Time to get rid of the 'Modular' in neuropsychology: A unified theory of anosognosia as aberrant predictive coding

Aikaterini Fotopoulou

Division of Psychology and Language Sciences, University College London

Please address all correspondence to:

Aikaterini (Katerina) Fotopoulou

Division of Psychology and Language Science
Floors 3-5, 1-19 Torrington Place, London, WC1E 6BT
Email: a.fotopoulou@ucl.ac.uk

Keywords: cognitive neuropsychology; lesion studies; epistemology; anosognosia; predictive coding; free energy

Word count: 7440

Abstract

Cognitive neuroscience, being more inclusive and ambitious in scope than cognitive neuropsychology, seems to have taken the place of the latter within the modern neurosciences. Nevertheless, recent advances in the neurosciences afford neuropsychology with epistemic possibilities that simply did not exist even 15 years ago. Human lesion studies still have an important role to play in shaping such possibilities, particularly when combined with other methods of enquiry. I first outline theoretical and methodological advances within the neurosciences that can inform and shape the rebirth of a dynamic, non-modular neuropsychology. I then use an influential computational theory of brain function, the free energy principle, to suggest a unified account of anosognosia for hemiplegia as a research example of the potential for transition from a modular, cognitive neuropsychology to a dynamic, computational and even restorative neuropsychology. These and many other adjectives that can flexibly, take the place of 'cognitive' next to 'neuropsychology' will hopefully designate the much needed rebirth and demarcation of a field, neuropsychology itself, that has somehow lost its place within the modern neurosciences and yet seems to have a unique and important role to play in the future understanding of the brain.

The Emerging Need for Integrative Neurosciences and the Place of Neuropsychology

On the basis of the idea that structure determines function, one of the most enduring aims of neuroscience has been the association of anatomically parcelled brain areas with specific functions. There has been remarkable and speedy progress in this regard at many different levels of analysis, ranging from the study of single synapses to the role of entire brain regions in complex cognitive functions. However, following decades of data accumulation at smaller scales, it is becoming increasingly clear that understanding the complex relations between different brain subsystems will require large-scale, integrative theories of brain function across different levels of analysis and different sub-fields of neuroscience (Grillner et al., 2005; Friston 2009a; Mesulam, 2012).

Given this increasingly acknowledged need for integration in the neurosciences, one would anticipate that neuropsychology, the long tradition of interdisciplinary, empirical studies of the relationship between the damaged brain and cognition, would have a clear contributing role in contemporary neurosciences. However, this field seems to have lost its former, prominent place within the modern neurosciences. Nowadays there is another, wider and prolific field studying the mind-brain interface; it is most commonly referred to as 'cognitive neuroscience'. A complete account of the professional and societal trends that may explain this change escapes the scope of this paper. Here I will focus on epistemic issues, tracing differences in epistemology between cognitive neuropsychology and neuroscience and other related fields. I will then call for a new, dynamic neuropsychology that combines the epistemological advantages of the various fields, while avoiding some of their limitations. Finally, I will use the syndrome of anosognosia for hemiplegia as an example of how dynamic, computational and therapeutic approaches to neuropsychology can be explored.

The Epistemology of Human Lesion Studies

In most psychological and neuroscientific methods, researchers intervene with behaviour or brain function in a predetermined way and then measure the effects of their intervention. Temporary lesions of certain brain areas can be induced in a controlled manner, for example by using transcranial magnetic stimulation (TMS), but in traditional human lesion studies, it is injury or disease that 'intervenes' with the normal function of the brain (Bechtel, 2012). This fact limits the control the neuropsychologist has over the phenomena in question, because the 'intervention' on the brain itself, its effects on the mind and the relation of the two, are all unknown and demand careful characterisation. Thus, traditional neuropsychological research has at least three corresponding aims: (a) to identify and measure behavioural or cognitive deficits; (b) to localize brain lesions; and (c) most importantly *to infer* the functional role of certain brain areas on the basis of the functional consequences of their damage. There are intrinsic limitations around these three aims. For instance, behavioural testing following brain damage is always subject to assumptions about, or at best post-hoc estimations of, an individual's corresponding, premorbid abilities. Moreover, some regions of the brain are highly susceptible to damage, while others are rarely affected by injury or disease. In addition, cytoarchitectonic studies have long shown that there is gradual transition between cortical areas and their demarcation is not absolute. Lastly, inferring the normal functional role of a given brain area on the basis of the functional consequences of its damage, is far more complex than it sounds and it depends on several assumptions about normal brain function and its dysfunction.

Despite these intrinsic limitations however, human lesion studies have seen several methodological developments. In terms of the first aim, psychometrically rigorous neuropsychological measures progressively enhanced mere clinical observations and both were more recently complemented by behavioural experiments. The second aim, i.e.

localization and characterization of brain lesions, has also progressed dramatically from postmortem studies to 3-D structural imaging techniques. For example, improved structural imaging technology, specialised software and related statistical analysis methods have allowed better specification of the location and extent of damage to grey matter cells, as well as to white matter fiber tracts, in groups of patients suffering from behavioural syndromes such as neglect, or amnesia (e.g. Karnath et al., 2009).

It is however the third aim of neuropsychological studies, i.e. inferring the functional role of certain brain areas on the basis of the functional consequences of their damage that constitutes the most important challenge of the method and has sparked several debates in the history of neuropsychology (see Deacon, 1989; Muller, 1992 for historical reviews). For example, the many pendulum swings in the history of neuropsychology between localizationist and anti-localizationist theories have informed the two central principles of brain structure-function relations that we use today, namely the principles of functional specialisation, or segregation, and functional integration, or convergence. Functional segregation, the conceptual roots of which can be traced back to the localisationist theories of the 19th century and even Franz-Josef Gall's 18th century phrenology, refers to the idea of functionally specialized neurons, grouped together in space to form segregated brain areas responsible for discrete mental functions. Functional integration, the conceptual origins of which can be traced back to holistic and anti-localisationist theories such as those of Pierre-Marie Flourens, John Hughlings Jackson, Karl Lashley, Alexander Luria and even the pre-psychoanalytic writings of Sigmund Freud, posits that complex mental functions are based on interactions or connectivity patterns among various interconnected, functionally diverse and structurally distributed components of the nervous system.

The relation between these two principles continues to be specified by neuroanatomical studies, as well as studies in several neuroscientific disciplines (for review

see Cloutman & Lambon Ralph, 2012), including human lesion studies (e.g. Catani & Ffytche, 2005; Ween, 2008; Seghier et al., 2010). And yet in most contemporary human lesion studies, inferences about the role of specific brain areas in cognition seem to be based more on the principle of functional segregation than that of functional integration. This preference seems to relate to the research agenda of 'cognitive neuropsychology', the epistemological marriage of cognitive science, artificial intelligence theory and human lesion studies that was formalised by the 1980s (Caramazza, 1984; Coltheart, 1985; Ellis & Young, 1988; Shallice, 1988) and in many ways continues to dominate the field of human lesion studies today. This approach criticised the localisation of behavioural abilities in particular brain areas, arguing that it lacked a consideration of the many cognitive operations that may be involved in any given behaviour. For example, observing general memory difficulties following temporal lobe damage may not be sufficient to determine which core, mnemonic function this area serves. In order to accomplish such inference one needs to have an *a priori*, *cognitive* model of the organisation of memory and its core, functional components. According to this approach human lesion studies mainly serve to test such models.

This position derives from the logic of models of artificial intelligence and regards the brain as the hardware of the biological mind and the mind as the software of the system. More specifically, cognitive theories of the mind were based on philosophical ideas of computationalism and functionalism. These specific forms of cognitivism that were proposed by Hilary Putnam, and developed most notably by Jerry Fodor (1975; Fodor & Pylyshyn, 1988), argued that the mind operates like an information processing system; a Turing machine that transforms information by performing a series of purely formal operations on symbolic (linguistic) representations. Thus, according to the so-called 'ultra-cognitive' neuropsychologists, brain localisation (the hardware) is of little importance to the cognitive model in question (the software).

