M., Simmons, A., Bartels, M., David, A.S., 2001. Recognizing one's own face. Cognition 78.
Lee, T.S., Mumford, D., 2003. Hierarchical Bayesian inference in the visual cortex. Journal of the Optical Society of America a-Optics Image Science and Vision 20.
Legrand, D., Ruby, P., 2009. What Is Self-Specific? Theoretical Investigation and Critical Review of Neuroimaging Results. Psychological Review 116.
Lenggenhager, B., Tadi, T., Metzinger, T., Blanke, O., 2007. Video ergo sum: Manipulating bodily self-consciousness. Science 317, 1096-1099.
Leopold, D.A., Logothetis, N.K., 1996. Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. Nature 379.
Li, L., Miller, E.K., Desimone, R., 1993. The representation of stimulus-familiarity in anterior inferior temporal cortex. Journal of neurophysiology 69, 1918-1929.
Liew, s.-l., ma, y., han, s., aziz-zadeh, l., 2011. Who's Afraid of the Boss: Cultural Differences in Social Hierarchies Modulate Self-Face Recognition in Chinese and Americans. Plos One 6.
Lumer, E.D., Friston, K.J., Rees, G., 1998. Neural correlates of perceptual rivalry in the human brain. Science 280.
Ma, Y., Han, S., 2012. Is the Self Always Better than a Friend? Self-Face Recognition in Christians and Atheists. Plos One 7.
Makin, T.R., Holmes, N.P., Ehrsson, H.H., 2008. On the other hand: Dummy hands and peripersonal space. Behavioural Brain Research 191.
Mars, R.B., Sallet, J., Schueffelgen, U., Jbabdi, S., Toni, I., Rushworth, M.F.S., 2012. Connectivity-Based Subdivisions of the Human Right "Temporoparietal Junction Area": Evidence for Different Areas Participating in Different Cortical Networks. Cerebral Cortex 22.
Mazzurega, M., Pavani, F., Paladino, M.P., Schubert, T.W., 2011. Self-other bodily merging in the context of synchronous but arbitrary-related multisensory inputs. Experimental Brain Research 213.
Mesulam, M.M., Mufson, E.J., 1982. Insula of the old-world monkey .3. Efferent cortical output and comments on function. Journal of Comparative Neurology 212, 38-52.
Mitchell, R.W., 1993. Mental models of mirror-self-recognition - 2 theories. New Ideas in Psychology 11.
Morecraft, R.J., Geula, C., Mesulam, M.M., 1992. Cytoarchitecture and neural afferents of orbitofrontal cortex in the brain of the monkey. Journal of Comparative Neurology 323, 341-358.

Morita, T., Itakura, S., Saito, D.N., Nakashita, S., Harada, T., Kochiyama, T., Sadato, N., 2008. The role of the right prefrontal cortex in self-evaluation of the face: A functional magnetic resonance imaging study. Journal of Cognitive Neuroscience 20, 342-355.

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 of the United States of America 105, 13169-13173.

Mufson, E.J., Mesulam, M.M., 1982. Insula of the old-world monkey .2. Afferent cortical input and comments on the claustrum. Journal of Comparative Neurology 212, 23-37.

Myers, A., Sowden, P.T., 2008. Your hand or mine? The extrastriate body area. Neuroimage 42.

Northoff, G., Heinzel, A., de Greck, M., Bennpohl, F., Dobrowolny, H., Panksepp, J., 2006. Self-referential processing in our brain - A meta-analysis of imaging studies on the self. Neuroimage 31, 440-457.

Pandya, D.N., Vanhoesen, G.W., Mesulam, M.M., 1981. Efferent connections of the cingulate gyrus in the rhesus-monkey. Experimental Brain Research 42, 319-330.

Pannese, A., Hirsch, J., 2010. Self-specific priming effect. Consciousness and Cognition 19.

