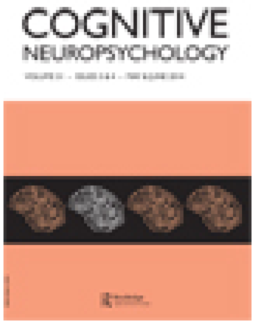

Cognitive Neuropsychology

ISSN: 0264-3294 (Print) 1464-0627 (Online) Journal homepage: https://www.tandfonline.com/loi/pcgn20

Hierarchical architecture of conscious processing and subjective experience

Theofanis I. Panagiotaropoulos, Liping Wang & Stanislas Dehaene

To cite this article: Theofanis I. Panagiotaropoulos, Liping Wang & Stanislas Dehaene (2020): Hierarchical architecture of conscious processing and subjective experience, Cognitive Neuropsychology, DOI: 10.1080/02643294.2020.1760811

To link to this article: https://doi.org/10.1080/02643294.2020.1760811

COGNITIVE NEUROPSYCHOLOGY
https://doi.org/10.1080/02643294.2020.1760811

ARTICLE COMMENTARY

Hierarchical architecture of conscious processing and subjective experience

Theofanis I. Panagiotaropoulos[a], Liping Wang[b] and Stanislas Dehaene[a,c]

[a]Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Sud, Université Paris-Saclay, Gif/Yvette, France; [b]Institute of Neuroscience, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, People's Republic of China; [c]Collège de France, Paris, France

ARTICLE HISTORY Received 10 December 2019; Revised 2 March 2020; Accepted 17 April 2020
KEYWORDS Conscious perception; prefrontal cortex; qualia; Global Neuronal Workspace; Attention-schema theory; internal models

How we consciously perceive the external world has troubled humans since the beginning of the scientific inquiry. Pre-Socratic philosophers attempted to explain visual perception through two opposing hypotheses of vision, the intromission theory (light from external objects enters the eye) versus the extramission theory (the eye emits light). Graziano et al. (2019) argue that the extramission theory and other popular beliefs that assign non-materialistic properties to consciousness originate from an implicit, automatic and inaccurate mental model of our own attention (Guterstam et al., 2019). Importantly, such a prescientific, folk-psychology model has the potential to bias not only the layman's views about the mind, but also the philosopher's views of consciousness such as the notion of qualia. Hence, it offers a potential explanation for the "meta-problem" of why so many people think that there is a hard problem of consciousness.

Graziano et al.'s proposal integrates Global Neuronal Workspace (GNW), Attention-schema (AST), higher-order (HOT) and illusionist theories of consciousness into a unified "standard model of consciousness". More specifically, it separates theoretical accounts of consciousness into those that pertain to i (information)-consciousness and to m (mysterious)-consciousness. I-consciousness concerns the selection and processing of sensory information that lead to conscious perception, while m-consciousness stands for the mysterious, experiential essence, commonly believed to accompany the contents of consciousness and formalized in the philosophical concepts of *res cogitans*, qualia, phenomenal consciousness etc. According to Graziano and colleagues, the belief in the existence of m-consciousness originates in an inaccurate, distorted and imperfect internal model that the human brain uses to represent i-consciousness, and which is also deployed for predicting the attention of others.

Currently, the main theoretical framework for i-consciousness is the Global Neuronal Workspace (GNW) hypothesis, a detailed neural implementation of the Global Workspace framework originally proposed by Bernard Baars (Baars, 1988). The theory predicts that sensory information gains access to consciousness by becoming available to a widespread neuronal workspace that involves higher-order association cortical areas, with the prefrontal cortex being a critical cortical node (Dehaene & Changeux, 2011). For GNW, consciousness serves a function: it evolved to break the modularity of non-conscious processing and broadcast information to a brain-wide network that makes information globally available for report (motor or verbal) and post-perceptual cognitive processing like working memory or decision making. There is no room, in the GNW, for any non-material qualia of experience; rather, the contents of the workspace *is* what we subjectively experience as a conscious feeling or experience.