Moreover, as the mind is modular in its core conception, i.e. it is organised in computationally autonomous, encapsulated and mostly serially organised domains of function, brain damage can result to a selective and encapsulated impairment of a component of cognitive processing without affecting other components (Caramazza, 1984). On the basis of these assumptions, cognitive neuropsychology studies in the 80s and in the early 90s aimed to identify behavioural dissociations in single case studies or case series. These dissociations suggested new modular divisions in a plethora of "box diagrams" in which cognitive information was shown to follow paths along, serially organised modules, any of which described as serving a different, core cognitive function, usually increasing in complexity.

Eventually progress in artificial intelligence theory itself, as well as in neuroanatomy, functional neuroimaging and computational neuroscience (see below) has increasingly raised doubts over these models. Even as far back as the late 80s and early 90s, the so-called connectionist theories introduced the idea of distributed representations and parallel processing in cognitive neuropsychology (e.g. Plaut & Shallice, 1993). Despite however the fruits of such epistemic changes (Lambon Ralph, 2004) to this day many cognitive neuropsychologists adhere to the original epistemological principles of the field either implicitly (see Harley, 2004 for a critical review), or explicitly (Caramazza & Coltheart, 2006). Moreover, while some neuropsychologists welcome the insights of other neuroscientific methods, such as functional neuroimaging (e.g. Cooper & Shallice, 2010), others reject their application to neuropsychology as a neo-localizationist attempt to elucidate the mind-brain relation (Coltheart, 2006; Harley, 2004; Page, 2006). This adherence to outdated principles of mental and brain functioning, and the associated reluctance to engage fully with recent methodological developments in the neurosciences may be at least partly

responsible for the non-prominent position of neuropsychology among the contemporary neurosciences.

The Advent of Cognitive Neuroscience and Functional Neuroimaging

Although it would be a mistake to assume that cognitive neuroscience shares no epistemological assumptions with cognitive neuropsychology (see below), cognitive neuroscientists differ from traditional cognitive neuropsychologists in both the 'what' and the 'how' they study the mind-brain interface. The advent of powerful methods of investigating the neural basis of the mind *in vivo* have allowed cognitive neuroscientists to expand their enquiries to topics that far exceeded the traditional topics of neuropsychology, e.g. language, semantic processing and memory. Instead, topics such as emotion and empathy are now considered mainstream areas of cognitive neuroscience research. At the theoretical level, the assumption prevailing until the early 90s to the effect that the human mind can be understood by examining *exclusively* cognitive functions has undergone considerable criticism (see for example Fotopoulou, 2010; Fotopoulou et al., 2012). Following some extraordinary discoveries, e.g. mirror neurons in the macaque monkey (Di Pellegrino et al., 1992), and other similar insights, a diverse and growing community of researchers views mental abilities as defined *also* by emotions and motivation, as embedded in the acting, sensing and feeling body, and as subject to intricate couplings between organisms and their interpersonal, social and technological environments (e.g. Damasio, 1994; LeDoux, 1996; Panksepp, 1998; Rizzolatti & Craighero, 2004; Knoblich et al., 2006; Decety and Ickes, 2009; Benedetti, 2010; Frith & Frith, 2010).

Perhaps more important to the change that took place in 'what' cognitive neuroscientists study, is the dramatic developments in 'how' they study the brain, and thus what kind of knowledge about brain-mind relations they can arrive at. Cognitive neuroscience does not need to depend on insights from the injured brain as neuropsychology

does. Instead, there have been substantial technological and mathematical advances in observing, measuring and visualizing functions of the 'healthy brain'. Such methods do not need to engage with the complex issue of inferring normal cognition on the basis of the structurally damaged brain. Instead, the researchers try to measure and correlate activity in a brain region with mental tasks being performed simultaneously, after *intervening,* in a predetermined and controlled way with behaviour and cognition (e.g. presentation of stimuli), or brain function itself (e.g. with magnetic or electrical stimulation).

Of course, such methods have their own epistemological challenges of inference. Correlations between mental tasks and surrogate brain signals (e.g. BOLD) in functional neuroimaging studies, for example, provide only indirect evidence of the involvement of certain brain location in any given task. It remains uncertain whether this particular area is necessary for the mental ability in question, and perhaps even more importantly, the precise neurobiological mechanisms by which this and other locations interact to generate such mental functions cannot be specified by such methods alone. Initial applications of functional neuroimaging in cognitive neuroscience seemed to underplay these challenges. Instead, they put forward rather simplistic, strict localisationist and modular arguments about the role of certain brain areas in complex mental functions. For example, during the first years of functional magnetic resonance imaging (fMRI), relatively simple experimental paradigms and statistical models (e.g. categorical designs, such as blocked subtraction paradigms) were used to infer the role of brain areas in cognition. In striking agreement with some of the aforementioned modular assumptions about cognition, these paradigms assumed that a single cognitive process can be selectively 'elicited' through specific stimuli and then 'subtracted' by a given system without affecting the function of the rest of the cognitive processes in the system (assumption of 'pure insertion') (Friston, 1994). Such subtractions were expected to reveal the spatially distinct organisation of the particular function in the brain.

Whereas mapping certain sensory functions (e.g. visual fields) into functionally specialised and hierarchically organised areas in the human cortex (spatial segregation) can benefit from tools such as fMRI (Wandell et al., 2007), assuming that a similar kind of strictly modular and one-to-one mapping would apply to complex cognitive and emotional functions such as empathy or awareness seems to constitute a naive return not only to the extreme modularity of cognitive neuropsychology, but also to the strict localisational logic of the 19th century neurologists. Indeed, brain scientists in the 19th century dealt with the lack of neuroanatomical and neurophysiological knowledge about connectivity in the human cortex by crudely designating mental faculties to specific brain areas (the origins of localisationism and functional specialisation) and inferring oversimplified neural connections to mediate the logical relationships between these mental processes (the origins of connectionist and associationist theories) (Steinberg, 2009). Unfortunately, several studies in cognitive neuroscience still portray extreme neo-localisationist and simplistic associationist assumptions.

Progress in Integrative Neuroscience: Beyond Cognitivism

Progressive efforts in cognitive and computational neuroscience focused on acknowledging some of the limitations of these methods and where possible, improving their methodological potential and correcting their theoretical inferences (e.g. Logothetis, 2008; Friston, 2009b). Important developments included a change in emphasis from functional segregation to parallel consideration of functional integration and a focus on methods that capture the dynamic, large-scale operations in the brain. As aforementioned, dynamic, large-scale network operations in the brain have been long anticipated in anti-localisational and holistic theorists in clinical neurology and physiology. Nevertheless, the technology that would allow quantification and computational inference of such large-scale network dynamics was not hitherto available.

Today, our neuroanatomical and physiological methods for *observing* structural connectivity (Mesulam, 2012) and our neuroimaging and statistical methods for *inferring* computational connectivity (Friston, 1994; Valdes-Sosa et al., 2011) have improved since the time of the so-called 'diagram makers' of the 19th century. For example, several large-scale distributed functional networks have been identified that are not task specific (e.g. responsible for the perception of faces) but rather context-driven. Such networks involve, for example, responses to salience, be that salience cognitive, emotional or homeostatic (Seeley et al., 2007; Sridharan et al., 2008), or reflect autonomous brain dynamics during rest (e.g. Raichle et al., 2001). These studies suggest a marked change in perspective from the traditional stimulus-based studies of cognitive science, and emphasise self-organising endogenous brain activity. Furthermore, the recent application of connectivity analysis (e.g. Bayesian hierarchical modelling and dynamic causal modelling), as well as neural field models (e.g. Laing et al., 2002) to cognitive neuroscience and even neuropsychology (see below) constitutes an important and unprecedented step towards understanding dynamic function-structure relations. Of course, the characterisation of such dynamic processes can still only be approximated by current neuroimaging techniques and computational modelling. Increased insight will depend on what we can learn about connectivity from other fields such as neurodynamics and neurophysiology (e.g. see Freeman, 2003; Coombes, 2010; Fuster, 2009; Mesulam, 2012).