Pannese, A., Hirsch, J., 2011. Self-face enhances processing of immediately preceding invisible faces. Neuropsychologia 49.

Passingham, R.E., Stephan, K.E., Kotter, R., 2002. The anatomical basis of functional localization in the cortex. Nature Reviews Neuroscience 3, 606-616.

Perrett, D.I., Hietanen, J.K., Oram, M.W., Benson, P.J., 1992. Organization and functions of cells responsive to faces in the temporal cortex. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 335, 23-30.

Perrett, D.I., Rolls, E.T., Caan, W., 1982. Visual neurones responsive to faces in the monkey temporal cortex. Experimental Brain Research 47, 329-342.

Petkova, V.I., Bjornsdotter, 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.

Petkova, V.I., Ehrsson, H.H., 2008. If I Were You: Perceptual Illusion of Body Swapping. Plos One 3.

Petrides, M., Pandya, D.N., 1999. Dorsolateral prefrontal cortex: comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. European Journal of Neuroscience 11, 1011-1036.

Petrides, M., Pandya, D.N., 2006. Efferent association pathways originating in the caudal prefrontal cortex in the macaque monkey. Journal of Comparative Neurology 498, 227-251.

Petrides, M., Pandya, D.N., 2009. Distinct Parietal and Temporal Pathways to the Homologues of Broca's Area in the Monkey. Plos Biology 7.

Phillips, M.L., Drevets, W.C., Rauch, S.L., Lane, R., 2003. Neurobiology of emotion perception I: The neural basis of normal emotion perception. Biological Psychiatry 54, 504-514.

Picard, N., Strick, P.L., 1996. Motor areas of the medial wall: A review of their location and functional activation. Cerebral Cortex 6, 342-353.

Pitcher, D., Charles, L., Devlin, J.T., Walsh, V., Duchaine, B., 2009. Triple Dissociation of Faces, Bodies, and Objects in Extrastriate Cortex. Current Biology 19, 319-324.

Pitcher, D., Walsh, V., Duchaine, B., 2011. The role of the occipital face area in the cortical face perception network. Experimental Brain Research 209, 481-493.

Pitcher, D., Walsh, V., Yovel, G., Duchaine, B., 2007. TMS evidence for the involvement of the right occipital face area in early face processing. Current Biology 17, 1568-1573.

Platek, S.M., Kemp, S.M., 2009. Is family special to the brain? An event-related fMRI study of familiar, familial, and self-face recognition. Neuropsychologia 47, 849-858.

Platek, S.M., Krill, A.L., Wilson, B., 2009. Implicit trustworthiness ratings of self-resembling faces activate brain centers involved in reward. Neuropsychologia 47, 289-293.