We agree with Graziano et al. that a second and largely orthogonal level of processing is necessary to accurately capture human mental representations: a set of metacognitive or self-monitoring processes whereby the brain models a subpart of itself. This higher-order level of cognitive processing and the emerging subjective experiences it entails are likely

CONTACT Theofanis I. Panagiotaropoulos theofanis.panagiotaropoulos@cea.fr Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin center, Gif/Yvette 91191, France; Stanislas Dehaene stanislas.dehaene@cea.fr Cognitive Neuroimaging Unit, CEA, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin center, Gif/Yvette 91191, France, Collège de France, 11 Place Marcelin Berthelot, Paris 75005, France

to be responsible for the recursive sense of consciousness, i.e., "being aware of oneself" which is available to humans and currently missing from machines (Dehaene et al., 2017). In a recent review (Dehaene et al., 2017), these two dimensions of conscious computations were termed "global availability" (C1) and "self-monitoring" (C2).

Graziano and colleagues emphasize that the internal self-model could lead to illusions of the non-materiality of consciousness, free will, etc. because it is necessarily simplified and inaccurate, without access to information about the physical, neuronal mechanisms underlying i-consciousness. We do not disagree, but would also like to stress that, in the vast majority of cases, internal models involved in self-monitoring of i-consciousness are quite right – as they should if this "forward model" is to be useful to guide decisions, and if it comprises highly efficient learning mechanisms that have had a lifetime to adjust the self-model. Empirical work contains many examples where human subjects can accurately introspect about:

- Their errors: the brain contains a circuit for internal comparison of the intended action plan with the ongoing action, which accurately predicts errors before they actually occur (Charles et al., 2014). Importantly, metacognitive knowledge about whether a response is correct or not requires a conscious representation of the intended action.
- Their reaction time on individual trials (with the interesting illusion that, in full agreement with GNW theory, participants are not aware of the waiting time that occurs during dual-tasking, when one task is delayed because the GNW is already occupied by other conscious contents; (Marti et al., 2010)
- The movement of their attention: in a visual search task, objective response times correlate with subjective reports of the number of attention movements, over and above the objective number of items on the screen (Marti et al., 2014).

These examples also emphasize that the human self-model goes beyond a mere "model of attention". We do not understand why Graziano and colleagues limit their proposal to attention, unless this term expands much beyond its original meaning, to encompass all aspects of self-monitoring. Monitoring one's errors, accurately choosing among several strategies (Shrager & Siegler, 1998), or modelling one's own body schema (Chang et al., 2017) require internal models much richer than a single « attention model ».

What are the neural correlates of C1 and C2 levels of conscious processing (Dehaene et al., 2017)? Our current understanding of the neural substrates of the GNW suggests that C1 critically involves higher-order associational cortical areas of the prefrontal, temporal and parietal cortices. This conclusion rests on empirical evidence from functional magnetic resonance imaging, EEG, MEG and intracranial recordings in humans (Dehaene & Changeux, 2011) and direct recordings of neuronal activity in non-human primates (Kapoor et al., 2020; Panagiotaropoulos et al., 2012; van Vugt et al., 2018). The bulk of the evidence indicates that sensory information becomes consciously perceived following an ignition event, inferred from a non-linearity in the amplitude of the brain responses when going from non-conscious to conscious perceptual processing. Although there is a current debate about the contribution of behavioural reports to the widespread activation of cortical networks, electrophysiological recordings of single neurons and neural populations during no-report paradigms suggest that neural activity in prefrontal and parietal cortical areas encodes the contents of conscious experience even in the absence of any overt report (Panagiotaropoulos et al., 2012).

As concerns C2, the brain areas for self-monitoring remain understudied and ill-understood. If subjective experience is indeed the outcome of a metacognitive, self-monitoring or predictive model of the C1 level, which in turn is reflected in the global availability of conscious information across associational cortical areas, one might predict that cortical areas hierarchically higher than the GNW should be involved – and indeed there is evidence that Brodmann area 10, for instance, in the frontal pole, plays a key role in introspection and self-confidence (Fleming et al., 2010). Graziano et al. (2019). suggest that the C2 network comprises the temporoparietal junction (TPJ), superior temporal sulcus and dorsomedial prefrontal cortex as possible substrates – and indeed these areas operate as a self-oriented network involved in modelling one's own mental representations as well as those of others ("theory of mind"), and whose stimulation, lesioning or disconnection can lead to severe distortions in the sense of self (Heatherton, 2011).