The Risk of 'Mindless' Reductionism

Taken together, the above developments increasingly convey a picture of neurocognitive organisation and function that surpasses the classical modular and computationalistic view of the mind as portrayed in cognitive sciences. Not surprisingly, these advances seem to have already taken the cognitive neuroscience community by storm, implicitly demanding that new epistemological criteria for cognitive theories are set, e.g.

modular, information-processing theories and computer metaphors are constantly re-evaluated (e.g. see Fuster, 2009; Piccinini & Scarantino, 2011). Nevertheless, as alternative psychological models that are capable of accommodating dynamic and complex mental processes are lacking within the models of classical cognitive psychology, simplistic notions on the nature of cognition and the localisation of complex mental functions in the brain are likely to persist for a few more years. There is encouraging progress in other fields, such as embodied cognition, psychodynamic and affective neuroscience and theoretical and computational neuroscience (see Fotopoulou, 2012b for review). However, assimilation of knowledge from these fields which use different psychological traditions (e.g. phenomenology, e.g. Varela et al., 1991; psychoanalysis, Fotopoulou, 2012b; Panksepp & Solms, 2012) and complex mathematical and statistical models, respectively, is likely to be slow.

It is perhaps not accidental that a large proportion of neuroimaging studies in cognitive neuroscience portray a return to behaviourism, or alternatively seem conceptualised in an atheoretical way. For example, several scientists set out to investigate the neural correlates of simple, everyday concepts such as ‘love’, ‘empathy’, ‘religious belief’, or ‘beauty’, without much consideration for the nature, taxonomy and functional role of such psychological states within a theory of the mind as a whole. As strictly modular, neurocognitive models struggle to account for dynamic, large-scale psychological phenomena, it seems highly unfortunate that the ‘psychological’ level of analysis is deemphasised in some atheoretical and reductionistic approaches within the neurosciences (see also Cooper & Shallice, 2010). For example certain fMRI studies disregard subjective states and meanings during scanning and make inferences about cognition exclusively on the basis of neural activation (e.g. certain studies give participants noxious stimuli and make inference about the neurobiology of pain but do not measure subjective pain ratings, nor the

cognitive and social context in which noxious stimulation occurs). These studies portray a radical materialism that leaves little causal room for the mental in brain-body relations. Such 'mindless' reductionism stands a chance of prevailing, unless and until 'mindful' theories and systematic studies of subjective experience provide novel insights about the mind-brain interface (Panksepp, 2007; Fotopoulou, 2012b).

Reclaiming the position of a new, dynamic and therapeutic neuropsychology within the integrative neurosciences

The above developments in neuroscience clearly call for a new, dynamic and integrative neuro*psychology* that can place the 'psychological' back into the neurosciences without needing to revert to extreme cognitivism/functionalism and strict modularity. There is no epistemological reason why lesion studies should be the only, or even the main method of a this new neuropsychology. On the contrary, as I have discussed above, increasing collaboration between neuroscientific methods can afford us with epistemic possibilities that simply did not exist even 15 years ago. However, lesion studies may still have an important role to play in shaping such possibilities, particularly when combined with other methods of enquiry. Such studies can abandon their exclusive attention to functional segregation and instead benefit from the older tradition of anti-localisationist theories in neuropsychology to incorporate notions of *structural* and *functional integration*, as well as *functional degeneracy* and *reorganisation* to the understanding of the damaged brain. I briefly outline recent advances in relation to these four notions below.

The advent of modern diffusion neuroimaging and probabilistic tractography, which can visualize white matter fiber tracts *in vivo* (Conturo et al., 1999), is a critical development for neuropsychology. Such techniques have increasingly been used to map connections even in regions of high anatomical complexity (Parker and Alexander,) and in relation to higher order mental abilities such as language and attention (see Cloutman & Lambon Ralph, 2012

for review). Their application of these methods to human lesion studies offers a unique opportunity to link behavioural or cognitive deficits with damaged structural connections and hence provide a more dynamic view of the brain abnormalities linked with specific neuropsychological syndromes (Catani & ffytche, 2005). This view in turn can allow greater understanding of the complex, interconnected networks that serve our cognitive abilities. Although these methods do not currently allow unequivocal conclusions on direct axonal connections (Mesulam, 2012), their continuous development holds the potential of increasing our understanding of *structural* connectivity and its role in mental functions and dysfunction.

Another set of studies has focused on how to study *functional* connectivity in patients with brain abnormalities (see Seghier et al., 2010 for review). In this context, deficits in functional integration or connectivity are assumed when the influence of one brain region on another is stronger or weaker in patients relative to control subjects (Price et al., 2006; Ween, 2008). This notion of 'dynamic diachisis' is important as it can allow future models of normal cognition to characterise not only which brain areas are necessary for certain mental functions but also how these areas are modulated by the activity of other areas during behaviour. These dynamic and hierarchical relations between brain areas cannot easily be induced with studies that focus only on behavioural dissociations and their corresponding structural damage in one or more brain areas.

Another small but increasing number of human lesion studies uses functional neuroimaging techniques to understand the role of functional degeneracy in language deficits (Noppeney et al., 2004). Degeneracy (see Edelman & Gally, 2001) refers to the ability of structurally-different elements to perform a similar function or achieve the same outcome (a similar, yet not identical concept is 'redundancy'). This principle can be traced back to holistic and anti-localisationist models that claimed that mental functions are performed by

the brain as a whole (Lashley, 1929), or at least by several, distributed and hierarchically organised systems in the brain (e.g. Luria, 1966). According to some of these theories, as there is a many to one relation between brain regions and mental functions, in case of damage to a particular part of the brain, other, parallel systems would take over the particular mental function (Lashley, 1929). The contemporary concept of degeneracy allows for some modularity and functional segregation but also accommodates a degree of functional redundancy and integration because it assumes that there are several, but limited in number, specialized systems for the same mental function (Price & Friston, 2002).

A final critical domain of the new, dynamic neuropsychology is the study of cognitive deficits in relation to brain plasticity and reorganisation following brain damage. Neuropsychological studies traditionally describe 'fixed' deficits resulting from irreversible damage to specialized brain modules. Indeed, only about fifty years ago, regrowth of connections after acute damage in the mature human brain was considered impossible. In the intervening years however, animal studies have overturned this dogma and replaced it with a model of the brain as a dynamic environment where "plasticity" of neural connections is the norm. It is increasingly recognised that the brain responds to brain injury by structural and functional reorganisation at a massive level. The latter changes include for example reorganization of functional circuits, leading to local expansion of cerebral activation areas and recruitment of parallel projecting cortical areas in the ipsilesional and contralesional hemispheres. Indeed, in the last 5 years, there has been particular progress in using functional neuroimaging techniques to measure such changes in the domains of motor function and language (Muellbacher & Hallett, 2006; Ward & Frackowiak, 2006). We still know very little about what drives and modulates these changes, but research in animals and preliminary research in humans suggest that they can be enhanced by environmental, behavioural, and pharmacological interventions. For example, recent studies have demonstrated that

neurological deficits previously regarded as intractable, e.g. stroke-induced hemiparesis, may be partly treated even by relatively simple and 'non-invasive' interventions, e.g. 'motor imagery' training (Seitz et al., 2004). Unfortunately, few studies exist in relation to high level cognitive and emotional processes following focal brain damage but it is clear that further research in this domain is now possible and warranted.

