Commentary on the 'Conscious Id' by Mark Solms

Title: *Beyond the Reward Principle: Consciousness as Precision Seeking*

A. Fotopoulou

King's College London, Institute of Psychiatry

Abstract: I use an influential computational theory of brain function, the free energy principle to suggest three points of added complexity to Solms' intriguing descriptions of the embodied mind: (a) the link between Ego and cognitive automaticity is not as straightforward as Solms suggests. Instead, cognition strives for both inference *and flexibility* in relation to the changing world and the inflexible drives; (b) affective consciousness may primarily map the degree of *uncertainty* (not pleasure) of internal bodily signals. Subcortical areas are the neurobiological source of this facet of consciousness that in itself is likely to be localised between many, distributed brain areas; and lastly (c) our innate motivational systems, the Id, ultimately serve the same optimisation principle as the Ego. However, unlike the later, they call for automaticity in behaviour, on the basis of innate unconscious priors that are fulfilled by instinctual e-motions and other reflexes, understood as evolutionary defined, primitive forms of active and perceptual inference.

Key words: Embodiment; Consciousness; Emotion; Motivation; Free Energy; Predictive Coding

Mark Solms' rich and provocative article weaves together classic concepts of Freudian metapsychology and insights from affective neuroscience into a novel, lucid neuropsychoanalytic account of embodiment and consciousness. Doing justice to the many research traditions and creative links this article invokes is not possible in this brief commentary. Moreover, I wish to take nothing away from the force and clarity by which Solms contrasts the subjectively felt body to the 'cognitivised' one, and questions the simplistic equation of consciousness with Ego. In this commentary I will use a computational theory of brain function, the free energy framework (Friston, 2005) to merely suggest three points of added complexity to Solms' intriguing descriptions of the embodied mind. These will relate to (a) the nature of the Ego (our cognition); (b) the distinction between phenomenal and perceptual consciousness and lastly (c) the Id (our innate drives).

A theoretical framework from computational neuroscience

The starting point of the free energy framework (Friston, 2005) is that the world is an uncertain place for self-organising biological agents to survive. This inherent ambiguity of the world threatens our need to occupy a limited repertoire of sensory states (e.g. humans need certain ranges in environmental temperature in order to survive). If however we cannot predict the causes of possible changes in the world with any certainty, we may find ourselves in surprising states for longer periods than those we could biologically sustain. We thus come up with a defiant solution. We base our predictions about our sensory states on unconscious inferences about their causes in the world (von Helmholtz, 1866). On the basis of limited or noisy information, our brain engages in probabilistic representations of the causes of our future states in an uncertain world so that it maintains hypotheses (''generative models'') of the hidden causes of sensory input. Theoretical neuroscientists use Bayesian theory to formalise this kind of inference and a number of other computational terms about probability distributions, such as 'free energy', 'uncertainty' and 'surprise' that have formal (mathematical) definitions. Here I attempt to find faithful 'psychological translations' for some of these concepts in order to examine the ideas put forward by Solms within this 'psychologised' version of the free energy framework.

According to the framework, the brain attempts to reduce the probability of being surprised by the world by reducing its own representational errors over time. These errors have been conceptualised as free energy, on the basis of the formal definition of the latter; a quantity from informational theory that bounds (is greater than) the evidence for a model of data (Feynmann 1972; Hinton & van Camp, 1993). When the data is sensory, free energy bounds the negative log-evidence (surprise) inherent in them, given a model of how the data were caused. Furthermore, our brain is assumed to achieve the minimisation of free energy by recurrent message passing among hierarchical level of

cortical systems, so that various neural subsystems at different hierarchical levels minimize uncertainty about incoming information by structurally or functionally embodying a prediction (or a prior) and responding to errors (mismatches) in the accuracy of the prediction, or prediction errors (Rao & Ballard, 1998). Such message passing is considered neurobiologically plausible on the basis of functional asymmetries in cortical hierarchies (e.g. Mesulam, 2012). Minimizing free energy corresponds to explaining away prediction errors following the principles of Bayes (Friston, 2010).