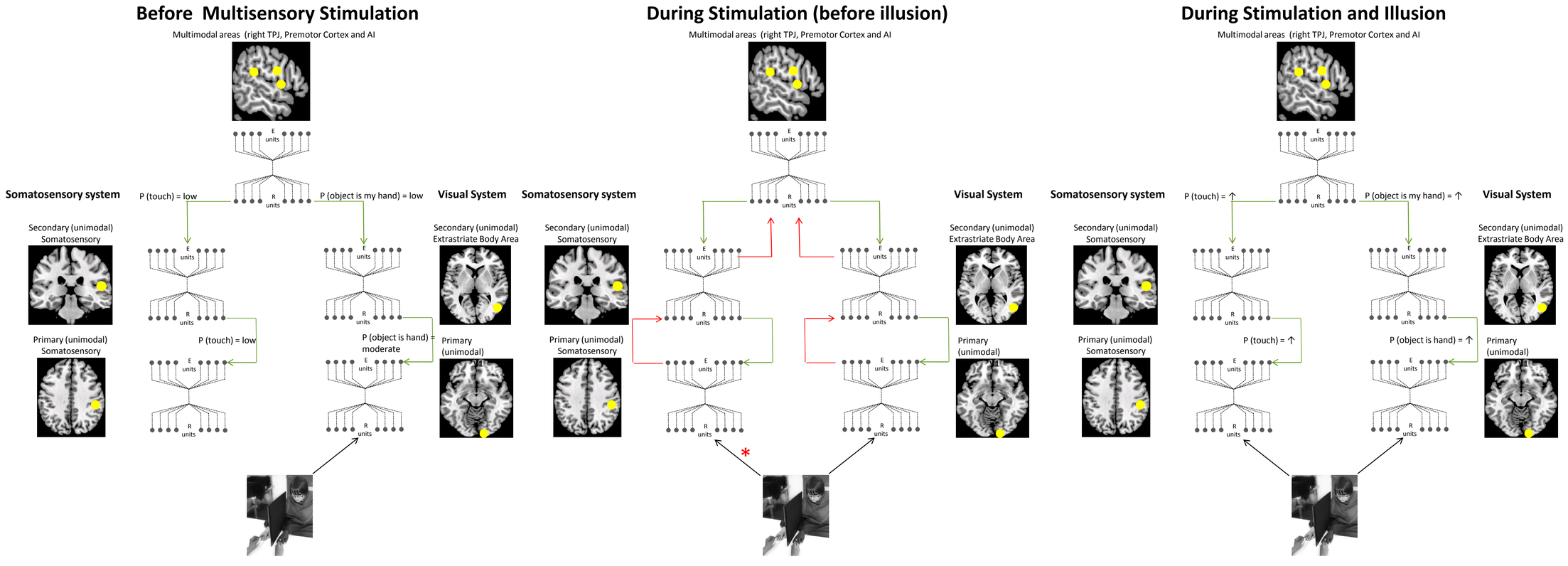
Platek, S.M., Loughead, J.W., Gur, R.C., Busch, S., Ruparel, K., Phend, N., Panyavin, I.S., Langleben, D.D., 2006. Neural substrates for functionally discriminating self-face from personally familiar faces. Human Brain Mapping 27, 91-98.
Platek, S.M., Thomson, J.W., Gallup, G.G., 2004. Cross-modal self-recognition: The role of visual, auditory, and olfactory primes. Consciousness and Cognition 13, 197-210.
Platek, S.M., Wathne, K., Tierney, N.G., Thomson, J.W., 2008. Neural correlates of self-face recognition: An effect-location meta-analysis. Brain Research 1232, 173-184.
Ramasubbu, R., Masalovich, S., Gaxiola, I., Peltier, S., Holtzheimer, P.E., Heim, C., Goodyear, B., MacQueen, G., Mayberg, H.S., 2011. Differential neural activity and connectivity for processing one's own face: A preliminary report. Psychiatry Research-Neuroimaging 194, 130-140.
Ramnani, N., 2006. The primate cortico-cerebellar system: anatomy and function. Nature Reviews Neuroscience 7, 511-522.
Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience 2.
Rauss, K., Schwartz, S., Pourtois, G., 2011. Top-down effects on early visual processing in humans: A predictive coding framework. Neuroscience and Biobehavioral Reviews 35, 1237-1253.
Reiss, D., Marino, L., 2001. Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. Proceedings of the National Academy of Sciences of the United States of America 98, 5937-5942.
Rotshtein, P., Henson, R.N.A., Treves, A., Driver, J., Dolan, R.J., 2005. Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nature Neuroscience 8, 107-113.
Rushworth, M.F.S., Mars, R.B., Summerfield, C., 2009. General mechanisms for making decisions? Current Opinion in Neurobiology 19, 75-83.
Schmahmann, J.D., Pandya, D.N., 1992. Course of the fiber pathways to pons from parasensory association areas in the rhesus-monkey. Journal of Comparative Neurology 326.
Schultz, W., Tremblay, L., Hollerman, J.R., 2000. Reward processing in primate orbitofrontal cortex and basal ganglia. Cerebral Cortex 10, 272-283.
Seltzer, B., Pandya, D.N., 1980. Converging visual and somatic sensory cortical input to the intraparietal sulcus of the rhesus-monkey. Brain Research 192.
Seltzer, B., Pandya, D.N., 1984. Further observations on parieto-temporal connections in the rhesus-monkey. Experimental Brain Research 55.
Seltzer, B., Pandya, D.N., 1986. Posterior parietal projections to the intraparietal sulcus of the rhesus-monkey. Experimental Brain Research 62.
Seltzer, B., Pandya, D.N., 1989. Frontal-lobe connections of the superior temporal sulcus in the rhesus-monkey. Journal of Comparative Neurology 281, 97-113.
Seltzer, B., Pandya, D.N., 1994. Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus-monkey - a retrograde tracer study. Journal of Comparative Neurology 343.
Seth, A.K., Suzuki, K., Critchley, H.D., 2011. An interoceptive predictive coding model of conscious presence. Frontiers in psychology 2.
Sforza, A., Bufalari, I., Haggard, P., Aglioti, S.M., 2010. My face in yours: Visuo-tactile facial stimulation influences sense of identity. Social Neuroscience 5.
Suarez, S.D., Gallup, G.G., 1981. Self-recognition in chimpanzees and orangutans, but not gorillas. Journal of Human Evolution 10, 175-188.
Sugiura, M., Sassa, Y., Jeong, H., Horie, K., Sato, S., Kawashima, R., 2008. Face-specific and domain-general characteristics of cortical responses during self-recognition. Neuroimage 42, 414-422.
Sugiura, M., Sassa, Y., jeong, H.J., Miura, N., Akitsuki, Y., Horie, K., Sato, S., Kawashima, R., 2006. Multiple brain networks for visual self-recognition with different sensitivity for motion and body part. Neuroimage 32, 1905-1917.