Taken together the above domains of study portray the potential for a dynamic and therapeutic neuropsychology. However, the labs that have the expertise to combine human lesion studies and other advanced neuroscientific techniques are certainly the exception rather than the rule in the field (e.g. see Mesulam, 2012; Price & Friston, 2002; Vuillemieur et al., 2001). Below I offer a brief historical account of a well-established neuropsychological syndrome, namely anosognosia for hemiplegia, as an example of how much the field has progressed thus far as well as what epistemological obstacles lie in the way of further progress and of integration with other neuroscientific developments. I will not attempt to offer a full account of the progress in the scientific understanding of this syndrome. Rather I will focus on developments that highlight some of the epistemological challenges of human lesion studies that I described above. Finally, I will use the recent computational modelling ideas of predictive coding and free energy minimisation to speculatively sketch how the understanding of motor awareness at psychological and neural levels can be advanced by taking into account some of the principles of such models and abandoning strict modularity and cognitivism.

The Neuropsychology of Anosognosia For Hemiplegia

Focal neurological damage may lead to abnormalities in the perception of and interaction with the external world, but it may also cause abnormalities in the perception of the patient's own body. The latter abnormalities can include primary somatosensory deficits such as tactile loss, or higher order deficits such as personal neglect. Following right

perisylvian lesions, and less often left perisylvian lesions (Cocchini et al., 2009) some patients may develop a striking disorder of body awareness termed ‘anosognosia for hemiplegia’ (AHP; lack of recognition or awareness of one’s paralysis). In the first decades following the naming of this symptom by Babinski (1914) several studies offered rich clinical descriptions of AHP and related symptoms (e.g. Joltrain, 1924; Waldenström, 1939; Gerstmann, 1942; Gilliat and Pratt, 1952 Critchley, 1955; Weinsten & Kahn, 1955). Such clinical descriptions portrayed a complex syndrome, including a varied pattern of deficits and manifestations. For example, some patients claim their limbs have moved even upon demonstration of the opposite (illusory movements, Feinberg et al., 2000; Fotopoulou et al., 2008), while others admit their on-line failure, but fail to update their long-term or, ‘off-line’ body awareness (Carruthers, 2008; see also Tsakiris & Fotopoulou, 2008). Other patients show implicit awareness of deficits despite explicit unawareness in verbal (Fotopoulou et al., 2010), or behavioural tasks (Nadrone et al., 2007; Cocchini et al., 2010; Moro et al., 2011), or portray a higher awareness of plegia in third-person versus first person tasks (Marcel et al., 2004; Fotopoulou et al., 2009; 2011).

It is currently unclear and debated to what extent these phenomena are manifestations of independent abnormalities, or the same primary deficit or a combination of deficits. Adding to the complexity is the fact that AHP appears in the context of a number of concomitant sensorimotor and cognitive impairments. During the 1980s and 1990s studies in cognitive neuropsychology attempted to establish whether any of these deficits or any given combination of deficits could explain the occurrence of one or more of the above anosognosic phenomena. While however several primary sensorimotor deficits and many higher order deficits such as intellectual impairment, memory loss, confusion, reasoning deficits, dysexecutive symptoms, visuospatial or, personal neglect, have all been reported frequently in patients with AHP, double dissociations between AHP and most of these deficits have

been noted in both acute and chronic AHP (e.g. Bisiach et al., 1986; Marcel et al., 2004). In response, some authors proposed multi-factorial theories of AHP, arguing for example that deficits in inferential reasoning may prevent sensorimotor deficits from being 'discovered' (Levine, 1990; Levine, Calvanio, & Rinn, 1991), or their discovery may not be 'remembered' (Cocchini et al., 2002). These explanations of AHP have now been tested in several studies (e.g. see Marcel et al., 2004; Vocat et al., 2010 for exceptionally well-conceived studies) and although they have not been equivocally supported, they remain relevant today (e.g. compare Prigatano & Schacter, 1991 with Prigatano, 2010).

The Cognitive Neuroscience of Anosognosia for Hemiplegia

This understanding of AHP as the secondary consequence of one or more concomitant neuropsychological deficits was however challenged by the progressive establishment of cognitive neuroscience during the 1990s. As topics such as consciousness, awareness, and the self entered the mainstream of cognitive neuroscience, scientists faced the challenge of a scientific understanding of self-consciousness. Advocates of what is generally known as the embodied cognition approach in philosophy of mind and cognitive neuroscience (e.g. Varela et al., 1991; Bermudez et al., 1995; Clark, 1996; Damasio, 1994; 2000; Gallagher, 2005), opted for distinguishing between several kinds and levels of self-consciousness and postulating a bodily "core" or "minimal" self, as the common denominator of all other facets of self-consciousness. It did not take long for researchers in cognitive neurology and neuroscience to realise that the rare and phenomenally rich body awareness aberration seen in anosognosia and related symptoms represents a unique source of insight into the neurocognitive processes of the bodily self.

As a consequence, the study of AHP and related disorders got a 'cognitive neuroscience' make over. This can be detected in at least four different developments: (i) new theoretical hypotheses were put forward, stemming from philosophical or computational

approaches on motor and embodied cognition, that view AHP as a specific disorder of motor awareness rather than a secondary consequence of deficits in other domains (e.g. Heilman et al., 1998; Frith *et al*., 2000; Berti et al., 2005); (ii) improvements in structural neuroimaging methods, software and statistics allowed lesion mapping studies to identify brain lesions selectively associated with AHP (Berti *et al*. 2005; Karnath et al., 2005); (iii) new diagnostic tests and meta-analysis of diagnostic criteria allowed group studies and statistical data about the prevalence of AHP (for reviews see Orfei *et al*., 2009; Jenkinson et al., 2011); and finally (iv) well-controlled, psychophysical experiments began to supplement clinical descriptions and neuropsychological assessments of patients (see Jenkinson and Fotopoulou, 2010 for review).

These developments have undoubtedly advanced our understanding of AHP. Yet, as aforementioned, it would be a mistake to assume that the epistemological premises of cognitive neuroscience are free from all the epistemological errors of cognitive neuropsychology. In the case of the study of AHP, contemporary studies seem to have inherited several epistemological premises from cognitive neuropsychology, most notably its strong emphasis on functional segregation and modularity. Simultaneously, and perhaps most unfortunately, some new studies of AHP portray some older limitations of traditional neuropsychology that cognitive neuropsychology had attempted to avoid, namely naive localisationism and reductionism.

For example, while progress in lesion mapping methods has allowed for a more precise identification of the lesion sites selectively associated with AHP, the results of such studies and their interpretations portray a return to strict modularity and a novel reductionism in the field. The labs of Berti (Berti et al., 2005) and Karnath (Karnath et al., 2005; Baier & Karnath, 2008) pioneered studies in which the anatomical extent of brain damage in groups with AHP was compared with that of matched control groups with hemiplegia and neglect.

These studies offer minimal details of their patients' unawareness symptoms, or of the subjective experience of their deficits. Perhaps most importantly, on the basis of these anatomical studies, these groups conclude that specific brain regions such as the premotor regions and the insula are involved in "conscious monitoring of motor acts" (Berti et al., 2005, p. 490) and (the posterior insula cortex) "is integral to self-awareness and to one's beliefs about the functioning of body parts" (Karnath et al., 2005, p. 7134). Yet these studies do not actually correlate any experimental measurement of motor monitoring, or self-awareness with their lesion data. Instead, psychometric measures are used to 'diagnose' anosognosia and classify patients to groups with or without the clinical symptom. Moreover, while the Karnath group note the high extent of white matter damage in their anosognosic groups, they do not place as strong emphasis on potential connectivity and functional integration interpretations, as they do on the functional segregation interpretations of their findings. Thus perhaps not surprisingly, the ensuing theories of awareness that both groups put forward are modular in their core conception; Berti and colleagues consider motor awareness and monitoring as a largely encapsulated function "implemented in the same neural network responsible for the process that has to be controlled" (Berti et al., 2005, p. 490), while Karnath and colleagues view awareness as a function that can be grossly and reliably disturbed due to damage to the posterior insula (Karnath et al., 2005).