However, perceptual inference cannot take us far in terms of our ultimate goal; surviving in an uncertain world. Psychologically speaking, we may become better in predicting ('mentalising') the changes in the environment that act to produce sensory impressions on us, but we cannot on this basis change the sensations themselves and hence ultimately their surprise. It is only by acting upon the world that we can 're-sample' the world to ensure we satisfy our predictions about the sensory input we expect to receive. Thus, action is understood as being elicited to fulfil prior expectations about proprioceptive sensations, not desired sensory states (as optimal motor control theory suggests). Both action and perception are governed by the imperative of free energy minimisation; action reduces free energy by changing sensory input, while perception reduces free-energy by changing predictions.

At this point we can follow Solms (and Carhart-Harris & Friston, 2010; Fotopoulou, 2012a; Hopkins, 2012) in considering this framework in parallel with the central Freudian 'economic' concepts of 'free' and 'bound' energy. Although the Freudian and Fristonian notions of free energy differ between them, they were both inspired by Helmholtzian ideas about thermodynamics and commonly convey the imperative to 'bind' and ultimately 'minimise' a quantity (formal surprise in Friston; nervous excitation and then psychical energy in Freud) that otherwise renders the functioning of a biological system suboptimal in an unpredictable world. As described above, the free energy framework places emphasis on an antagonism between an unpredictable world and a predicting mind. Freud however placed emphasis on an antagonism between *an inflexible body* and an unpredictable world; he claimed the mind (the Ego) is formed on the basis of an antagonism between the organism's biological needs (and corresponding inherited drives, the Id) and an unaccommodating world. Of course it should be clear that the two frameworks bear on the same single, ultimate principle; a biological system with given constraints needs to reduce its modes of exposure to the unpredictable world in order to maintain and prolong its evolutionary constrained existence (e.g. Friston, 2011; Friston & Ao, 2012). I fear however that Solms' new division of labour between the Ego and the Id and his denigration of drives (innate motivational priors) to affective consciousness and of Ego (cognition) to mere representation and automatisation, risks de-

emphasising the central place of the Freudian antagonism between an inflexible body and an unpredictable world. In fact, Solms says that much when he regards his new Conscious, subcortical Id as the seat of the novel, the salient, the emotional in the brain and his unconscious, cortical Ego as the driver of automaticity.

The Flexible Ego and Consciousness as Precision Seeking

In the free energy framework the challenge of the organism is to navigate the world by sustaining a set of prior beliefs, sufficiently robust that she does not react reflexively to incoming sensory stimuli. At the same time, and contrary to what Solms claims about automaticity (p. 17), our generative models of the world must not be so immutable that our responses become fixed, stereotypical and insensitive to unpredictable change. Indeed, an intrinsic component of the free energy framework is that our generative models need to maintain an optimal, dynamic balance between their robustness and flexibility. In Bayesian terms, organisms need to probabilistically infer two properties of the world; its states (content; mathematically this can be thought of as the centre of a probability distribution) and the uncertainty (context; the dispersion of such distribution) about such states. It is perhaps Solms' apparent disregard of the latter that lead him to equate the Ego with the driver of automaticity and to claim that the reduction of salience constitutes one of the aims of the Ego (p. 17). Increases in salience, novelty and motivational value do not oppose the principle of minimisation of free energy. In fact, the opposite applies; optimal inference in both perception and action requires optimising the precision (mathematically inverse dispersion or variance, and hence the inverse of uncertainty) of sensory signals (Feldman & Friston, 2010; Friston et al., 2012a). Uncertainty is thought of as encoded mainly by synaptic gain that encodes the precision of random fluctuations about predicted states. It follows that neuromodulations of synaptic gain (such as dopamine and acetylcholine), do not signal (reward or pleasure) prediction errors about sensory data but the context in which such data were encountered. In other words, such neuromodulators report the salience of sensorimotor representations encoded by the activity of the synapses they modulate. This is important, especially in hierarchical schemes, where precision controls the relative influence of bottom-up prediction errors and top-down predictions.

In psychological terms, processing of salience expectancy allows the organism to control the significance it attributes to the sensory data it uses to update its predictions or, explain away prediction errors. As regards exteroception, this processing of salience can be seen as attention in perceptual inference (Feldman & Friston, 2010), and as affordance (latent action possibilities of cues in the environment) in active inference (Friston et al., 2012a). In interoception, optimizing the

precision of internal body signals can be seen as increased *interoceptive sensitivity and related feelings of arousal* in perceptual inference (N.B. this is not synonymous to increased prediction error about interoceptive signals, see below) and as increased *seeking behaviours* in active inference (see also Friston et al., 2012c). Understanding the 'objectless', so-called 'SEEKING' system (Panksepp, 1998) as the driver of a kind of enacted search for increased precision regarding internal body priors fits with what we know about the neurobiology of dopamine and related, bottom-up neuromodulators (Pfaff & Fisher, 2012; Friston et al., 2012a). Viewing the SEEKING system as supporting precision seeking also has intuitive meaning; we are motivated to sample the world when we do not know where surprise will come from and vice versa (Anselme, 2010).