Sui, J., Liu, C.H., Han, S., 2009. Cultural difference in neural mechanisms of self-recognition. Social Neuroscience 4.
Sui, J., Zhu, Y., Han, S.H., 2004. Electrophysiological evidence for self processing: Effects of faces and words. International Journal of Psychology 39.
Summerfield, C., Egner, T., 2009. Expectation (and attention) in visual cognition. Trends in Cognitive Sciences 13, 403-409.
Tajadura-Jimenez, A., Grehl, S., Tsakiris, M., 2012. The Other in Me: Interpersonal Multisensory Stimulation Changes the Mental Representation of the Self. Plos One 7.
Taylor, M.J., Arsalidou, M., Bayless, S.J., Morris, D., Evans, J.W., Barbeau, E.J., 2009. Neural Correlates of Personally Familiar Faces: Parents, Partner and Own Faces. Human Brain Mapping 30.
Tong, F., Nakayama, K., Vaughan, J.T., Kanwisher, N., 1998. Binocular rivalry and visual awareness in human extrastriate cortex. Neuron 21.
Tsakiris, M., 2008. Looking for Myself: Current Multisensory Input Alters Self-Face Recognition. Plos One 3.
Tsakiris, M., 2010. My body in the brain: A neurocognitive model of body-ownership. Neuropsychologia 48, 703-712.
Tsakiris, M., Carpenter, L., James, D., Fotopoulou, A., 2010a. Hands only illusion: multisensory integration elicits sense of ownership for body parts but not for non-corporeal objects. Experimental Brain Research 204, 343-352.
Tsakiris, M., 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., Haggard, P., 2005. The rubber hand illusion revisited: Visuotactile integration and self-attribution. Journal of Experimental Psychology-Human Perception and Performance 31.
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.
Tsakiris, M., Longo, M.R., Haggard, P., 2010b. Having a body versus moving your body: Neural signatures of agency and body-ownership. Neuropsychologia 48, 2740-2749.
Tsakiris, M., Prabhu, G., Haggard, P., 2006. Having a body versus moving your body: How agency structures body-ownership. Consciousness and Cognition 15.
Uddin, L.Q., Kaplan, J.T., Molnar-Szakacs, I., Zaidel, E., Iacoboni, M., 2005. Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: an event-related fMRI study. Neuroimage 25, 926-935.
Uddin, L.Q., Molnar-Szakacs, I., Zaidel, E., Iacoboni, M., 2006. rTMS to the right inferior parietal lobule disrupts self-other discrimination. Social Cognitive and Affective Neuroscience 1, 65-71.
Verosky, S.C., Todorov, A., 2010. Differential neural responses to faces physically similar to the self as a function of their valence. Neuroimage 49, 1690-1698.
Vocks, S., Busch, M., Groenemeyer, D., Schulte, D., Herpertz, S., Suchan, B., 2010. Differential neuronal responses to the self and others in the extrastriate body area and the fusiform body area. Cognitive Affective & Behavioral Neuroscience 10.
Vogt, B.A., Pandya, D.N., 1987. Cingulate cortex of the rhesus-monkey .2. Cortical afferents. Journal of Comparative Neurology 262, 271-289.
von Kriegstein, K., Kleinschmidt, A., Sterzer, P., Giraud, A.L., 2005. Interaction of face and voice areas during speaker recognition. Journal of Cognitive Neuroscience 17.
Wacongne, C., Changeux, J.-P., Dehaene, S., 2012. A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. Journal of Neuroscience 32.
Winkler, I., Denham, S., Mill, R., Bohm, T.M., Bendixen, A., 2012. Multistability in auditory stream segregation: a predictive coding view. Philosophical Transactions of the Royal Society B-Biological Sciences 367.
Zheng, Z.Z., MacDonald, E.N., Munhall, K.G., Johnsrude, I.S., 2011. Perceiving a Stranger's Voice as Being One's Own: A 'Rubber Voice' Illusion? Plos One 6.