Subsequent lesion studies in AHP continue to lack clinical description depth (e.g. Fotopoulou et al., 2010 offer little description of the potential clinical variability of anosognosic behaviours in the patients they group together in their study) but introduce some methodological rigour against extreme reductionism and strict modularity. For example, studies by Fotopoulou et al. (2010) and Moro et al. (2011) correlated the extent and location of brain lesions in anosognosic and control groups with behavioural data from well-controlled experiments. Similarly, Vocat and colleagues (2010) take into account in their voxel-based

lesion-symptom mapping both extensive neuropsychological and psychological assessments, as well as continuous scores of unawareness on the basis of a detailed awareness questionnaire. Not surprisingly, these studies point to a heterogeneous and multi-component disorder occurring due to lesions affecting a distributed set of brain regions, including subcortical structures. Importantly, the latter study demonstrated that the neuropsychological and neural profile of patients' changes in time, and different lesion patterns are associated with AHP at different time points.

A Unified Account of the Heterogeneity and Variety of Anosognosia for Hemiplegia

How is the dynamic, heterogeneous and multifaceted nature of AHP to be accounted for? In response to this question, several groups (e.g. Vuilleumier, 2004; Davies et al. 2005; Orfei et al., 2009; Cocchini et al., 2010; Garbarini *et al*., 2012) suggest a revival of cognitive theories that implicate two or more contributory factors, usually some higher order, top-down impairment superimposed on some sensory deficit (*cf*. the discovery theory of Levine *et al*., 1991). However useful these accounts, the envisioned relation between the various critical factors seems to be that of a pure addition. To the best of my knowledge there are no theories on AHP inspired by biological or psychological insights on neurocognitive architecture and dynamic, hierarchical relations between networks and there are no studies on structural or functional connectivity in this syndrome.

Here I propose the clinical variability of AHP can be best understood on the basis on a single, psychologically and neurobiologically-plausible formulation that takes into account both bottom-up and top-down mechanisms of perception and belief formation (see also Fotopoulou, 2012a). Specifically, anosognosic phenomena can be linked to an antagonism between 'prior beliefs' (predictive internal models of the world formed on the basis of previous learning and genetics; Friston, 2005) and 'prediction errors' (discrepancies between

expected and actual inputs based on ascending interoceptive and exteroceptive signals, e.g. Schultz and Dickinson, 2000) at different levels and domains of the neurocognitive hierarchy (Mesulam, 2012). A dynamic balance needs to be maintained between the two so that we can filter and organise new incoming information based on our robust expectations, but the latter cannot be so robust that we do not allow for new learning and flexible adjustment to a changing world. Anosognosic behaviours, experiences and delusions can be hypothesized to involve abnormalities in the dynamic balance between these two poles (Fotopoulou, 2010; Fotopoulou, 2012a).

In the aforementioned model of Berti and colleagues (2005) a similar antagonism is described. However, as this model was inspired by a computational model of motor control (Wolpert, 1997), this antagonism is limited to the domain of action and concerns only efferent (predictive) and afferent (feedback) sensorimotor signals. By contrast, more recent theories of brain function have put forward the (arguably reductionistic) notion that the brain as a whole works as an Helmholtzian inference machine (Helmholtz, 1878/1971) that is trying to optimize its own Bayesian probabilistic model of the world by actively predicting the causes of its sensory inputs (Rao & Ballard, 1999; Friston, 2005). The essence of such Bayesian, 'predictive coding' frameworks is that neurobiological message-passing in the brain is achieved by structurally or functionally embodying (neurobiologically representing) a prediction (or a prior) and responding to errors (mismatches) in the accuracy of the prediction, or prediction errors. The idea that perception is an unconscious inferential process is not new to psychology (Gregory, 1966), neither is the idea that what is already 'known' or 'learned' in the mind shapes the perception and learning of new experiences (e.g. Bartlet, 1932). What is new about these frameworks is that they unify these ideas in one mathematically formulated framework that makes specific neurobiological predictions about the function of the brain (Friston, 2010). In this framework, perception and action both serve

this optimization (reduction of prediction error, surprise or free energy, see Friston, 2010) by changing predictions, or the signals being predicted, respectively. Furthermore, some of these models (e.g. Friston, 2010) place particular emphasis on the hierarchically organised large-scale networks that perform competing functions in the brain, conveying prediction errors via feedforward connections from lower to higher levels to optimize representations in the latter and transferring higher-order predictions via feedback connections that can suppress prediction errors in lower levels. The reciprocal but asymmetric characteristics of this hierarchy (Mesulam, 2012) allow for an optimization that makes every level in the hierarchy accountable to the others, delivering an internally consistent re-representation of sensory causes at multiple levels of the neurocognitive hierarchy. Thus, these models can envision a mismatch between expectation and experience in various levels of the neurocognitive hierarchy and in relation to several cognitive and emotional domains.

Hence, these models can explain more facets of anosognosia than previous models on the basis of a single dynamic balance between prior expectation of bodily signals and current experiences of the body, implemented in different domains and levels of brain-mind organisation. For example, they can explain the motor illusions of patients who claim they have moved their arms as planned even upon demonstration of the contrary (Fotopoulou et al., 2008), but they can also explain the more general, obstinate adherence of other patients to their premorbid everyday habits (‘Of course, I can walk’) despite implicit knowledge of their paralysis (Fotopoulou et al., 2010). Specifically, I speculatively propose that AHP can be caused by at least five kinds of disruptions in the dynamic relation between expectation and experience. These functional disruptions are not mutually exclusive and thus they can be combined in different ways in different patients, suggesting a potential, novel computational focus on detailed, case-based, neuropsychological enquiries.

First, a source of disruption is the mere fact that patient can no longer update their representation of their affected body parts by actively sampling sensory states (i.e. moving their affected limbs). Of course, this *lack of active inference* does not seem sufficient to cause AHP as the syndrome occurs in a minority of patients with hemiplegia and it is more common in patients with left rather than right-sided hemiplegia. However, this disruption may nevertheless dynamically contribute to the phenomenology of AHP and hence it needs to be taken into account, together with the other possible disruptions, in a computational model of the syndrome.

Second, *aberrant perceptual inference* (suboptimal synaptic activity, Friston, 2010) can be caused by deficits that lead to weak, or absent signals about prediction errors. Such deficits may occur in relation to exteroceptive signals about the left side of the body as represented in the connections of right hemisphere subcortical areas (e.g. the thalamus), or re-represented and organised in cortical functional networks of the right-hemisphere (Berti et al., 2005; Fotopoulou et al., 2010; Vocat et al., 2010; Moro et al., 2011). In addition, interoceptive, emotional signals normally also represented in subcortical areas and the insula, and their connections may also be compromised (Karnath et al., 2005; Fotopoulou et al., 2010; Vocat et al., 2010; Moro et al., 2011), leading to an obstinate adherence to past expectations of how the affected body parts should 'feel' and related aberrant beliefs. Thus, absent, weak, emotionally blunted or neglected prediction errors will be incapable of updating motor awareness, particularly in the presence of intact motor predictions and other prior beliefs (Frith et al., 2000; Fotopoulou et al., 2008).

Such 'sensory' bottom-up deficits have long been associated with AHP, but because of the observed double dissociations between such deficits and AHP (e.g. Bisiach et al., 1986; Marcel et al., 2004), the logic of modular neuropsychological inference required that such deficits are not considered necessary for anosognosia to occur; hence, they have been

theoretically de-emphasised in favour of top-down explanations (Bisiach & Berti, 1987; Heilman et al., 1998; Berti et al., 2005), or 'dual-factor' theories (see above). However, based on the proposed dynamic conceptualisation of anosognosia, severe or combined deficits in one or more of these domains may lead some patients to produce anosognosic behaviours about their affected limbs, without the requirement that these deficits are necessary components for the occurrence of all types of AHP. The relative weighting of such deficits in relation to the other types of predicting coding disruptions described here (and the issue of whether they are sufficient for any type of AHP to occur) remains to be computationally modelled and empirically tested.