One core aspect of consciousness may serve to register the aforementioned quality of 'uncertainty' and its inverse quality, precision. This view goes against the intuitive, long-standing view of core affective consciousness as monitoring hedonic quality, expressed by Solms in Freudian terms as the pleasure-unpleasure series. Instead, I propose that the core quality of this aspect of consciousness (as oppose to perceptual consciousness, see below) is a kind of certainty-uncertainty, or disambiguation principle. Certainty in this sense is not synonymous to prediction; i.e. it is not a measure of what was predicted, nor what occurred. Nor is it first and foremost the mental process that tells us what is good or bad for us homeostatically (although because of our innate constraints this is one common derivative of the certainty-uncertainty principle, see section on drives below). In this sense, consciousness is the process that tells us that we feel increased desire for and show approach tendencies towards unfamiliar, exotic and unpredictable foods, destinations and sexual partners not because we are predicting particularly rewarding experiences, but rather because we cannot predict such experiences with sufficient certainty. As I will argue below, it is instinctual emotions (the Id) and other innate priors that oppose this uncertainty principle and instead call for a relative automaticity and inflexibility in the system. If we leave Ego to its own devices, including its capacity for both perceptual inference and conscious disambiguation, it will lead the organism not to automaticity but rather to a never-ending and ultimately resource-draining, self-destructive cycle of seeking and cognitively finding (predicting and learning) of endless random fluctuations in the environment.

Before returning to the unconscious Id however, it is worth mentioning that a second type of consciousness can be conceived. Perceptual consciousness, both interoceptive and exteroceptive, may be instantiated as an instance of otherwise unconscious processes of perceptual inference about the causes of sensations. Indeed, it has recently been proposed that subjective feeling states arise from predictive inferences on the causes of interoceptive signals (Seth, Suzuki & Critchley, 2011). This ''interoceptive predictive coding'' model is compatible with the so-called James-Lange

theory of emotions to the degree that it claims that feelings are understood to arise from *perceptions* of physiological changes. Starting with the precise interpretation of James's work, classic debates in psychology have unfolded about whether bottom-up, direct bodily signals and/or top-down cognitive representations, categories or evaluations of physiological changes are responsible for feeling states. This model can specify the dynamic balance between bottom-up and top-down signals in interoception at various hierarchical levels, yet the interoceptive bodily self in this theory is always an inference (like Solms' objective body), i.e. it is inferred on the basis of generative models about the likely causes of one's interoceptive signals.

Contrary to Seth and colleagues, Solms views the core of affective consciousness as non-representational. As I proposed above, this aspect of consciousness can be best characterised as interoceptive sensitivity and precision seeking. Moreover, the dynamic source of affective consciousness and its most raw psychological manifestations may well depend on activity in the upper brainstem and limbic areas that Solms mentions. This implies a certain degree of functional segregation or modularity (see Fotopoulou, 2012b for discussion), and indeed as Solms' suggests a given hierarchy between more raw aspects of affective consciousness and more cognitivised aspects of consciousness. Nevertheless, the neural basis of the various affective qualities of consciousness is most likely generated at multiple and different levels of the hierarchy due to the functional integration (Friston, 1994), or the synchronisation (Engel, Fries, Singer, 2001) of *activity between such areas and cortical areas.*

The Inflexible Drives and Instinctual Emotions as Primitive Forms of Active Inference

This section stresses that our inherited motivational systems should not be equated with affective consciousness or salience, and moreover it is the drives (the Freudian Id, not the Ego) that call for a relative automaticity in both cognition and behaviour. Interestingly, this was the very point that Freud put forward about drives and the nirvana principle in 1920. The aim of minimizing free energy (and hence surprise) is to ensure that agents spend most of their time in a small number of 'valuable' states. Valuable states are not first and foremost conscious, pleasurable states as Solms implies (and Freud thought until 1920) but unsurprising states, i.e. states that evolution informs us our species most frequently occupied. Value, like free energy, depends on an organism's generative model and its implicit, heritable priors, optimised at different, evolutionary time scales; their job is to specify the innate value of certain attractive sensory states. These expectations thus include the prior that the organism itself (as part of the environment) occupies an invariant (attracting) set of physical (including internal) states. Valued states are therefore expected states. In other terms,

evolution equips an organism with optimised prior expectations about the states the organism is likely to encounter (these ideas are related to neural Darwinism; Friston, 2010).