Graziano and colleagues' proposal also hints at the possibility of a partial overlap and interlocking of the neural substrates for C1 and C2, which is an interesting possibility since it is known that prefrontal neuronal ensembles can encode simultaneously the contents of consciousness (C1) and abstract information about task phase (perhaps pertaining to C2) (Kapoor et al., 2018). Furthermore, recent evidence from recordings of single neurons in macaque monkeys suggest the existence of a hierarchically organized network involving dorsomedial prefrontal and anterior cingulate cortex as a substrate for hierarchical processing of information related to causal inference. Neural populations in both of these areas signal confidence for consciously perceived stimuli (Sarafyazd & Jazayeri, 2019) with anterior cingulate being downstream of dorsomedial prefrontal cortex and representing causal inference on longer time scales. Interestingly, prefrontal areas in the macaque brain also signal metacognitive self-monitoring of memory states (Miyamoto et al., 2017).

In another related line of research, premotor neurons encoding a subjective belief in hand ownership ("is the hand in this video my hand or not?") were recently recorded in the macaque monkey (Fang et al., 2019). This remarkable study provides evidence that a Bayesian causal inference model operates in premotor neural populations and is responsible for the rubber hand illusion. All these results from recordings in frontal cortical areas reinforce the general role of prefrontal cortex in multisensory causal inference and could indicate that self-awareness may also be best modelled in a Bayesian computational framework (Ehrsson & Chancel, 2019).

It is likely that self-monitoring processes are sensitive to the variability in the representation of conscious contents itself, resulting in perceived differences in the quality of subjective perception, for example each time we see the same red cup. We suggest that the source of this perceptual variability could lie in the variability of the responses of the neural populations that are involved in the representation of this cup. For example Bernasconi et al. (2011) found that noise in the EEG responses in frontoparietal and temporal areas of the human brain correlated with the perceived pitch and duration of auditory stimuli. These differences in the perceptual quality of identical stimuli are well predicted from signal detection theory (Micheyl & Dai, 2009; van Vugt et al., 2018). Due to these internally generated fluctuations, GNW should probably not be viewed as a completely stable neural network, but one whose contours vary even for a fixed visual input, therefore resulting in the activation of slightly different ensembles. This variability may be the source of differences in the subjective experience of the same object. Indeed, a common finding in neural recordings from cortical areas during paradigms of multistable perception is that the activity of some individual neurons is significantly correlated with subjective perception during the perception of competing visual stimuli, but not during perception without visual competition. Most importantly, variability in these neural representations can be identified not only in the same brain but also across brains, contributing further to differences in subjective experience.

In conclusion, we agree with Graziano and colleagues that, rather than looking for explanations of equivocal philosophical constructs such as "qualia", which probably originate from the fallacies of our intuitive self-monitoring mental models, a systematic investigation of the interactions between different hierarchical levels of cognitive processing, especially self-monitoring and conscious contents, may provide important insights into the nature of consciousness.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.

Bernasconi, F., Lucia, M. D., Tzovara, A., Manuel, A. L., Murray, M. M., & Spierer, L. (2011). Noise in brain activity engenders perception and influences discrimination sensitivity. *Journal of Neuroscience*, *31*(49), 17971–17981. https://doi.org/10.1523/JNEUROSCI.3715-11.2011

Chang, L., Zhang, S., Poo, M., & Gong, N. (2017). Spontaneous expression of mirror self-recognition in monkeys after learning precise visual-proprioceptive association for mirror images. *Proceedings of the National Academy of Sciences*, *114*(12), 3258–3263. https://doi.org/10.1073/pnas.1620764114