Third, *perceptual learning* (i.e. synaptic efficacy and plasticity, Friston, 2010) may be affected by certain lesions, such as the recently discovered limbic lesions in AHP patients (Fotopoulou et al., 2010; Vocat et al., 2010). These may lead to deficits in updating and learning processes per se, leading to an obstinate adherence to past expectations of the state of the affected body parts and related aberrant beliefs.

Fourth, dopamine-depleting lesions in fronto-striatal circuits (Venneri & Shanks, 2004; Fotopoulou et al., 2010; Vocat et al., 2010; Moro et al., 2011) may have a modulatory role in AHP, leading to a more general *difficulty in optimizing the precision (uncertainty) of prediction errors* (Friston *et al*., 2012), affecting their salience and ultimately both short- and long-term learning (suboptimal synaptic gain and plasticity, Friston, 2010). This can be linked with both specific instances of aberrant motor monitoring in functionally specialised systems (Berti et al., 2005), or more generally in global error monitoring (Venneri & Shanks, 2004; Vuillemier, 2004; Davies et al., 2005), mental flexibility (Levine et al., 1991) and 'surprise detection' (Ramachandran, 1995) deficits.

Finally, *premorbid priors* at various neurocognitive levels may be particularly strong and resistant to change. For example, different individuals have different premorbid traits of

adherence to past self schemata and experiences, including their experience of and attitude towards their own body (e.g. Gainotti, 1975). Although some of these attitudes and emotional factors have been regarded as purely psychogenic traits in the past (Weinstein & Kahn, 1955), and criticised as such (Bisiach & Geminiani, 1991), it is possible that the weakening of predictions errors described above may have particularly strong effects in a brain system that premorbidly requires large and sustained prediction errors in order to update its priors (see also Fotopoulou, 2010).

Clearly this speculative sketching of a model of AHP requires proper computational modelling and empirical testing in several neural and behavioural levels. The relative contribution of some of the above deficits and hypothesised networks, as well as some of multiple dynamic relations and connectivity patterns between them, may prove less important than others. However, the above speculative model seems capable of addressing the wide, clinical variability of AHP. Importantly, the model could potentially account for the spontaneous (Vocat *et al*., 2010), or intervention-based (Fotopoulou *et al*., 2009; Fotopoulou et al., 2011) changes of unawareness in time, as progressive updating of priors based on accumulating or, alternative signals (e.g. third-person perspective mirror feedback or of video-based, off-line feedback) about prediction errors, respectively. It may also lead to novel predictions about the potential functional restoration of AHP through behavioural (e.g. encouraging the processing of the sensorimotor or emotional evidence for an anosognosic belief rather than challenging the belief itself), or pharmacological (Corlett et al., 2009) intervention.

More generally, it is hoped that tolerance for the speculative, exploratory and computational nature of such encompassing and dynamic neuropsychological hypotheses, and rejection of the potentially misleading robustness of modular explanations, may lead to a new, more dynamic neuropsychological understanding of the mechanisms of motor and body

awareness and other psychological phenomena. Although the precise measurement of the hypothesised brain functions and the related large-scale network operations is still in its infancy in all fields of neuroscience, it seems that for neuropsychology to continue to contribute to contemporary neurosciences it needs to carefully step away from its comfort zone and embrace new epistemological assumptions about mind-brain relations. Simultaneously for as long as mentalistic concepts are less present in neuroscientific theories than they are in neuropsychological theories, neuropsychology will have a unique and important role to play in the neurosciences.

References

Babinski, J. (1914). Contribution à l'étude des troubles mentaux dans l'hémiplégie organique cérébrale. *Revue Neurologique, 27:*845-847.

Baier, B., Karnath, H.O. (2008). Tight link between our sense of limb ownership and self-awareness of actions. *Stroke*; 39(2):486-8.

Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge: Cambridge University Press.

Bechtel, W. (2012). The Epistemology of Evidence in Cognitive Neuroscience. In R. Skipper Jr., C. Allen, R. A. Ankeny, C. F. Craver, L. Darden, G. Mikkelson, and R. Richardson (eds.), *Philosophy and the Life Sciences: A Reader*. Cambridge, MA: MIT Press. In press.

Benedetti F. (2010). *The Patient's Brain: The Neuroscience Behind the Doctor–Patient Relationship*. Oxford, UK: Oxford University Press.

Bermúdez, J., Marcel, A.J., Eilan. N. (1995). The Body and the Self, MIT Press, Cambridge, MA.

Berti, A., Bottini, G., Gandola, M., Pia, L., Smania, N., Stracciari, A., Castiglioni, I., Vallar, G., Paulesu, E. (2005). Shared cortical anatomy for motor awareness and motor control. *Science*; 309: 488-91.

Bisiach, E. & Berti, A. (1987). Dyschiria: An attempt at its systematic explanation. In M Jeannerod (Ed), *Neurophysiological and Neuropsychological Aspects of Spatial Neglect*. Amsterdam: Elsevier, pp. 183-201.

Bisiach, E., Vallar, G., Perani, D., Papagno, C., Berti, A. (1986). Unawareness of disease following lesions of the right hemisphere: Anosognosia for hemiplegia and anosognosia for hemianopia. *Neuropsychologia*; 24:471-482.

Bisiach, E. & Geminiani, G. (1991). Anosognosia related to hemiplegia and hemianopia. In GP Prigatano and DL Schacter (Eds), *Awareness of Deficit After Brain Injury*. New York, Oxford University Press, 1991, pp. 17-39.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature, 397*, 756.

Caramazza, A. (1984). The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and Language*, *21*, 9–20.

Caramazza, A., & Coltheart, M. (2006). Cognitive neuropsychology twenty years on. *Cognitive Neuropsychology*, 23, 3–12.

Carruthers, G. (2008). Types of body representation and the sense of embodiment. *Consciousness and Cognition*, 17, 1302-1316.

Catani, M. & Ffytche, D. H. (2005). The rises and falls of disconnection syndromes. *Brain, 128*, 2224–2239.

Clark, A. (1996). Being There, MIT Press, Cambridge, MA.

Cocchini G, Beschin N, Della Sala S. (2002). Chronic anosognosia: a case report and theoretical account. *Neuropsychologia*; 40: 2030-8.

Cocchini, G., Beschin, N., Cameron, A., Fotopoulou, A., & Della Sala, S. (2009). Anosognosia for motor impairment following left brain damage. *Neuropsychology, 23(*2), 223-230.

Cocchini G, Beschin N, Fotopoulou A, & Della Sala S. (2010). Explicit and implicit anosognosia for upper limb motor impairment. *Neuropsychologia*, 48(5):1489-94.

Coltheart, M. (1985). Cognitive neuropsychology. In M. Posner & O. S.M. Marin (Eds.), *Attention and performance, Vol. XI* (pp. 3–37). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Coltheart, M. (2006). What has functional neuroimaging told us about the mind (so far)? Cortex, 42, 323–331.

Coombes, S. (2010). Large-scale neural dynamics: Simple and complex. *NeuroImage*, *52*, 731–739.

Cooper, R.P., Shallice, T. (2010). Cognitive neuroscience: The troubled marriage of cognitive science and neuroscience. *Topics in Cognitive Science 2*; 398–406.

Corlett, P.R., Taylor, J.R., Wang, X.-J., Fletcher, P.C., Krystal, J.H. (2009). Toward a neurobiology of delusions. *Progress in Neurobiology*; 92: 345–369.

Conturo, T.E., Lori, N.F., Cull, T.S., Akbudak, E., Snyder, A. Z., Shimony, J.S., McKinstry, R. C., Burton,H., and Raichle, M. E.(1999). Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences U.S.A. 96 10422–10427*.