However, as mentioned above, as priors are a mere hypotheses, the agent is evolutionary primed to test them by using sensory samples from the environment. Our primary expected states are therefore specified genetically but in one's lifetime they are fulfilled behaviourally, under active inference. Unlike the more object-less, exploratory SEEKING system mentioned above (Panksepp, 1998), the other instinctual, object-specific, primary e-motions described by Panksepp (1998) seem to fit exactly the role of primitive active inference in relation to innate priors. Reflexive, sensorimotor patterns are elicited to fulfil prior expectations about attractive sensory states of the organism, in the same way that classic reflexes elicit movement to fulfil prior expectations about proprioceptive sensations. In Freudian terms, it is the Id that calls for a relative automaticity and reduction of states to a minimum. This minimisation of non-evolutionary subscribed sensory states seems to be the ultimate guiding principle of our drives (innate priors), rather than the pleasure principle (homeostatically rewarding values). Indeed, as Freud suggested in 1920, the pleasure principle seems to be secondary to this minimisation imperative that governs the Id (the Nirvana principle that Solms now attributes to the Ego). In Friston's words, "the problem of finding sparse rewards in the environment is nature's *solution* to the problem of how to minimize the entropy (average surprise or free energy) of an agent's states: by ensuring they occupy a small set of attracting (that is, rewarding) states" (Friston, 2010 p. 135, emphasis added). It thus falls upon the Ego, or cognition, to tailor this inflexible, inherited minimisation imperative to the demands of the unpredictable world during one's life-time. Under perceptual and active inference it thus builds empirical priors on the foundations of innate priors. The Ego's 'cognitivised' generative models allow for a more flexible and efficient, yet motivationally constrained, relation with the ambiguous world. This includes retaining an optimal degree of instability in perceptual inference that allows it to explore alternative hypotheses about the causes of sensory states (Friston et al., 2012). Thus, while the world is ambiguous and potentially surprising and the Id strives to minimise the states the organism encounters to the very few that would satisfy basic, homeostatic needs, the Ego strives for an optimal balance between the two.

Conclusion

Thus in summary, I have used an influential computational theory of brain function, the free energy principle to suggest three points of added complexity to Solms' intriguing descriptions of the embodied mind: (a) most of the Ego may well be unconscious but the link between Ego and cognitive automaticity is not as straightforward as Solms suggests. Instead, cognition strives for both

inference *and flexibility* in relation to the changing world and the inflexible drives; (b) affective consciousness may primarily map the degree of *uncertainty* (not pleasure) of internal bodily signals. Subcortical areas are the neurobiological *sources* of this facet of consciousness that in itself is likely to be localised between many, distributed brain areas; and lastly (c) our innate motivational systems, the Id, ultimately serve the same optimisation principle as the Ego but unlike the later, they call for automaticity in behaviour, on the basis of innate unconscious priors that are fulfilled by instinctual e-motions and other reflexes, understood as evolutionary defined, primitive and inflexible forms of active and perceptual inference.