Charles, L., King, J.-R., & Dehaene, S. (2014). Decoding the dynamics of action, intention, and error detection for conscious and subliminal stimuli. *The Journal of Neuroscience*, *34*(4), 1158–1170. https://doi.org/10.1523/JNEUROSCI.2465-13.2014

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, *358*(6362), 486–492. https://doi.org/10.1126/science.aan8871

Ehrsson, H. H., & Chancel, M. (2019). Premotor cortex implements causal inference in multisensory own-body perception. *Proceedings of the National Academy of Sciences*, *116* (40), 19771–19773. https://doi.org/10.1073/pnas.1914000116

Fang, W., Li, J., Qi, G., Li, S., Sigman, M., & Wang, L. (2019). Statistical inference of body representation in the macaque brain. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1902334116

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*(5998), 1541–1543. https://doi.org/10.1126/science.1191883

Graziano, M. S., Guterstam, A., Bio, B. J., & Wilterson, A. I. (2019). Toward a standard model of consciousness: Reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cognitive Neuropsychology*, 1–18. https://doi.org/10.1080/02643294.2019.1670630

Guterstam, A., Kean, H. H., Webb, T. W., Kean, F. S., & Graziano, M. S. (2019). Implicit model of other people's visual attention as an invisible, force-carrying beam projecting from the eyes. *Proceedings of the National Academy of Sciences*, *116*(1), 328–333. https://doi.org/10.1073/pnas.1816581115

Heatherton, T. F. (2011). Neuroscience of self and self-regulation. *Annual Review of Psychology*, *62*(1), 363–390. https://doi.org/10.1146/annurev.psych.121208.131616

Kapoor, V., Besserve, M., Logothetis, N. K., & Panagiotaropoulos, T. I. (2018). Parallel and functionally segregated processing of task phase and conscious content in the prefrontal cortex. *Communications Biology*, *1*(1), 215. https://doi.org/10.1038/s42003-018-0225-1

Kapoor, V., Dwarakanath, A., Safavi, S., Werner, J., Besserve, M., Panagiotaropoulos, T. I., & Logothetis, N. K. (2020). Decoding the contents of consciousness from prefrontal ensembles. https://doi.org/10.1101/2020.01.28.921841

Marti, S., Bayet, L., & Dehaene, S. (2014). Subjective report of eye fixations during serial search. *Consciousness and Cognition*, *33C*, 1–15. https://doi.org/10.1016/j.concog.2014.11.007

Marti, S., Sackur, J., Sigman, M., & Dehaene, S. (2010). Mapping introspection's blind spot: Reconstruction of dual-task phenomenology using quantified introspection. *Cognition*, *115*(2), 303–313. (20129603). https://doi.org/10.1016/j.cognition.2010.01.003

Micheyl, C., & Dai, H. (2009). Likelihood ratio, optimal decision rules, and relationship between proportion correct and d' in the dual-pair AB-versus-BA identification paradigm. *Attention, Perception & Psychophysics*, *71*(6), 1426–1433. https://doi.org/10.3758/APP.71.6.1426

Miyamoto, K., Osada, T., Setsuie, R., Takeda, M., Tamura, K., Adachi, Y., & Miyashita, Y. (2017). Causal neural network of metamemory for retrospection in primates. *Science*, *355* (6321), 188–193. https://doi.org/10.1126/science.aal0162

Panagiotaropoulos, T. I., Deco, G., Kapoor, V., & Logothetis, N. K. (2012). Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron*, *74*(5), 924–935. https://doi.org/10.1016/j.neuron.2012.04.013

Sarafyazd, M., & Jazayeri, M. (2019). Hierarchical reasoning by neural circuits in the frontal cortex. *Science*, *364*(6441). https://doi.org/10.1126/science.aav8911

Shrager, J., & Siegler, R. S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*(5), 405–410. https://doi.org/10.1111/1467-9280.00076

van Vugt, B., Dagnino, B., Vartak, D., Safaai, H., Panzeri, S., Dehaene, S., & Roelfsema, P. R. (2018). The threshold for conscious report: Signal loss and response bias in visual and frontal cortex. *Science*. https://doi.org/10.1126/science.aar7186