Cloutman, L.L., Lambon Ralph, M.A. (2012). Connectivity-based structural and functional parcellation of the human cortex using diffusion imaging and tractography. *Frontiers in Neuroanatomy* ;6:34

Critchley, M. (1955). Personification of paralysed limbs in hemiplegics. *British Medical Journal 2*:284–286

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain.* New York: G.P. Putnam's Sons.

Damasio, A, (2000). The Feeling of What Happens: Body and Emotion in the Making of Consciousness, Harvest Books, New York.

Davies, M., Davies, A.A., Coltheart M. (2005). Anosognosia and the two-factor theory of delusions. *Mind and Language*; 20: 209-236.

Deacon, T.W. (1989). Holism and Associationism in Neuropsychology: An Anatomical Synthesis. in E. Perecman (Ed.), *Integrating Theory and Practice in Clinical Neuropsychology.* Erlbaum. Hilsdale, NJ. 1-47.

Decety, J. & Ickes, W.J. (2009). *The social neuroscience of empathy*. MA: MIT Press.

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

Edelman, G.M., Gally, J.A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences U.S.A.* 98 , 13763– 13768.

Ellis, A. W., & Young, A. W. (1988). *Human Cognitive Neuropsychology*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Feinberg, T. E., Roane, D. M., & Ali, J. (2000). Illusory limb movements for anosognosia for hemiplegia. *Journal of Neurology, Neurosurgery and Psychiatry, 68*, 511-513.

Fodor, J. (1975). *The language of thought.* USA: Thomas Y. Crowell Company, Inc.

Fodor, J.A., Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture. Cognition **28**, 3–71.

Fotopoulou, A. (2010). The affective neuropsychology of confabulation and delusion. *Cognitive Neuropsychiatry*, 15, 1/2/3, 1-13.

Fotopoulou, A. (2012a). Illusions and Delusions in Anosognosia for Hemiplegia: From Motor Predictions to Prior Beliefs. Commentary. *Brain* 135; 1344–1347.

Fotopoulou, A. (2012b). Towards Psychodynamic Neuroscience. In A. Fotopoulou, M. Conway, & D. Pfaff. (Eds.) *From the Couch to the Lab: Trends in Psychodynamic Neuroscience*. *Oxford University Press*.

Fotopoulou, A., Conway, M.A. & Pfaff, D. (2012) *From the Couch to the Lab: Trends in Psychodynamic Neuroscience*. *Oxford University Press*.

Fotopoulou, A., Tsakiris, M., & Haggard, P., Rudd, A., & Kopelman M. (2008). The role of motor intention in motor awareness: An experimental study on anosognosia for hemiplegia, *Brain* 131: 3432-42.

Fotopoulou, A., & Jenkinson, P. M. (2010). Unawareness of deficit in acute stroke: Neuropsychological therapy matters. *Stroke Matters, 9*, 8-9.

Fotopoulou, A., Jenkinson, P. M., Tsakiris, M., Haggard, P., Rudd, A., & Kopelman, M. D. (2011). Mirror-view reverses somatoparaphrenia: Dissociation between first- and third-person perspectives on the body. *Neuropsychologia, 49(14)*, 3946-3955.

Fotopoulou, A., Pernigo, S., Maeda, R., Rudd, A., & Kopelman, M. A. (2010). Implicit awareness in anosognosia for hemiplegia: unconscious interference without conscious re-representation. *Brain, 133(Pt 12)*, 3564-3577.

Fotopoulou, A., Rudd, A., Holmes, P., & Kopelman, M. (2009). Self-observation reinstates motor awareness in anosognosia for hemiplegia. *Neuropsychologia, 47(5)*, 1256-1260.

Freeman, W.J (2003). Neurodynamic models of brain in psychiatry. *Neuropsychopharmacology* 28, S54–S63.

Friston, K. (1994). Functional and effective connectivity in neuroimaging: a synthesis. *Human Brain Mapping, 2*, 56-78.

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

Friston, K. (2009a). The free-energy principle: a rough guide to the brain? *Trends*

in Cognitive Sciences. 13, 293–301.

Friston, K. J. (2009b). Modalities Modes, and Models in Functional Neuroimaging. *Science, 326*, 399-403.

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

Friston, K. J. & Dolan, R. J. (2010). Computational and dynamic models in neuroimaging.

NeuroImage, 52, 752–765.

Friston, K. J., Daunizeau, J., Kilner, J. & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biological Cybernetics, 102*, 227-260.

Friston, K.J., Shiner, T., FitzGerald, T., Galea, J.M., Adams, R., Brown, H., Dolan, R.J., Moran, R., Stephan, K.E., Bestmannm S. (2012). Dopamine, affordance and active inference. PLoS *Computational Biology* ;8(1):e1002327.

Frith, C.D., Blakemore, S.J., Wolpert, D.M. (2000) Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London, Series B. Biological Sciences* 355:1771–1788

Frith, U. & Frith, C. (2010). The social brain: allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society of London, Series B. Biological Sciences* 365, 165-176.

Fuster, J.M. (2009). Cortex and Memory: Emergence of a New Paradigm. *Journal of Cognitive Neuroscience*, 21 (11): 2047-2072.

Gainotti, G. (1975). Confabulation of denial in senile dementia: an experimental study. *Psychiatria Clinica, 8*, 99-108.

Garbarini F, Rabuffetti M, Piedimonte A, Pia L, Ferrarin M, Frassinetti F, Gindri P, Cantagallo A, Driver J, Berti A.(2012). 'Moving' a paralysed hand: bimanual coupling effect in patients with anosognosia for hemiplegia. *Brain*;135(Pt 5):1486-97.

Gerstmann J. (1942). Problem of imperception of disease and of impaired body territories with organic lesion. *Archives of Neurologyand Psychiatry, 48:* 890-913.

Gallagher, S. (2005). How the Body Shapes the Mind. New York: Oxford University Press

Gilliat, R.W. and Pratt R.T.C..(1952). Disorders of perception and performance in a case of right-sided cerebral thrombosis. *Journal of Neurology, Neurosurgery and Psychiatry, 15:* 264- 271, 1952.

Gregory, R. (1966). *Eye and Brain: The Psychology of Seeing*. London: Weidenfeld and Nicolson.

Grillner, S., Kozlov, A. &Kotaleski, J.H. (2005). Integrative neuroscience: linking levels of analyses. *Current Opinion in Neurobiology, 15*:614–621

Harley, T. A. (2004). Does cognitive neuropsychology have a future? *Cognitive Neuropsychology*, 21, 3–16.

Heilman KM, Barret AM, Adair JC. Possible mechanisms of anosognosia: a defect in self-awareness. *Philosophical Transactions of the Royal Society of London, Series B. Biological Sciences* 1998; 353:1903–9.

Helmholtz, H.von. (1878/1971). The facts of perception. In: Kahl, R. (Ed.), Selected

Writings of Herman von Helmholtz. Weslyan University Press.

Joltrain E. (1924). Un nouveau cas d'anosognosie. *Revue Neurologique, 2:* 638-640, 1924.

Jenkinson P, Fotopoulou A. (2010). Motor awareness in anosognosia for hemiplegia: Experiments at last. *Experimental Brain Research*, 204(3):295-304.

Jenkinson, P. M., Preston, C., & Ellis, S. J. (2011). Unawareness after stroke: A review and practical guide to understanding, assessing, and managing anosognosia for hemiplegia. *Journal of Clinical and Experimental Neuropsychology, 33(10),* 1079-1093.

Karnath HO, Baier B, Nägele T. (2005). Awareness of the functioning of one's own limbs mediated by the insular cortex? *Journal of Neuroscience* 3;25(31):7134-8.

Karnath, H-O., Rorden, C. & Ticini L.F. (2009). Damage to white matter fiber tracts in acute spatial neglect. *Cerebral Cortex* 19 (10):2331-2337.

Knoblich, G., Thornton, I. M., Grosjean, M. & Shiffrar, M. (2006). *Human Body Perception from the Inside Out*. New York: Oxford University Press.