Of course the above speculative view of the motivated and embodied mind leaves unanswered more theoretical and empirical questions than those it attempts to answer. At least one important point of complexity I did not touch upon is the role of other agents in both perceptual and 'precision seeking' consciousness. Similarly, I cannot possibly do justice to complex notions such as 'repression' and the 'dynamic unconscious' in this brief commentary. Nevertheless, a few implications could be highlighted. It is easy to infer from what I wrote above how conflict between the demands of different innate priors (see Hopkins, 2012 for further discussion), as well as between the unconscious Id (which seeks to reduce all non 'prescribed' evolutionary states) and the conscious Ego (which seeks to represent and learn all novel signals in the internal and external environment) is therefore unavoidable. The conflict between the Ego and the Id for example may be why risk and danger both attract and scare us. It is also easy to see why Freud insisted on an antithesis between unconscious drives and the conscious feelings that originate in relation to them. Drives themselves (innate priors) are unconscious, minimally reflective (they are reflexively fulfilled by instinctual e-motions) and hence they can never be fully 'updated' by the Ego according to the changes in the external world (perceptual inference and learning). On the contrary, it is important that the Ego registers the core feelings that relate to the specificity of such innate predictions (the bottom-up modulation of the certainty of such predictions) so that the cognitive resources available for scanning the world and the body for novelty and salience are always constrained by, and in competition with, the high precision of our innate expectations. These speculative ideas do of course require further specification, proper modelling and empirical testing, but I hope they at least hold the potential of contributing some added 'precision' to Solms' rich, wide-ranging and thought-provoking view of the embodied mind.

References

Anselme P (2010) The uncertainty processing theory of motivation. Behav Brain Res 208: 291–310.

Carhart-Harris, R. L. and Friston, K. J. The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. Brain 2010: 133; 1265–1283

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

Engel, A.K., Fries, P., Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top–down processing. *Nature, 2*; 704-16.

Feldman H, Friston KJ. Attention, uncertainty, and free-energy. Front Hum Neurosci. 2010 Dec 2;4:215.

Feynman, R.P. (1972). Statistical mechanics. Benjamin, Reading, UK.

Fotopoulou, A. (2012a). Towards Psychodynamic Neuroscience. In A. Fotopoulou, M. Conway, & D. Pfaff. (Eds.) *From the Couch to the Lab: Trends in Psychodynamic Neuroscience*. *Oxford University Press*. pp. 25-46.

Fotopoulou, A. (2012b).Time to get rid of the 'Modular' in neuropsychology: A dynamic, unified theory of anosognosia as aberrant predictive coding. *Journal of Neuropsychology*, in press.

Freud, S., 1920. Beyond the Pleasure Principle. *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Vol. 18.* London: The Hogarth Press.

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. (2008) Hierarchical models in the brain. PLOS Comput. Biol. 4, e1000211

Friston, K. (2009a). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences.* 13, 293–301.

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

Friston K. (2011). Policies and Priors. Springer Series in Computational Neuroscience, Computational Neuroscience of Drug Addiction 1:10; 237-283.

Friston, K. Ao, P. (2012). Free-energy, value and attractors. Computational and mathematical methods in medicine, Article ID 937860.

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. (2012a). Dopamine, affordance and active inference. PLoS *Computational Biology* ; 8(1): e1002327.

Friston, K., Breakspear, M., & Deco, G. (2012b). Perception and self-organized instability. *Frontiers in Computational Neuroscience, 6*; 44: 1-19.

Friston, K., Adams, R.A., Perrinet, L., & Breakspear, M. (2012c).Perceptions as hypotheses: saccades as experiments. *Frontiers in Psychology, 3*; 151: 1-20.

Hinton GE, van Camp D (1993) Keeping neural networks simple by minimising the description length of weights. In: Proceedings of COLT-93, pp 5–13.

Hopkins, J. Psychoanalysis, representation and neuroscience: the Freudian unconscious and the Bayesian brain. In A. Fotopoulou, M. Conway, & D. Pfaff. (Eds.) *From the Couch to the Lab: Trends in Psychodynamic Neuroscience*. *Oxford University Press*. pp. 230-265.

Kaplan-Solms, K. & Solms, M. (2000). *Clinical studies in neuropsychoanalysis: Introduction to a depth neuropsychology.* London: Karnac Books.

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

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

Pfaff, D.W. & Fisher, H.E. (2012). Generalized brain arousal mechanisms and other biological, environmental and psychological mechanisms that contribute to libido. In A. Fotopoulou, M. Conway, & D. Pfaff. (Eds.) *From the Couch to the Lab: Trends in Psychodynamic Neuroscience*. *Oxford University Press*. pp. 64-84.

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

Seth, A.K., Suzuki, K. & Critchley, H.D. (2012). An interoceptive predictive coding model of conscious presence. Frontiers in psychology, 2; 395: 1-16.

von Helmholtz, H. (1866). Concerning the perceptions in general. In: Treatise on physiological optics, vol III, 3rd edn (translated by J.P.C. Southall 1925 Opt Soc Am Section 26, reprinted New York, Dover, 1962).