Figure Legend.

Fig.1 Predictive coding of 'surprise' and 'explaining away' during the RHI. The green lines indicate top-down predictive information explaining away the bottom-up unexplained prediction errors or surprise indicated by the red lines. The solid black lines indicate a sensory input. In predictive coding this architecture dynamically reconciles predictive and unexpected information when a sensory event is unexpected. In all three panels the information is organised hierarchically within the sensory systems and this information converges on multimodal areas. In the left panel, before synchronous stimulation, the sensory input to the visual system has instantiated predictions in the visual system that one is seeing a rubber hand that has a low probability of being a real hand (but higher than a non-corporeal object) and has a low probability of being one's own hand (but higher than if the hand was placed in a spatial location that is removed from the body). In the middle panel, the experience of touch evokes surprise in the somatosensory system and its temporal and specular congruency with touch on the rubber hand causes surprise in the visual system. This surprise is explained away by the top-down influence from multimodal areas and perceptual learning processes in the unimodal areas. As a result, the probability that the visually perceived rubber hand is part of 'my' body and is also a real hand increases. In parallel, the probability that this object is part of the body updates the probability that touch on the rubber hand will result in a somatosensory experience. As such, during the experience of the illusion (right panel) touch on the rubber hand is no longer surprising, as the object is perceived visually as part of one's body and it is an object that touch upon evokes a somatosensory event.

Figure


Before Multisensory Stimulation
During Stimulation (before illusion)
During Stimulation and Illusion
Multimodal areas (right TPJ, Premotor Cortex and AI
Somatosensory system
Visual System
Secondary (unimodal) Somatosensory
Primary (unimodal) Somatosensory
Secondary (unimodal) Extrastriate Body Area
Primary (unimodal)
P (touch) = low
P (object is my hand) = low
P (touch) = low
P (object is hand) = moderate
P (touch) = ↑
P (object is my hand) = ↑
P (object is hand) = ↑
E units
R units
*