Laing, C. R., Troy, W. C., Gutkin, B. & Ermentrout, G. B. (2002). Multiple bumps in a neuronal model of working memory. *SIAM Journal of Applied Mathematics*, *63*, 62–97.

Lambon Ralph, M.A. (2004). Reconnecting cognitive neuropsychology: commentary on Harley's 'Does cognitive neuropsychology have a future?' *Cognitive Neuropsychology*, 2004, 21 (1), 31–35.

Lashley, K.S. (1929) *Brain Mechanisms and Intelligence*. Chicago: University of Chicago Press.

LeDoux, J. (1996). *The emotional brain*. New York: Simon & Schuster.

Levine, D.N. (1990). Unawareness of visual and sensorimotor defects: a hypothesis. *Brain and Cognition, 13:* 233-281.

Levine, D.N., Calvanio, R., Rinn, W.E. (1991). The pathogenesis of anosognosia for hemiplegia. *Neurology*;41:1770-1781

Logothetis, N.K. (2008). What we can and cannot do with fMRI. *Nature, 453,* 869-878.

Luria, A.R. (1966).Higher cortical functions in man, London: Tavistock.

Marcel AJ, Tegner R, Nimmo-Smith I. (2004). Anosognosia for plegia: specificity,

extension, partiality, and disunity of bodily unawareness. *Cortex*; 40:19–40.

Mesulam¸ M. (2012). The evolving landscape of human cortical connectivity: Facts and inferences. *Neuroimage*. In press.

Moro V, Pernigo S, Zapparoli P, Cordioli Z, Aglioti SM. (2011). Phenomenology and neural correlates of implicit and emergent motor awareness in patients with anosognosia for hemiplegia. *Behavioural Brain Research* 225(1):259-69.

Müller, R. (1992). Modularism, holism, connectionism: Old conflicts and new perspectives in aphasiology and neuropsychology, *Aphasiology, 6*:5, 443-475.

Muellbacher W, Hallett M. (2006). Reprogramming surviving motor cortex after stroke. In: Lomber S, Eggermont J, editors. Reprogramming the cerebral cortex: Plasticity following central and peripheral lesions. Oxford University Press; New York: 2006. pp. 257–72.

Nardone IB, Ward R, Fotopoulou A, Turnbull, O.E. (2007) Attention and emotion in anosognosia: Evidence of implicit awareness. *Neurocase,* 13: 438 - 445.

Noppeney, U., Friston, K.J., Price C.J. (2004). Degenerate neuronal systems sustaining cognitive Functions. *Journal of Anatomy,* 205 , pp. 433–442

Orfei, M.D. Caltagirone, M., Spalletta, G.(2009). The Evaluation of Anosognosia in Stroke Patients. Cerebrovascular Disease, 27:280–289.

Page, M.P. (2006). What can't functional neuroimaging tell the cognitive psychologist? *Cortex*, 42(3):428-43.

Panksepp, J. (1998). *Affective Neuroscience*. New York: Oxford University Press.

Panksepp, J. (2007). Neuro-psychoanalysis may enliven the mindbrain sciences. *Cortex, 43,* 1106-1107.

Panksepp, J., Solms, M. (2012).What is neuropsychoanalysis? Clinically relevant studies of the minded brain. Trends in Cognitive Sciences 16(1):6-8.

Parker,G.J.M.,and Alexander,D.C. (2005). Probabilistic anatomical connectivity derived from the microscopic persistence angular structure of cerebral tissue. *Philosophical Transactions of the Royal Society of London, Series B. Biological Sciences* 360, 893–902.

Piccinini, G. & Scarantino, A. (2011). Information processing, computation, and cognition.

Journal of Biological Physics, *37,* 1–38.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10, 377–500.

Price, C. J., Crinion, J., and Friston, K. J. (2006). Design and analysis of fMRI studies with neurologically impaired patients. *Journal Magnetic Resonance Imaging* 23, 816–826.

Price, C.J., Friston, K.J. (2002) Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences* **6**, 416–421.

Raichle, M., MacLeod, A., Snyder, A., Powers, W., Gusnard, D. & Shulman, G. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences U.S.A., 68,* 676-682.

Ramachandran, V. S. (1995). Anosognosia in pariental lobe syndrome. *Consciousness and Cognition, 4,* 22-51.

Rizzolatti, G. and Craighero, L. (2004), 'The mirror neuron system', *Annual Review of Neuroscience*, **27**, pp. 169–92.

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

Prigatano, G.P. & Schacter, D.L. (1991). *Awareness of Deficit after Brain Injury (Eds)*. New York: Oxford University Press.

Schultz, W., Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience* 23, 473–500.

Seeley, W., Menon, V., Schatzberg, A., Keller, J., Glover, G, Kenna, H., Reiss, A. & Greicius. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *The Journal of Neuroscience, 27,* 2349-2356.

Seitz RJ, Bütefisch CM, Kleiser R, Hömberg V. (2004). Reorganisation of cerebral circuits in human ischemic brain disease. Restorative Neurology and Neuroscience, 2004;22(3-5):207-29.

Seghier, M.L., Zeidman, P., Neufeld N.H., Leff A.P, & Price C.J. (2010). Identifying abnormal connectivity in patients using dynamic causal modeling of fMRI responses. *Frontiers in Systems Neuroscience 4*; 142: 1-14.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Shridharan, D., Levitin, D., & Menon, V. (2008). A critical role for the right frontoinsular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences U.S.A.,105,* 12569-12574.

Steinberg, D.A. (2009) Cerebral localization in the nineteenth century — the birth of a science and its modern consequences. *Journal of the History of the Neurosciences: Basic and Clinical Perspectives*, *18*:3, 254-261

Stone, T., & Young, A. W. (1997). Delusions and brain injury: The philosophy and psychology of belief. *Mind and Language*, *12*, 327_364.

Tsakiris, M., & Fotopoulou, A. (2008). Is my body the sum of online and offline body representations? *Consciousness and Cognition*, 17(4):1317-20.

Valdes-Sosa, P. A., Roebroeck, A., Daunizeau, J. & Friston, K. (2011). Effective connectivity: Influence, causality and biophysical modelling. *Neuroimage*, 15;58(2):339-61.

Varela, F. Thompson, E. Rosch, E. (1991). The Embodied Mind, MIT Press, Cambridge, MA.

Venneri, A. & Shanks, M.F. (2004). Belief and awareness: reflections on a case of persistent anosognosia. Neuropsychologia 42 (2004) 230–238

Vocat R, Staub F, Stroppini T, Vuilleumier P. (2010). Anosognosia for hemiplegia: a clinical-anatomical prospective study. *Brain*,133(Pt 12):3578-97.

Vuilleumier, P. (2004). Anosognosia: The neurology of beliefs and uncertainties. *Cortex*, 40: 9-17.

Vuilleumier P, Sagiv N, Hazeltine E Poldrack, R.A., Swick, D., Rafal, R.D., & Gabrieli, J.D.E. (2001). Neural fate of seen and unseen faces in visuospatial neglect: a combined event-related functional MRI and event-related potential study. *Proceedings of the National Academy of Sciences U.S.A*; 98: 3495–500.

Waldenström J. (1939). On anosognosia. *Acta Psychiatrica, 14:* 215-220.

Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron, 56*, 366-383.

Ward NS, Frackowiak RS. (2006). The functional anatomy of cerebral reorganisation after focal brain injury. *Journal of Physiology* Paris, 99(4-6):425-36.

Ween, J. E. (2008). Functional imaging of stroke recovery: an ecological review from a neural network perspective with an emphasis on motor systems. *Journal of Neuroimaging* 18, 227–236.

Weinstein EA, Kahn RL (1955).*Denial of Illness*. Springfield: Charles C. Thomas.

Wolpert, D. M. (1997). Computational models of motor control. *Trends in Cognitive Sciences, 1*(6), 209-216.